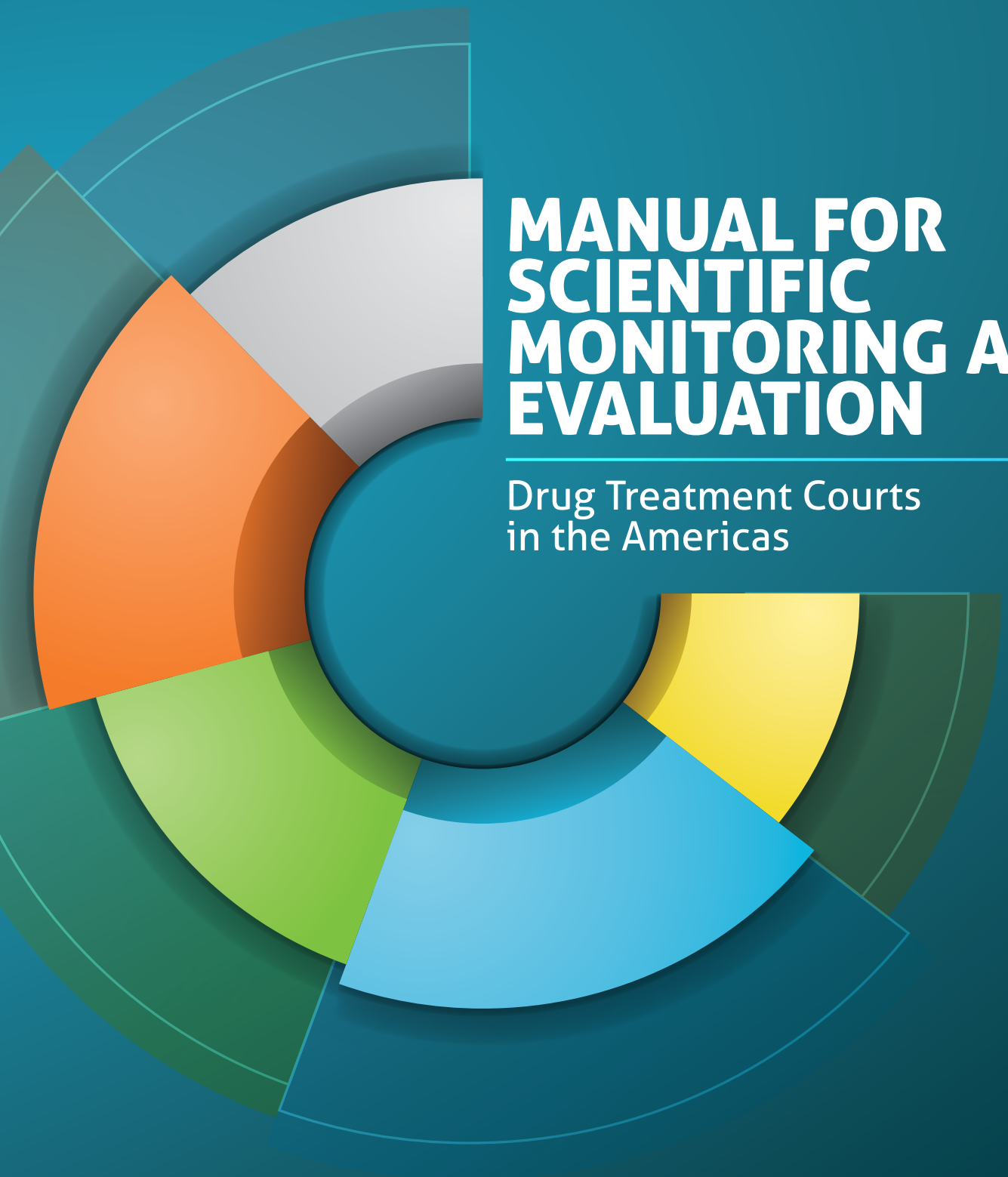


MANUAL FOR SCIENTIFIC MONITORING AND EVALUATION

Drug Treatment Courts in the Americas

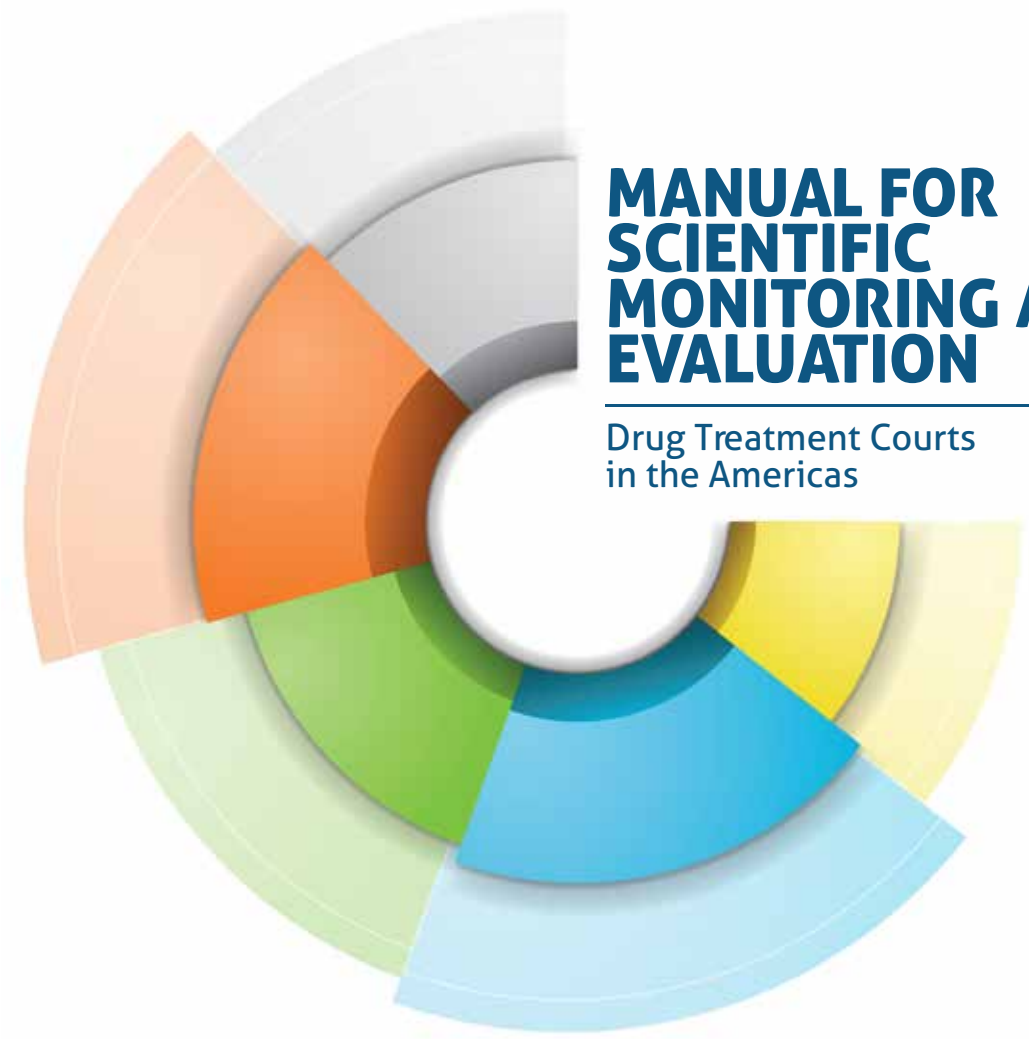
OAS | CICAD

OAS Cataloging-in-Publication Data

Inter-American Drug Abuse Control Commission.
Manual for scientific monitoring and evaluation [of] Drug Treatment Courts in the Americas.
v. ; cm (OAS. Official records ; OEA/Ser. L/XIV.6.61)
ISBN 978-0-8270-6830-8
1. Drug addiction--Treatment. 2. Substance abuse--Treatment. 3. Drug addiction--America. 4. Drug courts.
I. Title. II. Title: Drug Treatment Courts in the Americas. III. Marlowe, Douglas B. IV. Drug Treatment Court Program for the Americas. V. Organization of American States. Secretariat for Multidimensional Security. VI. Series.
OEA/Ser. L/XIV.6.61

ISBN 978-0-8270-6830-8

This publication was prepared by the Executive Secretariat of the Inter-American Drug Abuse Control Commission (ES-CICAD), Secretariat for Multidimensional Security of the Organization of American States (SMS/OAS). It was developed under the aegis of the Drug Treatment Court Program for the Americas, an initiative coordinated by ES-CICAD/SMS/OAS. This publication was made possible through the financial contribution of the Government of the United States of America, through the Bureau of International Narcotics and Law Enforcement Affairs (INL) of the U.S. Department of State, and of the Government of Canada, through the Anti-Crime Capacity Building Program (ACCBP) of Global Affairs Canada. The contents expressed in this document are presented exclusively for informational purposes and do not necessarily represent the opinion or official position of the Organization of American States, its General Secretariat, or its Member States.



MANUAL FOR SCIENTIFIC MONITORING AND EVALUATION

Drug Treatment Courts in the Americas

ACKNOWLEDGEMENTS

Special thanks to CICAD Assistant Executive Secretary Angela Crowdy, for her tireless support of CICAD's mission and OAS member states; to Ambassador Paul Simons, for initiating the dialogue that led to the drafting of this document during his service as CICAD Executive Secretary; to Douglas Boyd, for his research assistance in the development of this manual; to the National Association of Drug Court Professionals, the Canadian Association of Drug Treatment Court Professionals, and Fundación Paz Ciudadana, for their continued support; and to the participants in our workshops in The Bahamas, Chile, and Mexico, whose contributions were invaluable for the creation of this manual.

TABLE OF CONTENTS

"Scientific evidence should underpin the development of drug policy: In the past, too many public policy decisions on drugs were taken based on 'good intentions,' rather than on scientific evidence, and set unrealistic compliance goals. Now our member states seek to establish and implement evidence-based drug policies and actions with attainable goals. They are also aware that to achieve this status, they need to build institutions capable of developing objective, reliable, and comparative information."

—CICAD Executive Secretariat, Introduction to the Hemispheric Drug Strategy, 2010

A WELCOME FROM THE ORGANIZATION OF AMERICAN STATES

Dear Reader,

Thank you for your interest in this guide on drug treatment court (DTC) monitoring and evaluation. We commissioned its publication because the Secretariat for Multidimensional Security (SMS) seeks to guarantee greater security for communities in the Western Hemisphere. In order to do so, we must address the effects of the world's drug problem in our hemisphere.

Our mission to assess, prevent, confront, and respond effectively to security threats includes a mandate to seek new solutions for reducing and preventing crime, and for protecting human rights. In the Western Hemisphere, a clear relationship exists between crime, insufficient protection of human rights, and corruption. In the area of drug policy, these issues have allowed criminal organizations that traffic drugs and conduct other illicit activities to prosper.

Consequently, the OAS, through the Inter-American Drug Abuse Control Commission (known by its Spanish language acronym, CICAD) of SMS supports and promotes evidence-based drug policies that reduce crime, protect human rights, focus on the well-being of the person, and do not characterize the drug user as an object of the criminal justice system. In so doing, CICAD recognizes the need to promote drug policies and dialogue that focus on public health and human rights, while taking into consideration evidence-based policies, regulatory frameworks, civil society participation, and gender.

The DTC model is an excellent example of this type of evidence-based drug policy. Over two decades of robust academic inquiry demonstrates that when DTC programs follow evidence-based practices, they are effective in reducing criminal recidivism, saving public funds, protecting human rights, and helping participants

recover from debilitating substance use disorders. It may thus not come as a surprise that fifteen OAS member states have either implemented the DTC model or are exploring its implementation. Moreover, DTCs are consistent with the OAS Hemispheric Drug Strategy and corresponding Plan of Action on Drugs 2016-2020, serving as an alternative to criminal prosecution or imprisonment, and as a safe-guard of human rights.

Nonetheless, as the term "evidence-based" suggests, the continued success of the DTC model depends on sustained, rigorous scientific monitoring and evaluation by practitioners and the research community. This guide exists to support these efforts, and to make the fruits of over 20 years of academic experience readily available to new researchers—especially those in countries that are in the process of adapting the model to their own national contexts. It is written by some of the best and the brightest in the field, and covers every major aspect of DTC processes and outcomes. I hope you will find it as useful as its author, contributors, editors, and funders have designed it to be.

I extend my sincere gratitude to Canada and the United States of America for providing the resources to write and publish this guide. Their citizens' commitment to, and success with, the DTC model has served as an inspiration across the Hemisphere.

Dr. Farah Urrutia
Secretary for Multidimensional Security

FOREWORD

From my vantage point as the Executive Secretary of CICAD, drug treatment courts (DTCs) are one of the most important programs we have, in strict accordance with the OAS Hemispheric Drug Strategy and corresponding Plan of Action on Drugs 2016-2020. These two fundamental documents call for the means of offering treatment, rehabilitation, and recovery support services to drug-dependent criminal offenders as an alternative to criminal prosecution or imprisonment, and in full respect of human rights.

As its title indicates, the manual you are reading exists to teach researchers how to monitor and evaluate DTC programs more effectively. To the lay reader, a need to monitor and evaluate might suggest a critical problem with the model that needs fixing. In the case of DTCs, however, the reality is very much the contrary.

The need to monitor and evaluate DTCs stems from their success. As you will see in the following pages, the DTC model rests upon foundations of empiricism and research. Looking back to the model's inception in 1989, we do not find lofty theoreticians designing programs perched atop an ivory tower. Instead, the intrepid professionals who created the first DTCs did so from the field, responding to a real problem: they were practitioners mired in the crack cocaine crisis of that era. Confronted by a wave of crime—fueled by substance abuse—that threatened to overwhelm the criminal justice system, these men and women were desperate for an alternative response to a serious dilemma.

They saw the same offenders—whose substance use disorders were the primary factor driving their criminal activity—pass through their courts again and again. Jail was clearly not resolving that underlying issue. These defendants would serve their time, but their substance abuse problem always followed them,

shadow-like, back into civilian life. Predictably, many of them committed new crimes, were re-arrested, and sent back to jail. Meanwhile, their substance abuse problem remained untreated and unabated.

To use a concrete example, consider someone who steals bicycles to pay for a drug habit. Sending this person to jail leaves the substance abuse problem untreated—and sometimes even undetected. He or she serves the sentence, ends up back on the street with the same need to buy drugs as before, and the vicious cycle continues.

Faced with this seemingly intractable problem, these local practitioners forged a novel solution using tools they had on hand: the criminal justice and health sectors would come together to offer voluntary, court-supervised substance abuse treatment to substance-dependent defendants interested in participating.

The concept behind this new, more collaborative approach was straightforward: provide treatment and rehabilitation instead of exacting retribution. Of course, putting the model into practice was more challenging. There was no precedent or guidelines to follow. The task of coordinating practitioners from the different worlds of criminal justice and public health was not simple. Judges, prosecutors, defense attorneys, doctors, and social workers had to learn to work together as a team to help the DTC participants, instead of seeing one another as adversaries.

As such, at its outset the model ran on improvisation and goodwill. Practitioners progressively improved the model, keeping what worked and modifying what could be improved, but they observed that it seemed to function overall. DTC graduates appeared to do better than those who went to jail.

Such initial success drove an ever-greater interest in this new alternative to incarceration. By the mid-1990s, the research community had taken note. Researchers applied their talents to more rigorous evaluation of the DTC model,

to see if practitioners' anecdotal observations held up to scientific scrutiny. This application of the scientific method proved powerful. More formal evaluation of the model hastened its refinement and solidified its evidentiary support.

By 2006, the scientific community had concluded, from advanced research techniques called meta-analyses, that the DTC model indeed reduced criminal recidivism. Other studies also showed that the model also saved money over traditional criminal justice approaches. Moreover, research from other nations like Canada and Chile, which had adapted the DTC model to their own circumstances, began to confirm its usefulness in other jurisdictions and national contexts. By the close of that decade, more research had been published on the effects of adult DTCs than on virtually all other criminal justice programs combined.

Since then, the scientific community has begun consolidating the available research into guides for practitioners, such as the National Association of Drug Court Professionals' two-volume Adult Drug Court Best Practice Standards. Such documents serve a key role in ensuring that the results of very sophisticated research do not end up forgotten in a dusty academic archive, but instead directly improve how DTC practitioners work with program participants.

That brings us to the document you are reading right now. For the first time, the scientific community is presenting a comprehensive guide about the effective study of DTCs for use by practitioners, academics, and evaluators. This serves two important purposes. First, continued research is essential to keeping the DTC model robust and functional. The success the DTC model enjoys rests on empiricism. This volume represents a "how-to" guide to spur more empirical research. Second, as more countries adapt the DTC model to their particular needs, their own academic communities will need to monitor and evaluate processes and outcomes. This guide will accelerate the learning curve for researchers that are new to the field.

In other words, this document distills more than two decades of scholarly effort into one publicly available, accessible volume. We asked some of the field's most accomplished figures to serve as its author and contributors. These experts have covered effective monitoring and evaluation practices from start to finish.

As such, on behalf of CICAD, I humbly posit that if you wish to learn how to monitor and evaluate DTCs, you have come to the right place—and I hope that when you finish, you will agree that this guide was worth the time you invested in reading it.

Ambassador Adam E. Namm
Executive Secretary
Inter-American Drug Abuse Control Commission (CICAD)

Why Were Drug Treatment Courts Created?

Drug treatment courts (DTCs) were created to address the problem of crime that is driven by underlying substance use disorders. Substance use disorders are a major contributor to crime in the Western Hemisphere and around the globe (OAS, 2010). Reduce drug and alcohol consumption, and crime declines; increase it, and crime rises (Chandler et al., 2009; Holloway et al., 2006). Yet despite rising incarceration rates for drug-related crimes in the Americas, consumption and distribution of illegal drugs has continued to rise precipitously (OAS, 2011, 2012a).

DTCs emerged as one alternative to incarceration designed to address this issue. DTCs emphasize intensive community-based substance use treatment and supervision of addicted individuals charged with drug-related offenses, under judicial oversight. The DTC judge leads a multidisciplinary team of professionals that commonly includes representatives from the prosecutor's office, defense bar, treatment agencies, police, and probation department. The team members meet frequently to review participants' progress in treatment and make recommendations to the judge about suitable responses to impose for progress or lack thereof in treatment. These responses may include: (A) rewards such as verbal praise, reduced supervision requirements or small prizes for productive achievements; (B) sanctions such as verbal reprimands, community service or brief intervals of jail detention for infractions; or (C) adjustments to participants' treatment regimens in light of insufficient clinical progress.

Additionally, participants are tested for drug and alcohol use on a random basis, are required to complete a substance use treatment program, and must often satisfy restorative justice requirements, such as making restitution to victims. In pre-adjudication DTCs, successful graduates have their criminal charge(s) dropped or withdrawn, and in post-

adjudication DTCs, graduates may avoid incarceration or reduce the length or conditions of probation.

This approach has proven to work empirically. Extensive research on the model and its effects indicates that most DTCs have generated significant reductions in recidivism and drug use, savings for taxpayers, and reduced victimization from avoided crimes. For example, rigorous studies in Australia (Jones, 2011); Canada (Latimer et al., 2006; Somers et al., 2011); and the United States (Mitchell et al., 2012; Rossman et al., 2011) have demonstrated that DTCs can significantly reduce crime, promote recovery from addiction, and produce significant cost benefits more effectively than traditional criminal justice approaches. Therefore, although DTCs are not the only solution to crime fueled by substance use disorders, they do represent a promising, evidence-based model that can be adopted, adapted, and evaluated in other local and national settings.

This strong empirical foundation has driven a significant expansion of DTCs in the United States since the first such program was founded in 1989 in Miami-Dade County, Florida. As of 2015, there were more than 3,100 DTCs in the United States (National Institute of Justice, 2018), and that growth will likely continue. A 2017 report by the President's Commission on Combating Drug Addiction and the Opioid Crisis recommended further expansion and funding for DTCs at the federal level, and support for DTCs at the state level remains strong (President's Commission on Combating Drug Addiction and the Opioid Crisis, 2018).

The success of DTCs in the United States has also created demand for the model internationally. Since the turn of the century, other nations have adapted the DTC model for implementation in their own particular legal, social, and political contexts. Starting in the late 1990s and into the following decade, Canada, Bermuda, Cayman Islands, Chile, Jamaica, and Mexico started DTC pilot projects. By 2010, DTCs had expanded to the point where the Organization of American States, through the Inter-American Commission for Drug Abuse Control (CICAD), launched the *Drug Treatment Courts Program in the Americas*,

with the aim of supporting the expansion of the model to other member states that request assistance.

This CICAD program has supported other nations' adoption and adaptation of the model. By 2019, at least fifteen nations and two territories in the Americas had explored, developed, or implemented some type of DTC model: Argentina, Barbados, Belize, Bermuda, Canada, Cayman Islands, Chile, Colombia, Costa Rica, Dominican Republic, Guyana, Jamaica, Mexico, Panama, Peru, United States, and Trinidad and Tobago. The DTC model has also spread across the ocean to nations in other continents.

In most nations outside the United States, DTC programs are still in their formative stages, and efforts to evaluate their outcomes have only recently been initiated. Nonetheless, initial results are promising. A survey conducted by American University on behalf of the CICAD/OAS analyzed responses from officials in several OAS member states, including Canada, Mexico, Chile, Suriname, Bermuda, and Jamaica (Cooper et al., 2010). The majority of respondents reported that DTCs in their country appeared to reduce crime better than traditional correctional dispositions, and approximately half of the respondents reported achieving notable cost savings. A 2018 impact evaluation from Chile, for example, showed that DTCs there reduced recidivism by almost 30 percent compared to a control group, and a 2014 study reported savings of almost 75 percent (US$500 per participant per month) (Fundación Paz Ciudadana, 2018; Fundación Paz Ciudadana 2014). These figures suggest that implementing DTCs in South American and Caribbean nations is feasible and potentially desirable. In addition, a recent process evaluation of the treatment court in Nuevo León, Mexico, concluded that the program delivers high quality, evidence-based services, and was perceived positively by participants (Rempel et al., 2014).

This expansion and interest has continued through the present day. Consequently, as more and more countries have expressed interest in collaborating with the Organization of American States and its partners, it has become increasingly more important to promote the development of monitoring and evaluation capabilities necessary to sustain evidence-based DTC programs

(Rempel, et. al. 2014). That need is the primary motivation for this manual, which aims to assist professionals across the world with understanding how to monitor and evaluate DTC programs effectively.

About This Evaluation Manual

In the summer of 2010, the Inter-American Drug Abuse Control Commission (CICAD) of the Organization of American States (OAS) adopted the Hemispheric Drug Strategy which, among other provisions, encourages member states to develop alternatives to incarceration for individuals that suffer from a substance abuse disorder and who have been charged with crimes related to that disorder (OAS, 2010). Since then, and as one of those alternatives, by the request of a growing number of OAS member states, ES/CICAD/OAS has been offering training and technical assistance to help plan, implement, and evaluate new Drug Treatment Court (DTC) programs.

Subsequently, representatives from several OAS member states convened to establish a preliminary framework for evaluating DTCs (OAS, 2012b). A consensus was reached at this meeting that each country should evaluate its DTC programs to ensure transparency and accountability, identify effective and ineffective practices, attract external funding, and promote the merits of the DTC model. It was further agreed that member states should consider collecting a common dataset of core performance indicators which could serve as the basis for monitoring DTC activities and impacts in the Americas. These performance indicators should be SMART (i.e., Specific, Measurable, Achievable, Relevant, and Time-bound). They should also be value-neutral because benchmarks for success have not yet been established for many countries; therefore, there is no basis for knowing what levels of performance should be expected from programs in those countries. Collecting performance data will enable OAS member states to develop and validate performance benchmarks for their programs.

In the years that followed, CICAD/OAS led an effort to create and consolidate guidance for DTC monitoring and evaluation. Through dialogue with experts and member states, and carrying out process evaluations in seven nations, a consensus was developed over the state of existing research and certain core performance indicators, and written down in this manual. The manual was then subjected to extensive peer review by experts from both government and civil society organizations.

Although no benchmarks for success have yet been established for many OAS member states, it was agreed that this manual should present lessons learned in Australia, Canada, and the United States. Performance benchmarks from these three nations, while not some sort of "gold standard" for other OAS member states to copy blindly, may still serve as useful starting points for evaluators to consider when creating and testing their own scientific hypotheses. These empirical foundations took decades of painstaking observation and research, as is the case with many new concepts. For example, although the first DTC was founded in 1989, it was not until 2005 (sixteen years later) that the U.S. government agency responsible for evaluating the success of federally-funded programs concluded there was sufficient scientific evidence to prove that DTCs reduced crime (U.S. Government Accountability Office, 2005). It was not until 2011, twenty-two years after the first DTC was created, that researchers in the United States concluded that DTCs produced other benefits beyond reducing crime, such as reducing substance use and family conflict (Rossman et al., 2011). And it was only in 2012 that scientists in the United States completed the body of studies that were required to identify the best candidates for DTCs and some of the best practices that produce better outcomes in DTC programs (Marlowe, 2012a).

Accordingly, this manual is designed to allow other OAS member states to take advantage of these decades of experience. Hopefully, it will allow them to establish the effectiveness of their programs and identify best practices for their national context more quickly and efficiently.

LOGIC MODEL FOR DTC PROGRAM EVALUATIONS

Figure 1 depicts a logic model for monitoring and evaluating DTCs. Detailed descriptions of the activities depicted in the logic model are provided in the next several chapters of this evaluation manual. The logic model also includes an approximate timeline for performing various types of monitoring and evaluation activities.

As the diagram indicates, when a DTC is first established, the initial task for evaluators is to perform a *process evaluation* aimed at ensuring the DTC is following the intended model, treating the appropriate participants, and delivering effective services. Many of these activities will be conducted in the first year of the program, and then repeated over subsequent years to ensure program operations are not departing from best practices. Part of the process evaluation may also include estimating the costs of administering the DTC (*cost analyses*) and understanding how those costs are incurred by different agencies, such as the court, probation department, and treatment programs (*cost allocation analyses*).

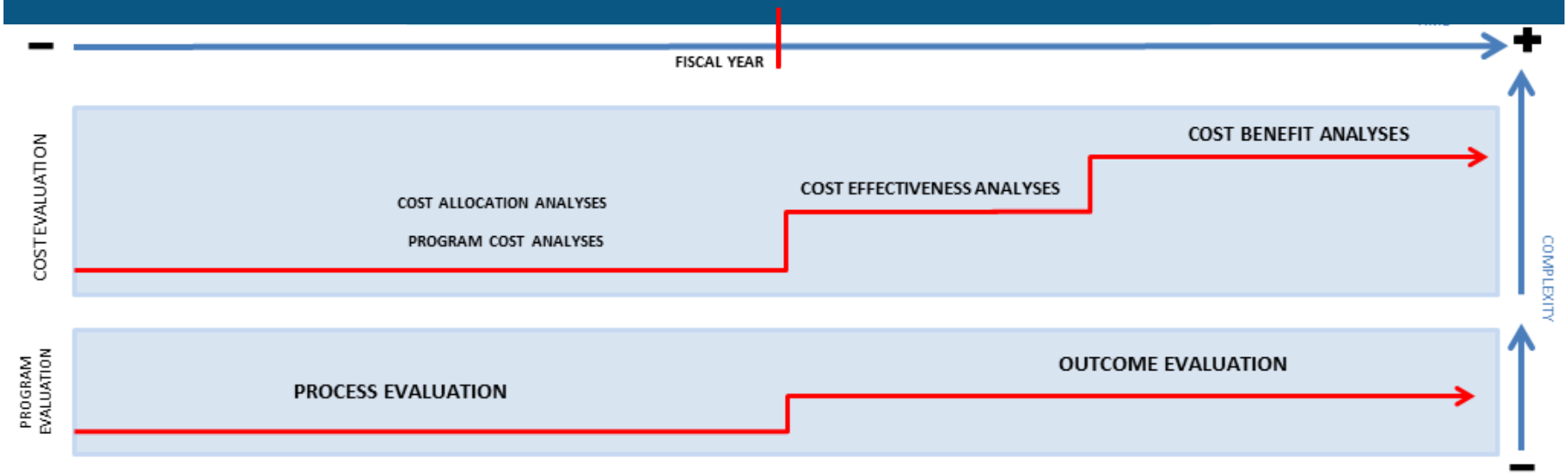
After the DTC has been in operation for at least a year (and often longer than that), evaluators may conduct an *outcome evaluation* to assess how participants are performing. Outcome evaluations may be conducted over the short term while participants are still enrolled in the program (*short-term outcome evaluations*), and/or after participants have been discharged from the program (*long-term outcome evaluations*). Finally, evaluators may determine over the longer-term whether the costs of administering the DTC are justified by resulting improvements in participants' adaptive functioning (*cost-effectiveness analyses*) and financial savings to participants and society, such as avoiding costs related to new arrests or new incarcerations or generating financial income from improved employment and education (*cost-benefit analyses*).

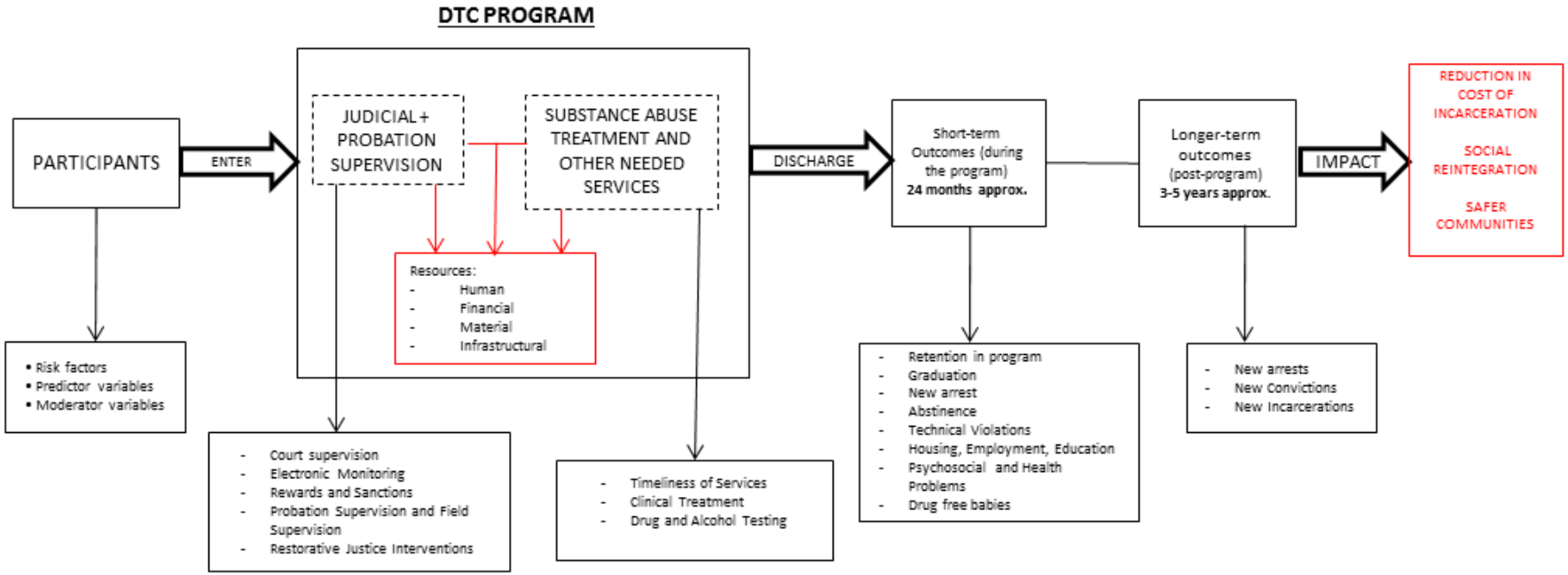
An important point to emphasize is that evaluators, DTC personnel, and policymakers should not become discouraged or overwhelmed by the potential scope of evaluation

activities. The adage "first things first" is particularly apt in this regard. Every DTC can begin at relatively modest cost and effort to measure the quality and quantity of its services.

As discussed in greater detail *infra*, information garnered from this process evaluation is likely to be highly predictive of long-term outcomes. If a DTC is delivering high-quality services and treating the right participants, it is very likely to reduce recidivism, enhance recovery from addiction, and generate cost benefits for society. Consequently, DTCs that lack sufficient resources to conduct high-quality outcome evaluations and cost-effectiveness evaluations can nonetheless begin the evaluation process by generating important and useful information. Later, if and when resources become available, they can increase the sophistication of its evaluation activities. Additionally, tracking participants' progress and recording/organizing data appropriately from the outset of the DTC program—a comparatively simple and low-cost proposition if done correctly—will greatly facilitate future monitoring and evaluation activities.

FIGURE 1: DRUG TREATMENT COURTS EVALUATION LOGIC MODEL





Finally, there are many ways to obtain useful information for DTC evaluations. For example, if evaluators want to measure how much treatment participants receive in the program, they can review the DTC's policies and procedures manual to determine how much treatment is intended to be provided, interview staff members or participants about how much treatment is commonly provided, or measure the actual number of sessions each participant receives. While the cost and effort required to obtain the information increases with the precision of the measurement, in many instances less precise measurements will still prove sufficient for a useful evaluation. The lesson for DTCs with limited budgets or evaluation expertise is that it is often possible to obtain adequate information at reasonable cost and effort.

TYPES OF PROGRAM EVALUATIONS

Program evaluations generally fall into one of four broad categories, depending on the scope and aims of the evaluation:

(1) Process evaluations,

(2) Short-term outcome evaluations,

(3) Long-term outcome evaluations, and

(4) Cost evaluations.

In addition, the term impact evaluation is commonly reserved for short-term and long-term outcome evaluations that include a comparison group. As will be discussed later, it is necessary to include a comparison group in order to determine whether outcomes produced by the DTC represented a significant improvement over what would have been achieved without the DTC. For example, if a DTC has a 60 percent re-arrest rate, this might seem high; however, if the comparison group had a re-arrest rate of 80 percent, then the DTC produced a significant positive impact.

Process Evaluations

Process evaluations indicate whether a program functions as planned, treats the intended target population of participants, and delivers the types and dosages of services that are likely to produce favorable outcomes. Questions that are commonly addressed in process evaluations include:

- What problems and barriers were encountered in implementing the program, and how were they resolved?
- How do DTC staff members, participants and stakeholders perceive the program's aims and effectiveness?
- What are the characteristics and needs of the participants, and are the services offered in the program adequate to meet those needs?

- Is the program serving its intended target population of participants?
- Does the program provide a sufficient dosage and quality of services that it is likely to produce favorable outcomes?
- What types of rewards and sanctions are typically delivered?
- What restorative justice requirements (e.g., victim restitution or community service) are being implemented, and do participants comply with those requirements?
- Are cases resolved more rapidly or less rapidly for DTC participants than for comparable individuals in the traditional criminal justice system?

As was noted earlier, several CICAD-sponsored stakeholder meetings were held in the course of developing this evaluation manual. Audience participants raised numerous additional questions of interest to their countries. Representatives from Chile, for example, stated that the following issues are critical for successfully installing a DTC in that country, and until such issues are resolved it is very difficult to engage in short-term and long-term planning for DTC programs:

- Is there sufficient political will to design and implement the DTC model in this country?
- Does the philosophy of the criminal justice system in this country focus predominantly on increasing social control of offenders, or does it also include an emphasis on providing adequate rehabilitation, treatment and social services?
- Is there an adequate budget available for purposes of developing and maintaining the DTC?
- Which staff members and stakeholders involved in the DTC are prone to implement the DTC model, and which ones may be reluctant or ambivalent about components of the model?
- What government sector(s) is in charge of administering the daily operations of the DTC?

- Is there a coordinating unit or agency with the political and technical support needed to lead development of the model during the implementation stage?
- Does this coordination unit have the power to develop and fund components of the DTC, convene other sponsors and stakeholders when needed, and intervene to resolve disagreements and barriers to effective implementation?
- Are the defense attorneys in the country reluctant to accept the DTC model (because, for example, they believe treatment and rehabilitation impose a higher penal burden than other approaches)?
- Do prosecutors believe the DTC model is too lenient compared with other approaches?
- Do judges consider the DTC model to be a function of the treatment or social service systems rather than the courts or justice system?
- Is there an adequate network of treatment and rehabilitation services for people suffering from substance abuse and mental health issues?
- Is there an adequate network of treatment and rehabilitation services willing to accept and treat persons charged with criminal offenses?
- Does the multidisciplinary team of professionals involved in the DTC possess sufficient knowledge and expertise related to both treatment and criminal justice interventions?
- Do treatment professionals involved with the DTC agree to coordinate their efforts with judicial authorities and share carefully defined treatment-related information?
- Is the DTC achieving the goals and milestones specified in the country's legislation or rules authorizing and establishing the program?

As stated earlier, process evaluations are usually the first order of business for DTC evaluators. They should ideally be performed early in the course of developing the program. Until it is established that DTC processes are being implemented correctly and that the program is serving the right people, there is little purpose in examining short- and long-term outcomes.

Fortunately, process evaluations can usually be performed at modest cost to the program because they often consist of descriptions of program operations and attitudes and skills of staff members—information that is relatively inexpensive to obtain.

Short-Term Outcome Evaluations

Short-term outcome evaluations (also referred to as proximal or intermediate outcome evaluations) measure participants' performance while they are still enrolled in the program. The evaluation typically focuses on outcomes during treatment that are likely to predict post-program performance. Examples of these so-called during-treatment outcomes that are commonly examined include:

- What percentage of participants successfully graduated from the program?
- What is the average length of stay in the program?
- What is the average attendance rate at treatment sessions?
- What proportion of urine drug tests was negative for all substances of abuse?
- What proportion of participants committed a new offense or technical violation while enrolled in the program?

These short-term outcomes are often meaningful in their own right to policymakers, the public, and other stakeholders of DTC programs. For example, one of the central aims of a DTC is to rehabilitate persons suffering from addiction. Therefore, helping participants successfully stay in and graduate from treatment is an important indicator of success.

Ultimately, however, the public and other stakeholders are likely to judge the merits of a DTC by how well it reduces crime, incarceration rates, and public expenditures. Therefore, it is important for at least some short-term outcome measures to be significant predictors of criminal reoffending and other long-term outcomes.

Fortunately, studies consistently find that longer retention in treatment and successful graduation from DTCs do, in fact, predict greater reductions in criminal recidivism (Carey et al., 2012; Gottfredson et al., 2007, 2008; Peters et al., 2002). Therefore, demonstrating

that a DTC increases treatment retention and graduation rates suggests that it will result in future reductions in crime.

Long-Term Outcome Evaluations

Long-term outcome evaluations (also referred to as distal outcome evaluations) measure participants' performance after they are no longer enrolled in the DTC program. Often, evaluators will report post-program outcomes for participants who successfully completed the DTC and those who were unsuccessfully discharged or voluntarily withdrew from the program. Common questions that might be addressed include:

- What percentages of participants were arrested for a new offense or convicted of a new offense after being discharged from the program?
- What percentage of participants was incarcerated for a new crime, and for how long?
- What percentage of participants was living in safe and drug-free housing at follow-up?
- What percentages of participants were gainfully employed or enrolled in educational programs?
- What percentages of participants were experiencing serious medical, psychiatric, or family problems at follow-up?
- What percentage of babies born to DTC participants was delivered healthy and drug-free?
- What percentage of participants was suffering from or engaged in domestic violence?
- What percentage of participants was still committing crime or abusing illicit drugs, but has not been in contact with the criminal justice system?

Long-term outcome evaluations are often difficult for DTCs to conduct on their own, because staff must locate participants after discharge to determine how they are doing. Few DTCs have sufficient resources to follow up with participants after discharge, and par-

ticipants are often reluctant to speak honestly with DTC staff about their performance. Participants may, for example, fear getting into trouble again with the criminal justice system if they admit to substance use or crime. Similarly, they may not want to disappoint their former therapists or other staff by acknowledging relapses or other treatment setbacks. In most instances, therefore DTCs will need to hire independent evaluators to conduct long-term outcome evaluations.

Cost Evaluations

Cost evaluations attach monetary values to the results of the above evaluations to estimate the net financial impacts of a DTC. In cost evaluations, distinctions are made between three concepts: (1) investment costs, (2) outcome costs, and (3) outcome savings. *Investment costs* represent the additional expenditures that were required to administer the DTC program, such as the added costs of treatment and frequent court hearings. *Outcome costs* are the expenditures that were incurred by taxpayers or the government to deal with the participants' subsequent behaviors, such as the costs of prosecuting new offenses or incarcerating participants for new crimes. *Outcome savings* are monies repaid to or reclaimed by society as a result of the improved functioning of DTC participants. For example, participants may find work and pay income taxes, contribute to the financial support of their children, or volunteer to work in charities or social service agencies. In Panama, for example, the Criminal Procedure Code specifies charitable work as a condition of rehabilitation, including volunteering to work in institutions of social welfare, education, public infrastructure, or local government with no payment from the state.

By comparing the financial expenditures and savings, economists estimate the net financial benefits of the DTC program. Questions that might be addressed in a cost evaluation include:

- What were the additional costs of providing treatment and supervision services for the DTC participants?

- What were the additional costs of holding frequent court hearings for DTC participants?
- How do the costs of the DTC compare to the costs of traditional approaches to case disposition?
- What were the outcome costs incurred by the DTC participants resulting from criminal recidivism or the use of public or government resources, compared to those of a control group?
- What outcome savings were produced as the result of increased employment or productivity on the part of the DTC participants?
- Were the investment costs of the DTC recouped by the outcome savings?

Many DTCs examine investment costs as part of a process evaluation or short-term outcome evaluation. Often, DTC personnel are capable of estimating how much staff time and other resources are required to administer the program.

It is more difficult, however, to estimate outcome costs and outcome savings based on reductions in crime and other long-term outcomes. Therefore, independent evaluators with expertise in cost evaluations are often needed to conduct cost-effectiveness analyses or cost-benefit analyses, which compare investment costs against long-term outcome costs and savings.

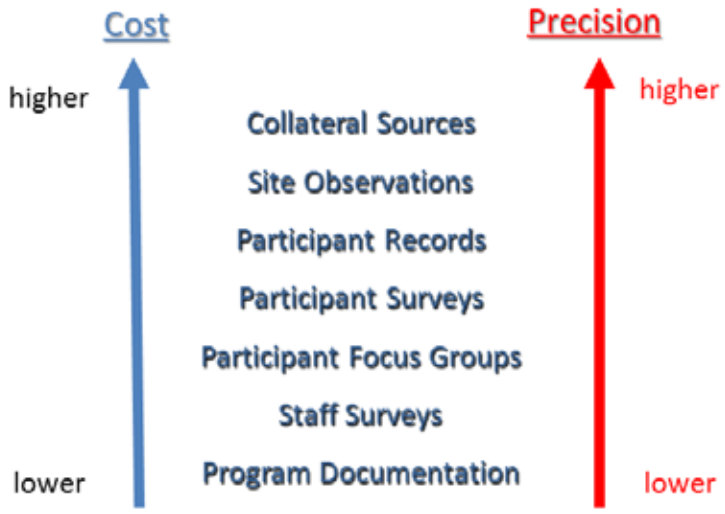
SOURCES OF INFORMATION

By some estimates, conducting a thorough program evaluation can require 10 percent to 25 percent of a DTC's operating budget (Peters, 1996). This is, of course, unsustainable for countries that are developing their DTC programs on a pilot basis or on a small budget as a proof of concept before committing substantial resources to the endeavor.

Fortunately, however, there are various ways to obtain information for a DTC program evaluation, and some methods will require considerably less effort and expense than others. Generally speaking, the more precise the source of the information, the more difficult and costly it will be for the evaluator to collect the data. Evaluators must decide how much effort and resources they are capable of devoting to measuring a given variable, and whether the added benefits of more precise measurement justify the additional time and expense that may be required.

Figure 2 depicts a hierarchy of data sources that may be used in a DTC evaluation, ranging from information that is relatively simple and inexpensive to collect at the bottom of the figure to information that is more costly and complex to collect at the top of the figure. The higher in the figure the source of the information, the more precise the data will be, but also the more difficult and costly it will be to collect that data.

FIGURE 2: SOURCES OF INFORMATION



Consider, for example, the various ways in which an evaluator could measure the frequency and impact of court hearings in a DTC. The evaluator could review the program's policies and procedures manual to determine how often court hearings are intended to be held, interview staff members or participants about how frequently court hearings are typically held, or extract data from program charts or records on actual attendance rates at court hearings. The evaluator could also observe court hearings directly or interview participants' friends and family members to determine how impactful the hearings tend to be.

Each of these approaches carries with it benefits and burdens that must be taken into account by an evaluator when planning an evaluation design. By carefully comparing the advantages and disadvantages of each in advance, the evaluator can design an evaluation that will both meet the necessary objectives and stay within its allocated budget.

Program Documentation

Perhaps the simplest and least costly method is to review the DTC's policies and procedures manual or other documentation to determine how frequently court hearings are planned to be scheduled. The policies and procedures manual might specify, for example, that court hearings should be scheduled every two weeks during the first phase of the program and monthly thereafter, unless there is a good reason for the judge to schedule hearings more or less frequently. Although this might seem like a blunt and imprecise method of measurement, studies have found that it can, in fact, significantly predict outcomes in DTCs and differentiate effective from ineffective programs (Carey et al., 2012; Shaffer, 2010; Rossman et al., 2011).

There are obvious advantages to this approach. First, the information can be collected quickly and with minimal effort by research staff. Second, it is unnecessary to wait for a substantial number of participants to enter and complete the program before being able to collect attendance information at court hearings. A review of program documentation can be accomplished before the DTC opens its doors or in the first year of operations.

Of course, programs do not always adhere to what is written in their policies and procedures manual, and some programs may not have a manual. Practices also have a tendency to change or "drift" over time. A DTC might start out with the intent of holding frequent court hearings, but then find itself scheduling hearings less frequently due to time or resource constraints. Greater precision is likely to be achieved by measuring what actually happens in the DTC rather than what people thought would happen when the program was first developed or the policies and procedures manual was last updated.

Staff Surveys

The next simplest and least costly approach is to survey staff members about how often court hearings are typically held in the program. Most DTCs have a core team of approximately six to ten staff members, including the judge, program coordinator, and representatives from the prosecution, defense, treatment programs, probation department, and law enforcement. It is usually possible to survey this small group of professionals with minimal expense and inconvenience.

In several large-scale studies in the United States, evaluators used web-based survey instruments (Carey et al., 2012; Shaffer, 2010) or computer-assisted personal interviews (CAPI) (Rossman et al., 2011) to collect information from staff members in dozens of geographically dispersed DTC programs. The web-based surveys gathered self-report information from staff members who logged in to encrypted databases using secured usernames and passwords. In the CAPI studies, the interviewers used computers with software that guided them through a structured interview process and allowed them to record responses directly into an electronic database.

In both types of studies, the staff members' responses were entered into a database that automatically tabulated and summarized the results. The Appendix to this manual provides information on how to obtain staff member surveys that have been used successfully in large-scale DTC program evaluations.

There are several limitations to this method, however. Staff members may not recall accurately how often court hearings were held, or they may be motivated to present the DTC in an undeservedly favorable light. Evaluators may also find it necessary to reconcile discrepancies in staff members' responses. For example, if the judge reported that court hearings typically occur every two weeks, but the prosecutor reported they occur every four weeks, the evaluator would need to contact the judge and prosecutor again to resolve the discrepancy.

Participant Focus Groups

Another method is to organize focus groups with a subset of participants to gauge their perceptions about the program. Participants may more likely than staff to offer neutral or negative observations about the DTC, and they may also be more likely to recall accurately how often they were required to appear before the judge in court. Another advantage to focus groups is that the evaluator does not need to interview every participant in the program. Information can be collected from a manageable number of participants at reasonable cost and effort.

Ordinarily, participants should be randomly selected to participate in focus groups to ensure they represent the views and experiences of most participants in the program. It is often not appropriate, for example, to select only successful graduates to participate in focus groups, because graduates may have very different perceptions about the DTC or experiences in the program than other participants. Graduates may have been successful precisely because they attended more frequent court hearings than other persons who were unsuccessfully discharged from the program. If this is the case, then relying solely on graduates' responses might lead the evaluator to conclude, erroneously, that the DTC holds more frequent hearings than it actually does for many participants.

Participant Surveys

A more precise, but also more costly, method is to survey all or most of the participants in the program. This approach ensures the evaluator will assess the full range of participants' experiences in the program, and therefore the results are more likely to be representative of how the DTC operates in practice.

Because each participant will report on his or her own attendance in court, the evaluator will obtain a range of scores for the entire sample of participants. Having a range of scores will allow the evaluator to calculate such measures as the average number of hearings that were attended, as well as variability in the number of hearings that were attended, such as the range or standard deviation. As will be discussed later, collecting a range of scores allows the evaluator to employ more sophisticated statistical techniques and answer more sophisticated questions.

Participant Records

Another more precise method is to examine data from court or treatment records concerning participants' attendance at court hearings. Assuming the information is entered accurately, this method may yield the most precise measurement of attendance at court hearings, and the evaluator would obtain a range of scores on the entire sample of participants. Measuring actual attendance rates avoids concerns that staff members or participants may not recall accurately how often hearings were held. It also avoids concerns that some staff or participants may, consciously or unconsciously, present the DTC in an unduly favorable or unfavorable light.

A major limitation of this approach is that it relies on information being entered accurately into participant records. Unfortunately, it is not uncommon for evaluators to encounter missing or incomplete information in DTC records. Staff members in DTCs are often very busy, and may not give the relatively less exciting task of data collection and entry the attention and care it requires. If information is not entered accurately into participant re-

cords, evaluators may have no option but to conduct surveys or focus groups in an attempt to make up for missing or incomplete information.

Site Observations

Evaluators often find it highly enlightening to observe DTC operations directly. For example, if an evaluator wishes to examine the quality and impact of court hearings, there is often no substitute for conducting on-site observations of court proceedings. The evaluator can assess, for example, how long the judge interacts with participants in court, whether the judge gives participants a chance to explain their views, and whether other team members, such as the defense attorney or prosecutor, contribute valuable input. This level of qualitative information is difficult or impossible to obtain from program records or participant surveys.

Needless to say, collecting this level of information can be costly and resource-intensive for evaluators. Not only must evaluators attend a sufficient number of hearings to get a reliable picture of how the program operates, but evaluators must also be trained carefully on how to record reliable and valid data concerning their observations. It is often necessary to use standardized rating scales to ensure information is recorded in an unbiased manner and useful format. Additional information about observer rating scales is provided in later chapters of this manual.

DTC staff should not conduct site observations themselves, as they cannot be expected to provide accurate and unbiased appraisals of how effectively they interact with participants or perform their roles. Instead, well-trained, independent observers are required for this task. Site observations are therefore generally viewed as an advanced evaluation procedure, and are rarely performed outside of well-funded evaluation projects.

Collateral Sources

Finally, evaluators may find it useful to collect information from participants' friends, family members, or other individuals who are knowledgeable about participants'

experiences in the program. For example, participants may view frequent court hearings as an unnecessary burden on their time, especially during the early phases of treatment. Family members, however, may have a very different view about the importance of court hearings—seeing them, for example, as indispensable to keeping their loved one law-abiding and abstinent from drugs and alcohol. If an evaluator relied solely on participant perceptions to assess the value of court hearings, he or she might reach incorrect conclusions (e.g., that court hearings should be scheduled less often, when in fact frequent hearings were a critical element for success).

Interviewing collateral sources can be difficult for many DTCs. Participants may be reluctant to have staff members speak with their friends or family members, and it may be difficult to convince friends or family to come to court and be interviewed by criminal justice authorities. Often it is necessary to have trained and independent evaluators conduct collateral interviews, which can be costly and difficult for many DTCs.

OPERATIONALIZING VARIABLES

In 1997, the National Association of Drug Court Professionals in the United States published what are commonly referred to as the 10 Key Components of DTCs (NADCP, 1997). Several years later, the International Association of Drug Treatment Courts (IADTC) adopted the 10 Key Components, and added three additional components:

(1) Focusing on case management to address ancillary needs of participants and promote their social reintegration;

(2) Ensuring individualized treatment to address the needs of special populations, such as women, participants with co-occurring disorders, indigenous populations, and ethnic minorities; and

(3) Provision of aftercare recovery services.

The IADTC principles are embodied in a document entitled the 13 Key Principles for Court-directed Treatment and Rehabilitation Programs ("13 Key Principles").

INTERNATIONAL ASSOCIATION OF DRUG TREATMENT COURTS (IADTC)
13 Key Principles for Court-Directed Treatment and Rehabilitation Programmes

(1) The programmes integrate substance dependency treatment services with justice system case processing.

(2) Using a non-adversarial approach, prosecution and defense lawyers promote public safety while protecting offenders' due process rights.

(3) Eligible offenders are identified early and promptly integrated into the programme.

(4) The programmes ensure access to a continuum of substance dependency treatment and other rehabilitation services.

(5) Compliance is objectively monitored by frequent substance abuse testing.

(6) A coordinated strategy governs responses of the court to programme non-compliance (and compliance) by offenders.

(7) Ongoing judicial interaction with each offender in a programme is essential.

(8) Monitoring and evaluation measure the achievement of programme goals and gauge effectiveness.

(9) Continuing interdisciplinary education promotes effective planning, implementation, and operations of these court-directed programmes.

(10) Forging partnerships among courts directing treatment programmes, public agencies, and community-based organizations generates local support and enhances programme effectiveness.

(11) Ongoing case management includes the social support necessary to achieve social reintegration.

(12) There is appropriate flexibility in adjusting programme content, including incentives and sanctions, to better achieve programme results with particular groups, such as women, indigenous people and minority ethnic groups.

(13) Post treatment and after-care services should be established in order to enhance long term programme effects.

The 13 Key Principles represent a revolutionary change in thinking about how to successfully rehabilitate individuals suffering from addiction who are charged with drug-related crimes. It should be immediately apparent, however, that many of these Principles are described in very general terms.

Consider, for example, Principle 7, which states that DTCs should provide "ongoing judicial interaction with each offender." How does one define "ongoing?" Is the frequency of face-to-face contacts sufficient to measure this variable? If it is, how frequently should the judge meet with participants? Should the judge hold hearings on a daily, weekly or monthly basis, or should the frequency of hearings be based on the needs of each participant? Similarly, what does "interaction" mean? Must the judge speak directly to participants, give them a chance to voice their own views, or is it sufficient to simply review each participant's progress and impose consequences?

Evaluators are faced with the difficult task of defining these Principles in a manner that permits them to be measured and examined objectively. The process of defining variables in measurable and objective terms is referred to as *operationalizing* the variables. Different evaluators are not required to agree on the same definition of a variable, but each must make it painstakingly clear how he or she is defining that variable. If the definition of a variable is not clearly described, it will not be possible for other evaluators to replicate the study or interpret the findings.

For example, it is possible to measure "ongoing judicial interaction" in a number of ways. An evaluator could measure how often participants appeared before the judge in court. Alternatively, the he or she could measure how long the judge spoke with each participant during court hearings, or could rate how attentive or encouraging the judge appeared to be during his or her interactions with participants. There is no one correct way to operationalize "judicial interaction" or any of the other 13 Key Principles of a DTC program. Instead, deciding how to operationalize a variable is essentially a value judgment that is often based on the following considerations:

- **Burden of Collection**: Some variables are considerably simpler and less costly to

measure than others. For example, it is much easier to measure the number of court hearings that participants attended than to use a stopwatch to time how long they appeared before the judge in court. Even greater effort would be required to rate the quality of the judge's interactions with the participants.

- **Reliability of Measurement**: Any evaluator should be able to reliably measure the number of court hearings that participants attended. However, two different evaluators could reach different conclusions about how "attentive" or "encouraging" the judge appeared to be during his or her interactions with participants. To avoid this problem, substantial effort and expense may be required to use rating instruments that yield comparable findings across different evaluators. This time-intensive process is referred to as ensuring *inter-rater reliability*.
- **Cultural Significance:** Variables may have greater or lesser significance in different contexts or cultures. For example, some cultures may place great value on showing respect for authority figures. Citizens in those cultures might be socialized from an early age to avoid looking authority figures in the eye, disagreeing with authority figures, or speaking frankly to authority figures. Rather than improving outcomes, longer and more personalized interactions with a judge might cause anxiety or discomfort for such individuals. Therefore, it is necessary to validate the impact of any variable in different countries or cultures. The fact that personalized interactions with a judge might improve outcomes in the United States does not mean necessarily they will have the same effect in other cultures.

Measurement Precision

Variables may also be operationalized differently depending on the precision of measurement that the evaluator desires. Generally speaking, the more precisely a variable is measured, the more sophisticated the questions it can address. Consider, for example, how an evaluator might measure the provision of substance use treatment in a DTC. Put aside for the moment the more complicated question of how to measure the quality of treatment, and consider how to measure the amount of treatment that was provided. This

variable could be measured in at least three ways:

- **Dosage**: The number of treatment sessions that were attended.
- **Attendance rate:** The ratio of the number of treatment sessions that were attended, divided by the number of treatment sessions that were originally scheduled.
- **Density**: The number of treatment sessions that were scheduled or attended per unit of time, such as per month or per phase of the program.

Measuring the dosage of treatment is relatively straightforward and requires only a simple tally. Measuring the attendance rate is a bit more complicated because it requires the evaluator to collect two pieces of information: the number of sessions that were attended, and the number of sessions that were originally scheduled. Although this might seem like a minor burden, experience indicates that many DTC programs do not do a good job of entering missed appointments into a database or participant records. This is sometimes referred to as the "problem of the missing denominator" because the denominator in the attendance ratio (the number of sessions that were originally scheduled) is unknown. To avoid this problem, it is essential for staff members in DTCs to record whether each scheduled appointment was kept, not kept, rescheduled, or excused. As a rule of thumb, sessions are counted as having been rescheduled or excused only if (1) a staff member gave permission in advance for the session not to occur, or (2) the participant provided objective documentation that he or she was unable to attend the session, such as providing a doctor's note confirming that the participant was ill or injured.

Measuring the density of treatment is even more complicated because it requires what is called *date-stamping*. This means that each entry must be linked to the date on which the event occurred or should have occurred. For example, attendance at each counseling session must be linked to the date on which the session was held. Date-stamping can be critical for predicting outcomes in DTCs. For example, a participant who has irregular attendance in treatment and takes a year to complete twenty-six sessions of counseling is likely to have poorer outcomes than another participant who completes the full sequence of twenty-six sessions on time within three months. Without date-stamping, there might be no

way to differentiate these two cases and critically important information could be lost.

Although these might seem like trivial matters, many DTC program evaluations have been unable to report appreciably more than dosage information because data on missed appointments or the dates on which appointments occurred were never recorded. Perhaps the evaluators did not anticipate this problem or perhaps they were unable to convince program staff members to record the information more precisely. Either way, greater efforts may be required to perform more sophisticated analyses in DTC evaluations.

Evaluators must decide for themselves whether the additional effort that will be required to measure a variable more precisely is justified by the ability to answer more sophisticated questions that the additional information provides. One important issue to consider in this regard is the *redundancy* of the information. For example, the number of treatment sessions that participants attended is part of the calculation for all three variables of dosage, attendance rate, and density. As a result, these three variables may be significantly, although only partially, correlated with each other. In other words, they may be partially redundant. Variables that are partially redundant might predict outcomes similarly, although not equally. Density might, for example, predict outcomes better than dosage, but perhaps the degree of prediction that is provided by dosage alone would be good enough for many evaluators' purposes.

Ideally, evaluators should look to empirical findings to determine whether more precise measurement is needed for their evaluations. Where information is available on the redundancy of alternative measures, that information is provided in this evaluation manual.

This section also underscores how important and valuable it is for DTC staff to make robust data collection, entry, and management a priority. Entering participant data correctly into DTC records as it is generated is relatively straightforward if the DTC has the proper systems in place to identify, collect, enter, and manage relevant data, trains staff appropriately, and periodically verifies that these systems and procedures are being properly implemented. Doing so is certainly less expensive and challenging than having to design subsequent monitoring and evaluation efforts around incomplete or missing data.

Participant Characteristics

Participant characteristics refer to attributes of participants that predate their entry into the DTC. These typically include:

- Socio-demographic variables, such as age, gender, race, and employment status;
- Clinical variables, such as primary substance(s) used, psychiatric or substance use diagnoses, and history of substance use or mental health treatment; and
- Criminal history variables, such as prior arrests, convictions and incarcerations.

Many DTCs and probation departments administer risk and need assessment tools to match participants to appropriate levels of treatment and supervision, and scores on these tools are also important variables for assessing participant characteristics. Further information about risk and need assessment tools is discussed in later chapters of this manual.

Participant characteristics are referred to as predictor variables if they correlate significantly with outcomes in DTC programs. They are referred to as risk factors if they correlate with poorer outcomes, such as higher recidivism rates. They may also be referred to as *moderator variables*, but only if they differentially predict outcomes in DTCs as compared to alternative programs, such as probation. For example, if females performed significantly better in a DTC than probation, but males performed better on probation, then gender would be a moderator variable for the effects of the DTC. As will be discussed later, determining whether a participant characteristic is a moderator variable as opposed to merely a predictor variable or risk factor requires the use of more sophisticated statistical techniques than simple correlations.

In this context, the DTC model as embodied in the 13 Key Principles makes two fundamental assumptions about participants in the program. It assumes that participants are (1)

addicted to illicit[1] drugs or alcohol, and (2) unlikely to improve their behavior in a program that is less intensive than a DTC, such as probation or a referral to treatment. Participants who satisfy both conditions are referred to as *high need* and *high risk*. This means they have a high need for substance use treatment and are also at high risk for failing to complete treatment unless a judge and team of other professionals hold them accountable for complying with treatment.

If a DTC serves participants who do not have these characteristics, it might find itself providing treatment to people who do not need treatment or providing judicial supervision to people who do not need judicial supervision. Studies from Canada and the United States have revealed that treating low-need or low-risk individuals in DTCs has the potential to waste scarce treatment and judicial resources (DeMatteo et al., 2006). It can also make outcomes worse for low-risk or low-need participants by exposing them to antisocial peers or interfering with their involvement in productive activities, such as work, school, homemaking or childcare (Lowenkamp & Latessa, 2004; Taxman & Marlowe, 2006).

The participant characteristics listed below have consistently been found to be predictor variables or risk factors in DTC program evaluations in several countries (Marlowe et al., 2003; Newton-Taylor et al., 2009; Somers et al., 2012).

- Current age
- Gender
- Education (number of academic grades completed)
- Employment (full time, part time, seasonal or temporary, recently unemployed, chronically unemployed)
- Marital status (married, never married, divorced, widowed, cohabitating)
- Race
- Ethnicity

1 Illicit drugs include prescription medications that are used for a non-prescribed or non-medically-indicated purpose.

- Nationality
- Number of prior criminal convictions
- Months of prior incarcerations
- Prior substance use treatment episodes
- Age of onset of substance use
- Age of onset of delinquent or criminal activity
- Stable and drug-free housing (yes or no)
- Proportion of time spent interacting with other individuals engaged in crime or substance use (from "none" to "most")
- Diagnosis of substance dependence / addiction vs. substance abuse
- Primary substance(s) used (e.g., opiates, cocaine, amphetamines, sedatives, hallucinogens, marijuana, alcohol)
- Co-occurring diagnosis of major psychiatric disorder (i.e. major depression, bipolar disorder, psychotic disorder, post-traumatic stress disorder)

Most of these variables are simple and inexpensive to collect and should be examined in any DTC program evaluation. Including these variables in moderator analyses can assist the evaluator to determine which participants were helped by the DTC, and which participants might not have been helped or perhaps even harmed. Performing moderator analyses can prevent an evaluator from reaching the unfounded conclusion that a DTC did not work, when in fact the real problem might have been that it treated the wrong people.

Better prediction is often achieved by using structured questionnaires or interviews that combine several of these predictor variables into scales or summative scores. *Risk assessment tools* measure the likelihood that participants will commit a new offense or fail in treatment unless they receive intensive supervision such as that provided in a DTC. *Clinical assessment tools* evaluate the diagnostic criteria for substance dependence or addiction, as well as the symptoms of other major psychiatric disorders. The Appendix provides links to risk assessment tools and clinical assessment tools that have been used successfully in DTC

program evaluations. Several of these tools have been translated into Spanish or validated for use in South American countries.

Note that the above list includes only variables that have been shown to correlate with outcomes in DTCs. As was noted earlier, most DTC studies have been conducted in the United States, Canada and Australia, and it is possible that other variables may emerge as significant predictors of DTC outcomes in South American and Caribbean nations. These variables may also be defined or measured differently in other countries.

Performance Indicators

Performance indicators differ from participant characteristics in that they reflect events occurring after participants entered the DTC program. There are two general types of performance indicators. *Program-level* performance indicators (also referred to as *inputs*) represent the services or interventions that were delivered in the DTC program. *Participant-level* performance indicators (also referred to as *outputs*) represent the impacts of the DTC on the participants. Examples of program-level performance indicators might include how often court hearings or treatment sessions were held. Examples of participant-level performance indicators might include how often participants tested negative for alcohol and illicit drugs, or the percentage of participants who graduated from the DTC program.

As previously mentioned, it can be challenging for evaluators to measure the wide range of services that DTCs provide and the diverse impacts these programs may have on their participants. The primary function of a performance indicator is to reduce this complex information into a manageable and analyzable set of numerical indexes, such as ratios, sums or percentages. There is no one correct way to operationalize a performance indicator, and OAS member states are free to define and measure these variables in the manner they deem most appropriate and informative.

At the same time, however, substantial benefits will be gained by having different

countries collect a common dataset of core performance indicators. This will allow member states to compare the performance of their DTCs to that of programs in other countries. It will also be possible to aggregate performance information and examine the regional impact of DTCs in the Western Hemisphere.

Core Performance Indicators: As noted in a previous section, representatives from several OAS member states convened on two occasions to provide guidance on the content and structure of this evaluation manual. They agreed that it would be desirable for member states to collect a core dataset of performance indicators. Collection of the core performance indicators will be voluntary and each country will decide for itself whether to share its findings with other nations. It was decided that the following criteria should guide the selection of core performance indicators:

- The core dataset should contain no more than ten performance indicators to avoid placing an onerous burden on member states.
- The performance indicators must be easy and inexpensive to collect, and objectively and reliably measurable.
- Program-level performance indicators should reflect the essential components of a DTC, without which a program would not be considered a DTC. These include, at a minimum, court hearings, substance use treatment, and drug and alcohol testing.
- Participant-level performance indicators should reflect outcomes of primary interest to many stakeholders of DTC programs. These include, at a minimum, retention in substance use treatment, use of illicit drugs and alcohol, and criminal recidivism.
- The performance indicators must have been demonstrated in prior research to predict post-program outcomes in DTCs or differentiate effective from ineffective DTC programs.

Recommended Performance Indicators: It was agreed further that alternative lists of

recommended and discretionary performance indicators should also be provided. Countries interested in conducting more in-depth evaluations can review these lists for further ideas about possible ways to measure performance in their programs. Performance indicators are listed as recommended if they have shown promise for predicting outcomes in DTCs, but are difficult to measure or are partially redundant with other performance indicators.

Discretionary Performance Indicators: Finally, performance indicators are listed as discretionary if they are theoretically relevant to DTC programs or have been recommended by leading research organizations, but have not yet been the subject of substantial research.

Performance Benchmarks

Performance benchmarks refer to specific thresholds for success on performance indicators that DTCs should strive to achieve. For example, a performance indicator might reveal that participants in a DTC attended an average of two court hearings per month. But whether this is a desirable level of attendance remains an unanswered question. Perhaps four court hearings per month would be more effective, or one hearing per month would be more cost-effective.

When programs are new, and research has not yet identified which practices produce better outcomes, performance benchmarks are based on educated hypotheses or anecdotal observations. For example, when the earliest DTCs were developed in the late 1980s and early 1990s, evaluators had little basis for knowing how the programs should operate and what services they should provide. The evaluators made educated hypotheses about the best ways to structure and deliver DTC services, operationalized variables to measure those services, and examined the effects of the services on outcomes.

Over time, evaluators determined empirically which practices produced better outcomes in DTC programs. For example, researchers in Australia and the United States found that scheduling court hearings every two weeks during the first phase of a DTC

produced significantly better outcomes than scheduling court hearings less frequently (Carey et al., 2012; Jones, 2011; Festinger et al., 2002; Marlowe et al., 2006, 2007; Zweig et al., 2012). Based on that body of research, one benchmark for success in the United States and Australia is to schedule court hearings every two weeks during the first phase of a DTC program.

The same findings might not generalize to new cultures or new countries. Research might reveal that court hearings have a greater or lesser impact on outcomes in South American or Caribbean Nations than they do in the United States or Australia. This does not, however, entitle evaluators in South American or Caribbean countries to ignore prior research. Especially when findings have been replicated by different researchers in different countries, evaluators in new nations have a responsibility to take note of those findings and determine empirically whether the same lessons might apply in their own programs.

This is not to suggest that evaluators in OAS member states should limit their inquiries to replicating what was found previously in other countries. All evaluators are encouraged to ask new questions, develop new benchmarks, and add to the international body of knowledge on DTC programs.

The point here is that science proceeds, in part, through careful replication of prior studies and assessing the generalizability of previous findings. Evaluators have an obligation to take into consideration findings that have come before when planning their research agenda and conducting program evaluations. Therefore, if a benchmark for success has been identified for a given performance indicator in at least one country, that information is provided in this manual for evaluators in OAS member states to consider when examining the performance of their programs.

CALCULATING PERFORMANCE INDICATORS

This chapter provides detailed information about calculating performance indicators for DTC program evaluations. Table 1 summarizes the data elements that are required to calculate each performance indicator, the formulas for the calculations, and possible interpretations of the results. The last two columns of Table 1 indicate whether studies have validated those performance indicators and whether benchmarks for success have been identified in other countries. For ease of identification, core performance indicators are designated as such in **red font**, recommended performance indicators are designated in **blue font**, and discretionary performance indicators are designated in **black font**.

TABLE 1: Recommended Performance Indicators & Potential Performance Benchmarks for DTC Evaluations in the Americas

*Core performance indicators reflect the defining features of DTCs, are reasonably inexpensive to collect, and significantly predict long-term outcomes. Recommended performance indicators significantly predict outcomes but may be difficult to measure or redundant with other performance indicators. Discretionary performance indicators are theoretically important to DTCs but have been insufficiently studied.

**Performance benchmarks are derived from studies conducted in Australia, Canada, and/or the United States. The same performance benchmarks may not be valid in South American or Caribbean countries.

Variable	Data Elements	Performance Indicators*	Sub-analyses	Interpretation	Prediction of Post-Program Recidivism or Costs	Potential Performance Benchmarks**
Court supervision	# hearings scheduled (HrgS) # hearings attended (HrgA) # hearings cancelled or rescheduled (HrgC) Date stamped	***Dose of hearings* = HrgA** ***Court appearance rate* = HrgA (HrgS – HrgC)**	By month or phase of program (*density*)	Intensity of judicial supervision Compliance with court hearings	Moderately predictive	Every 2 weeks during first phase Every 4 weeks thereafter
Substance use treatment	# treatment sessions scheduled (TxS) # treatment sessions attended (TxA) # treatment sessions cancelled, excused or rescheduled (TxC) Date stamped	***Dose of treatment* = TxA** ***Treatment attendance rate* = TxA (TxS – TxC)**	By month or phase of program (*density*) By level of care (e.g., outpatient vs. residential) By modality (e.g., cognitive-behavioral vs. insight-oriented)	Intensity of treatment provided Compliance with treatment conditions	Moderately predictive	6 to 10 hours per week during first phase 100 to 200 hours total depending on the severity of the case

TABLE 1: Recommended Performance Indicators & Potential Performance Benchmarks for DTC Evaluations in the Americas

*Core performance indicators reflect the defining features of DTCs, are reasonably inexpensive to collect, and significantly predict long-term outcomes. Recommended performance indicators significantly predict outcomes but may be difficult to measure or redundant with other performance indicators. Discretionary performance indicators are theoretically important to DTCs but have been insufficiently studied.

**Performance benchmarks are derived from studies conducted in Australia, Canada, and/or the United States. The same performance benchmarks may not be valid in South American or Caribbean countries.

Variable	Data Elements	Performance Indicators*	Sub-analyses	Interpretation	Prediction of Post-Program Recidivism or Costs	Potential Performance Benchmarks**
Drug & alcohol testing **-and-** **Abstinence from drugs & alcohol**	# urine, saliva or breath analyses scheduled (UaS) # urine, saliva or breath samples provided (UaP) # samples negative for all illicit drugs and alcohol (UaN) # samples invalid or adulterated (Ual) # samples excused (UaE) Date stamped Alternative: # days on continuous drug surveillance; e.g., anklet, sweat patch (DayCDS) # days with no positive reading for illicit drugs or alcohol (DayCDSneg)	***Testing dose* = UaP** - or - **DayCDS** ***Testing compliance rate* = (UaP – Ual) (UaS – UaE)** ***Abstinence rate* = (UaN – Ual) (UaS – UaE)** - or - **DayCDSneg**	By month or phase of program (*density*) By substances detected	Intensity of drug & alcohol testing; Compliance with drug & alcohol testing; Abstinence from drugs & alcohol	Moderately predictive	Twice per week for urine or saliva testing Minimum of 90 days on continuous monitoring Minimum of 90 consecutive days of abstinence; ideally 180 consecutive days of abstinence
Graduation or retention	# participants entered program (Ent) # participants graduated (Grad) # neutral discharges (Neu) # still enrolled (Enr) Alternative: # days from entry to discharge or last in-person contact (DayED)	***Graduation rate* = Grad (Ent – Neu)** - or – ***Retention rate* = (Grad + Enr) (Ent – Neu)** ***Length of stay* = DayED**	Reasons for termination	Ability of DTC to retain participants long enough to meet rehabilitative goals	Highly predictive	Avg. graduation rate ~ 60% 18 – 24 months of enrollment in the DTC Minimum of 90 days in treatment, and ideally 9- 12 months in treatment

TABLE 1: Recommended Performance Indicators & Potential Performance Benchmarks for DTC Evaluations in the Americas

*Core performance indicators reflect the defining features of DTCs, are reasonably inexpensive to collect, and significantly predict long-term outcomes. Recommended performance indicators significantly predict outcomes but may be difficult to measure or redundant with other performance indicators. Discretionary performance indicators are theoretically important to DTCs but have been insufficiently studied.

**Performance benchmarks are derived from studies conducted in Australia, Canada, and/or the United States. The same performance benchmarks may not be valid in South American or Caribbean countries.

Variable	Data Elements	Performance Indicators*	Sub-analyses	Interpretation	Prediction of Post-Program Recidivism or Costs	Potential Performance Benchmarks**
Recidivism	# participants entered program (Ent) # neutral discharges (Neu) # arrested for any new offense (Arr) # convicted of any new offense (Conv) Date stamped Alternatives: # of new arrests per participant (NoArr) # of new convictions per participant (NoConv) # of days from entry to first new arrest (DayArr)	***Re-arrest rate* = Arr (Ent - Neu)** - and/or - ***Re-conviction rate* = Conv (Ent-Neu)** ***Time to re-arrest* = DayArr**	Level of offense (felony vs. mis-demeanor vs. summary) Type of offense (e.g., drug-related, theft, property, or violent) In-program vs. post-program recidivism	Resumption of criminal activity	Highly predictive	No information
Incarceration	# participants entered program (Ent) # neutral discharges (Neu) # incarcerated for any new offense or technical violation (Inc) # days incarcerated (DayInc)	***New incarceration rate* = Inc (Ent - Neu)** **Days incarcerated = DayInc**	In-program vs. post-program incarceration Prison vs. jail	Extent of incarceration	Highly predictive	No information
Timeliness of services	# days from arrest or violation to entry (DayAE) # days from entry to first treatment contact (DayET)	***Intake efficiency* = DayAE** ***Treatment entry efficiency* = DayET**		Ability of program to reduce delays in case resolution and capitalize on motivation from the arrest event	Moderately predictive	Maximum of 50 days from arrest to entry; ideally within 20 days

TABLE 1: Recommended Performance Indicators & Potential Performance Benchmarks for DTC Evaluations in the Americas

*Core performance indicators reflect the defining features of DTCs, are reasonably inexpensive to collect, and significantly predict long-term outcomes. Recommended performance indicators significantly predict outcomes but may be difficult to measure or redundant with other performance indicators. Discretionary performance indicators are theoretically important to DTCs but have been insufficiently studied.

**Performance benchmarks are derived from studies conducted in Australia, Canada, and/or the United States. The same performance benchmarks may not be valid in South American or Caribbean countries.

Variable	Data Elements	Performance Indicators*	Sub-analyses	Interpretation	Prediction of Post-Program Recidivism or Costs	Potential Performance Benchmarks**
Probation supervision	# probation sessions scheduled (ProbS) # probation sessions attended (ProbA) # probation sessions cancelled or re-scheduled (ProbC) # probation field visits; e.g., home, work, school (ProbFV) Date stamped	***Probation dose* = ProbA + ProbFV** ***Probation attendance rate* = ProbA (ProbS – ProbC)**	By month or phase of program (*density*) Field vs. office visits	Intensity of probation supervision	Insufficiently studied	No information
Rewards & sanctions	# rewards administered per participant (Rew) # sanctions imposed per participant (Sanc) # infractions per participant (Infr) # achievements per participant (Ach)	***Balance of reinforcement* = Rew Sanc** ***Certainty of rewards* = Rew Ach** ***Certainty of sanctions* = Sanc Infr**	By magnitude of rewards or severity of sanctions	Consequences imposed for participants' conduct	Inconsistently predictive	4 : 1 ratio of rewards to sanctions
Restorative justice interventions	# hours community service performed Payment of court fines & fees (none, partial, complete) Restitution to victims (none, partial, complete) Satisfaction of other financial obligations; e.g., child support, alimony (none, partial, complete) # hours in victim impact panels	Not defined in the literature		Degree to which the program applies restorative justice principles and interventions	Inconsistently predictive	No information

TABLE 1: Recommended Performance Indicators & Potential Performance Benchmarks for DTC Evaluations in the Americas

*Core performance indicators reflect the defining features of DTCs, are reasonably inexpensive to collect, and significantly predict long-term outcomes. Recommended performance indicators significantly predict outcomes but may be difficult to measure or redundant with other performance indicators. Discretionary performance indicators are theoretically important to DTCs but have been insufficiently studied.

**Performance benchmarks are derived from studies conducted in Australia, Canada, and/or the United States. The same performance benchmarks may not be valid in South American or Caribbean countries.

Variable	Data Elements	Performance Indicators*	Sub-analyses	Interpretation	Prediction of Post-Program Recidivism or Costs	Potential Performance Benchmarks**
Employment	# participants entered program (Ent) # neutral discharges (Neu) # employed at entry (EmpE) # employed at discharge (EmpD)	***Discharge employment rate* = EmpD (Ent – Neu)** ***Employment improvement rate* = (EmpD – EmpE) (Ent – Neu)**	Full time vs. part time Excluding temporary, seasonal or off the books employment Type of work (e.g., clerical, manual labor, supervisory)	Degree to which the program improves the employment of participants	Highly predictive	DTC offers vocational services as needed Employment (or educational enrollment) is a condition of graduation
Education	# participants entered program (Ent) # neutral discharges (Neu) # with high school diploma or equivalency at entry (DipE) # with high school diploma or equivalency at discharge (DipD) # enrolled in educational program at entry (EdE) # enrolled in educational program at discharge (EdD)	***Educational improvement rate* = (DipD + EdD) – (DipE + EdE) (Ent – Neu)**		Degree to which the program improves the education of participants	Inconsistently predictive	Educational degree or enrollment (if unemployed) is a condition of graduation
Housing	# participants entered program (Ent) # neutral discharges (Neu) # in stable housing at entry (HouE) # in stable housing at discharge (HouD)	***Discharge housing rate* = HouD (Ent – Neu)** ***Housing improvement rate* = (HouD – HouE) (Ent – Neu)**	By type of housing; e.g., living with family or friends, recovery house	Degree to which the program improves the housing stability of participants	Highly predictive	Sober housing is a condition of graduation

TABLE 1: Recommended Performance Indicators & Potential Performance Benchmarks for DTC Evaluations in the Americas

*Core performance indicators reflect the defining features of DTCs, are reasonably inexpensive to collect, and significantly predict long-term outcomes. Recommended performance indicators significantly predict outcomes but may be difficult to measure or redundant with other performance indicators. Discretionary performance indicators are theoretically important to DTCs but have been insufficiently studied.

**Performance benchmarks are derived from studies conducted in Australia, Canada, and/or the United States. The same performance benchmarks may not be valid in South American or Caribbean countries.

Variable	Data Elements	Performance Indicators*	Sub-analyses	Interpretation	Prediction of Post-Program Recidivism or Costs	Potential Performance Benchmarks**
Emotional Health	# participants entered program (Ent) # neutral discharges (Neu) # reported serious emotional problems at entry (EmotE) # reported serious emotional problems at discharge (EmotD)	***Discharge emotional problems* = EmotD (Ent – Neu)** ***Emotional improvement rate* = (EmotE – EmotD) (Ent – Neu)**	By type of symptoms; e.g., depression, hostility, paranoia	Degree to which the program improves the emotional health of participants	Moderately predictive	DTC offers mental health counseling as needed Emotional problems have substantially improved prior to discharge
Medical and dental Health	# participants entered program (Ent) # neutral discharges (Neu) # reported serious medical or dental problems at entry (MedE) # reported serious medical or dental problems at discharge (MedD)	***Discharge medical or dental problems* = MedD (Ent – Neu)** ***Medical and dental improvement rate* = (MedE – MedD) (Ent – Neu)**		Degree to which the program improves the medical and dental health of participants	Insufficiently studied	DTC offers medical and dental services as needed
Family Relationships	# participants entered program (Ent) # neutral discharges (Neu) # reported serious family problems at entry (FamE) # reported serious family problems at discharge (FamD)	***Discharge family problems* = FamD (Ent – Neu)** ***Family improvement rate* = (FamE – FamD) (Ent – Neu)**		Degree to which the program reduces the family problems of participants	Insufficiently studied	DTC offers family counseling as needed Family problems have substantially improved prior to discharge
Social Relationships	# participants entered program (Ent) # neutral discharges (Neu) # reported serious social problems at entry (SocE) # reported serious social problems at discharge (SocD)	***Discharge social problems* = SocD (Ent – Neu)** ***Social improvement rate* = (SocE – SocD) (Ent– Neu)**		Degree to which the program reduces the social or interpersonal conflicts of participants	Insufficiently studied	No information
Birth of Drug-free babies	# drug- and alcohol-free babies delivered or fathered by DTC participants (Bab) Date stamped	*Drug-free babies* = Bab	In-program vs. post-program births	Degree to which the program contributes to the health of newborns	Insufficiently studied	No information

Court Supervision

In many ways, court supervision is the defining ingredient of a DTC. Many correctional rehabilitation programs provide substance use treatment, probation supervision, or drug and alcohol testing for drug-addicted persons charged with crimes; however, only DTCs are supervised by a judge and require participants to appear frequently in court for status review hearings. Numerous studies in Australia and the United States have found that outcomes in DTCs were significantly influenced by how often participants appeared before the judge in court (Carey et al., 2012; Festinger et al., 2002; Jones, 2011; Marlowe et al., 2006, 2007; Mitchell et al., 2012; Zweig et al., 2012).

Court supervision is commonly measured in terms of the "dosage" or number of court hearings that participants attended. Dosage is generally considered to be a program-level performance indicator because it indicates how frequently the judge supervised participants from the bench. It is proposed as a core performance indicator (indicated by **red font** below and in Table 1 because it: (1) reflects a defining ingredient of DTC programs, (2) is simple and inexpensive to collect, (3) is objectively and reliably measured and (4) has been found consistently to predict outcomes in DTC evaluations.

Dose of hearings = # of hearings attended

Evaluators may also wish to examine participants' appearance rate at scheduled court hearings. The appearance rate is a participant-level performance indicator because it indicates the degree to which participants were compliant with their requirements to attend court hearings. This variable is slightly more complicated to measure than dosage because it requires the evaluator to examine both the number of hearings that were attended and the number of hearings that were originally scheduled. It is calculated by dividing the number of hearings that were attended by the number that were scheduled, minus any hearings that may have been cancelled or rescheduled in advance or with the court's approval.

$$\textit{Court appearance rate} = \frac{\textit{\# of hearings attended}}{\textit{\# of hearings scheduled} - \textit{\# of hearings cancelled or rescheduled}}$$

In many studies, court appearance rates have not been significantly more predictive of outcomes than the dosage of court hearings. Perhaps because the repercussions for missing court sessions can be severe and may include incarceration, appearance rates have typically exceeded 85 percent. For mathematical reasons, variables tend to be less predictive when there is a restricted range of scores on those variables. Therefore, the court appearance rate is listed as a discretionary performance indicator (represented by black font in the formula above and in Table 1). OAS member states might find this performance indicator to be more predictive of outcomes if there is greater variability on the measure in their DTCs.

Finally, by date-stamping the information related to participants' attendance at court hearings, evaluators can measure the density of court hearings that were scheduled or attended during each month or phase of the program. In terms of performance benchmarks, evidence from the United States suggests court hearings should be held every two weeks during the first phase of the program, but may subsequently be reduced to monthly for participants who are compliant with their treatment obligations (Carey et al., 2012; Marlowe et al., 2007). If OAS member states wish to similarly determine when it is safe and effective to reduce the frequency of court hearings in their programs, it will be necessary to date-stamp the attendance information at court hearings.

Substance Use Treatment

Substance use treatment is another defining ingredient of a DTC. The basic assumption of the DTC model is that addiction is causing or exacerbating participants' criminal activity. Therefore, substance use treatment is believed to be essential to achieve long-term desistance from crime. In fact, if an individual can desist from crime without

substance use treatment, he or she is generally not considered to be a suitable candidate for a traditional DTC program (Marlowe, 2012b).

Similar to measuring court supervision, the provision of substance use treatment is commonly measured in terms of dosage (i.e., the number of sessions that were attended) or participants' attendance rate at scheduled treatment sessions. The dosage of treatment is considered a program-level performance indicator because it reflects the amount of treatment services that were delivered. The attendance rate is a participant-level performance indicator because it reflects the degree to which participants complied with their treatment obligations. Finally, by date-stamping treatment information, evaluators can measure the density of treatment that was scheduled or delivered during each month or phase of the program.

The dosage of substance use treatment is proposed here as a core performance indicator because it is a defining ingredient of DTC programs, is simple and inexpensive to collect, can be objectively and reliably measured, and significantly predicts outcomes in DTC programs. Studies in DTCs have consistently found that the longer participants remained in substance use treatment and the more sessions they attended, the better their outcomes (Banks & Gottfredson, 2003; Gottfredson et al., 2007; Gottfredson et al., 2008; Peters et al., 2002; Shaffer, 2010; Taxman & Bouffard, 2005).

Dose of treatment = # of treatment sessions attended

Participants' attendance rate in treatment is listed as a **recommended** performance indicator for OAS member states (reflected by **blue font** below and in Table 1). Similar to treatment dosage, this indicator significantly predicts post-program outcomes. However, the attendance rate is a bit more difficult to measure than dosage because it requires the evaluator to examine both the number of sessions that were attended and the number of sessions that were originally scheduled. In addition, because the dosage of treatment and the attendance rate are partially redundant, it is unclear whether the attendance rate

predicts outcomes significantly better than dosage alone. Evaluators in OAS member states might find that the dosage of treatment is adequate to characterize the provision of treatment and predict post-program outcomes in their DTC programs.

The treatment attendance rate is calculated by dividing the number of treatment sessions that were attended by the number originally scheduled, minus any sessions that were cancelled, excused or rescheduled in advance or with the approval of treatment staff.

$$\textit{Treatment attendance rate} = \frac{\textit{\# of sessions attended}}{\textit{\# of sessions scheduled- \# of sessions cancelled or rescheduled}}$$

In terms of performance benchmarks, studies in Canada and the United States have reported that the best outcomes were achieved when participants attended approximately six to ten hours of treatment per week during the first phase of the program, for a total of 100 to 200 hours of treatment over nine to twelve months (Bourgon & Armstrong, 2005; Huebner & Cobbina, 2007; Landenberger & Lipsey, 2005; Peters et al., 2002; Sperber et al., 2013). The studies found that 100 hours of treatment was generally sufficient for the less seriously addicted or antisocial participants, whereas 200 hours was required for high-risk participants who had more serious addictions or criminal histories. It is unknown whether these same performance benchmarks will be applicable in South American or Caribbean countries. Local evaluators in OAS member states are encouraged to examine the optimum dosage of substance use treatment in their own DTC programs.

It is also important to categorize substance use treatment by the modality or level of care. For example, residential or inpatient treatment is usually considerably more expensive than outpatient treatment and may be valued differentially in cost evaluations. In addition, the time participants spend in residential or inpatient treatment is likely to reduce their *time at liberty* significantly. This means that participants are likely to have reduced

opportunities to engage in substance use, crime, and other behaviors of interest to evaluators while they are being treated in a residential setting. As will be discussed later, the evaluator may need to statistically adjust for the number of days participants were in residential treatment when conducting certain types of outcome analyses. Evaluators are strongly encouraged, therefore, to date-stamp the days on which participants enter and are subsequently discharged from residential and inpatient treatment programs.

It may also be useful to characterize the orientation or philosophy of treatment being delivered. Studies in Canada and the United States have determined that cognitive-behavioral therapies (CBT) produced significantly better outcomes among criminal justice populations than insight-oriented or psychodynamic therapies (Andrews & Bonta, 2010; Gendreau, 1996). A number of rating tools have been developed to measure the quality and orientation of treatment and other services delivered in correctional rehabilitation programs. These include the Correctional Program Assessment Inventory (CPAI) which was developed in Canada, and the Correctional Program Checklist (CPC) developed in the United States. Higher ratings of treatment quality on these instruments have been shown to predict higher graduation rates and lower recidivism rates among DTC participants (Gutierrez & Bourgon, 2012) and participants in other correctional treatment programs (Lowenkamp et al., 2006, 2010). Importantly, raters must be trained carefully on how to administer these instruments. Information on training curricula for these instruments is provided in the Appendix.

Finally, the quality of the relationship between the participant and counselor has also been shown to significantly influence outcomes in substance use treatment. As will be discussed later, several instruments have been developed to assess participants' satisfaction with substance use treatment and the quality of the therapeutic alliance between the participant and treatment staff. Evaluators may wish to consider using these tools, which are listed in the Appendix, in their DTC program evaluations.

Drug and Alcohol Testing

Drug and alcohol testing is another key component of DTC programs. Drug and alcohol test results often serve as the principal basis for assessing participants' response to treatment, adjusting the conditions of treatment where indicated, and administering rewards for abstinence and sanctions for continued substance use. Participants in DTCs commonly attribute their success, in part, to the frequency of drug and alcohol testing (Goldkamp et al., 2002; Kleinpeter et al., 2010; Wolfer, 2006).

Some drug tests, such as urine or saliva tests, must be administered repeatedly because they can only detect drug or alcohol use over short periods of approximately two to four days. For these types of tests, the core program-level performance indicator is the dosage or number of tests that were administered. Other tests, such as sweat patches or ankle monitors, can provide continuous surveillance of drug or alcohol use over extended periods of several weeks. For these tests, the core performance indicator is the number of days participants were subjected to continuous surveillance.

Testing dose = # of urine, breath, or saliva tests administered

- and/or –

Testing dose = # of days on continuous surveillance

Evaluators may wish to examine participants' compliance rate with scheduled drug tests. The compliance rate is calculated by dividing the number of tests that were provided, minus any tests that were invalid or adulterated, by the number of tests that were scheduled minus tests that were excused in advance or with the court's approval.

$$\textit{Testing compliance rate} = \frac{\textit{\# of drug tests provided-\# of drug tests invalid or adulterated}}{\textit{\# of drug tests scheduled-\# of drug tests excused}}$$

Few studies have examined the impact of compliance with drug testing on outcomes in DTCs. Compliance with drug testing is often highly correlated or redundant with the dosage

of drug testing (discussed above) and with the results of drug testing (discussed below); therefore, the compliance rate is unlikely to provide significant prediction beyond other performance indicators. For this reason, it is listed as a discretionary performance indicator.

Studies in the United States have revealed that to achieve the most effective outcomes, drug tests with short detection windows should be administered no less frequently than twice per week for at least the first several months of the program (Carey et al., 2012; Gibbs & Wakefield, 20142014; Kleinpeter et al., 2010) and continuous surveillance testing should be conducted for a minimum of ninety consecutive days (Flango & Cheesman, 2009). It remains to be determined whether similar performance benchmarks will emerge in other OAS member states.

Abstinence

Achieving abstinence from illicit[2] drugs and alcohol is one of the primary goals of a DTC. As discussed earlier, the DTC model assumes substance use is causing or exacerbating criminal activity. Therefore, abstinence from addictive and intoxicating substances is believed essential for reducing criminal recidivism.

For drug tests with short detection windows, the core participant-level performance indicator for abstinence is the percentage of tests that were negative for all illicit drugs and alcohol. Samples that are invalid or adulterated are *not* counted as drug-negative. For tests involving continuous surveillance, the core performance indicator is the number of days participants were on continuous surveillance without a positive test reading.

$$\textit{Abstinence rate} = \frac{\textit{\# of drug tests negative for all illicit substances} - \textit{\# of drug tests invalid or adulterated}}{\textit{\# of drug tests scheduled} - \textit{\# of drug tests excused}}$$

- and/or –

Abstinence = # of days with no positive reading on continuous monitoring

2 Illicit drugs include prescription medications that are used for a non-prescribed or non-medically-indicated purpose.

In terms of performance benchmarks, studies from the United States suggest ninety consecutive days of abstinence is a minimum threshold for achieving positive outcomes after graduation (Carey et al., 2012). Ideally, programs should strive for at least one hundred and eighty days of consecutive abstinence prior to graduation in order to increase the odds of stable, long-term sobriety (McLellan et al., 2000).

The failure to provide a drug or alcohol test has a nuanced significance that merits particular attention. Because participants in DTCs are sometimes rewarded for providing negative drug tests and sanctioned for providing positive drug tests, they are considerably more likely to provide a specimen if they expect it to be drug-negative than if they expect it to be drug-positive.

For this reason, the failure to provide a specimen cannot be assumed to be a random or ignorable event, but rather is likely to reflect an effort to conceal substance use. In fact, many DTCs view the failure to provide a valid specimen as a separate infraction from substance use, and one that may receive a more severe sanction than providing a drug-positive specimen, which often receives a light sanction or adjustment to the participant's treatment plan (Marlowe, 2008, 2011). This is because participants who fail to provide a specimen are reasonably assumed both to have engaged in substance use and to have concealed the usage. Thus, the DTC might give participant two responses or a more severe sanction for this compound issue.

The generally recommended course of action is to assume that missing specimens are drug-positive. Of course, if a participant was excused by staff from providing a specimen for an acceptable reason (e.g., because of illness or an employment obligation), then it is appropriate to treat the missing specimen as missing data. Evaluators may elect to examine the data both ways, first by treating missing specimens as drug-positive and secondly by treating them as missing data. However, if the results are significantly discrepant in the two analyses, the former results are more likely to reflect the true levels of drug use.

There are advanced statistical imputation procedures that can sometimes be used to

compensate for missing drug-test data. Some imputation procedures examine the pattern of drug test results immediately before and after a missing specimen. For example, if a participant had several negative specimens just before and just after a missed specimen, the missing specimen might be presumed to be drug-negative. Alternatively, some procedures may impute the average or most prevalent result for the participant or for the population as the most likely result in place of a missing specimen. Selecting an appropriate imputation procedure is complicated and expert statistical consultation is usually required to apply these procedures correctly.

Retention and Graduation

Retention in substance use treatment is one of the most significant predictors of long-term outcomes in DTC programs. The longer participants remain in treatment, the better their outcomes (Gottfredson et al., 2008; Peters et al., 2002; Taxman & Bouffard, 2005). Similarly, by far the best outcomes are achieved by those who successfully complete or graduate from a DTC program (Carey et al., 2012; Gottfredson et al., 2007; Mitchell et al., 2012; U.S. Government Accountability Office, 2005).

Most DTCs have similar requirements for graduation and hold formal graduation ceremonies that mark successful completion of the program. DTCs that do not have formal criteria for graduation or graduation ceremonies have significantly poorer outcomes than those that do (Carey et al., 2012). Graduates are typically required to complete a substance use treatment program, maintain abstinence from all illicit drugs and alcohol for at least ninety to one hundred and eighty days, find employment or enroll in an educational program, be housed in sober living quarters, pay applicable fines and fees, and avoid new criminal charges or serious technical violations. Participants who can achieve these goals are far more likely than others to remain drug-free and crime-free after graduation.

Some DTCs may assign a *neutral discharge* to participants who were withdrawn from

the program for reasons beyond the control of the participant and the program. A neutral discharge is assigned most commonly to participants who are withdrawn from the program because they moved out of the jurisdiction (with the court's permission), enlisted in the military, or became too medically or psychiatrically unstable to provide continuing competent consent to participation. A neutral discharge may also be assigned if it is discovered after the fact that the participant had been admitted to the program erroneously; for example, if the participant had a prior conviction that precluded eligibility or resided in a judicial district that was not within the jurisdictional boundaries of the DTC. A neutral discharge should never be applied to cases in which termination was related in any way to a participant's performance in the DTC.

Participants who receive a neutral discharge are removed from the denominator when calculating the graduation rate. The formula is therefore the number of participants who graduated from the program, divided by the number who entered the program minus the number who received a neutral discharge. Participants who are still enrolled in the DTC are not included in this calculation. In terms of performance benchmarks, the average graduation rate is approximately sixty percent in the United States, with most DTCs graduating between fifty and seventy-five percent of their participants (Marlowe et al., 2016).

$$\textit{Abstinence rate} = \frac{\textit{\# of participants who graduated}}{\textit{\# of participants who entered DTC program = \# of neutral discharges}}$$

As will be discussed later, evaluators sometimes measure outcomes based on cohorts of DTC participants. A *cohort* is defined as a group of individuals who entered the DTC during the same specified time period, usually an interval of twelve months (Heck, 2006; Rubio et al., 2008). For example, all participants who entered the DTC between January 1 and December 31 of a given year might be defined as a cohort. Outcome analyses may then be conducted separately for cohorts defined, for example as having entered the program between January 1 and December 31, 2012, between January 1 and December 31, 2013,

and so forth. This provides a series of "snapshots" indicating how well the DTC performed over successive years of operation.

When evaluating cohorts, it is often necessary to examine participants who are still enrolled in the DTC. In this case, the retention rate is calculated by dividing the number of participants who graduated from the program plus the number that are still enrolled, by the number who entered the program minus the neutral discharges:

$$\textit{Abstinence rate} = \frac{\textit{\# of participants graduated + \# of participants still enrolled}}{\textit{\# of participants that entered program - \# of neutral discharges}}$$

An alternative way to measure retention is the *length of stay* in the program. This is defined as the number of days from a participant's entry into the DTC to his or her discharge or last in-person contact with DTC staff. The length of stay is listed as a recommended performance indicator because it is highly predictive of outcomes but is also likely to be redundant with the retention rate or graduation rate. In terms of performance benchmarks, studies in the United States suggest DTCs have better outcomes when the planned length of stay for the program is between approximately eighteen and twenty-four months (Carey et al., 2012; Shaffer, 2010).

$$\textit{Length of stay = date of discharge or last personal contact with staff - date of entry}$$

It is not always obvious how to define the entry date or discharge date for a DTC participant. Most DTCs have a formal entry hearing at which participants enter a plea, voluntarily waive certain legal rights, and are officially enrolled in the program. The date of this hearing is typically counted as the entry date because it marks the specific point in time when the program gained authority over the individual to impose conditions of treatment and supervision. Some participants may stop attending treatment for several weeks or months before being formally discharged from a DTC. Under such circumstances, the last in-person contact with staff is the best indicator of when the participant left the program.

Recidivism

For many policymakers and members of the public, reducing criminal recidivism is the primary goal of a DTC. Recidivism is defined here as any return to criminal activity after the participant entered the DTC program. If a participant is arrested for or charged with a crime for alleged activity that occurred before he or she entered the DTC, this is *not* counted as recidivism because the event occurred prior to entry.

Recidivism is typically measured by new (1) arrests, (2) convictions, (3) incarcerations or (4) self-reported criminal activity. Each measure has its own advantages and disadvantages that the evaluator will need to take into consideration. Incarceration is discussed separately below because it has distinct public safety and cost implications.

There are several potential benefits to analyzing new arrests as a measure of criminal recidivism. First, arrests are usually substantially closer in time to the alleged criminal activity than convictions. In some countries, it may take months or years to conduct plea negotiations, hold a criminal trial, and determine guilt or innocence. Evaluators can usually report arrest outcomes in a much shorter time than waiting for lengthy legal proceedings to be resolved. Second, criminal cases may be 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 severity of the offense that actually occurred.

On the other hand, individuals may be arrested for crimes they did not in fact commit. This measure could seriously overestimate the level of criminal recidivism. Relying on conviction data rather than arrest data may provide greater assurances that the crimes did in fact occur.

Evaluators must also consider the timeliness and accuracy of information that is available in criminal justice databases. In some countries such as the United States, arrest data may be entered in a more timely and faithful manner than conviction data. In other

countries such Canada or Jamaica, conviction data may be equally or more accurate than arrest data. Evaluators must familiarize themselves with how and when information is entered into national, state, and local administrative databases.

Self-reported information can potentially provide the most accurate assessment of criminal recidivism because it does not require detection or prosecution by law enforcement. Because most crimes go unreported by victims and undetected by the authorities, arrest and conviction data often underestimate true levels of criminal activity. For obvious reasons, however, individuals cannot be relied upon to acknowledge crimes unless they receive strict assurances that the information will be kept confidential and will not be used against them in a criminal proceeding. This ordinarily requires a DTC to hire an independent interviewer who has no connection to the court or criminal justice system to confidentially survey the participants. Few DTCs are likely to have adequate resources to hire such independent interviewers.

Most DTC evaluations in the United States have relied on arrests as the primary measure of recidivism, whereas studies in Canada and Australia have relied primarily on convictions. With the notable exception of one well-funded national study in the United States (Rossman et al., 2011), few DTC evaluations have examined self-report data as the measure of recidivism.

Arrest rates and conviction rates are calculated by dividing the number of participants who were arrested or convicted of a new offense by the number who entered the program minus neutral discharges. Technical violations, defined as violations of a court order that do not constitute a crime *per se*, are usually counted separately from new offenses. For example, drinking alcohol is legal for most adults in most countries, but may be a technical violation for a DTC participant and could lead to a new arrest or conviction for a probation violation. Sanctions for technical violations are often counted as investment costs for the DTC program, whereas sentences or dispositions for new offenses are often counted as outcome costs.

$$\text{Re-arrest rate} = \frac{\text{\# arrested}}{\text{\# entered program - \# of neutral discharges}}$$

- and/or –

$$\text{Re-arrest rate} = \frac{\text{\# convicted}}{\text{\# entered program - \# of neutral discharges}}$$

Evaluators may also wish to examine recidivism in terms of the average number of new arrests or new convictions per participant. This is usually not very informative within the first twelve to twenty-four months of a participant's entry. It usually takes a substantial time for participants to re-engage in criminal conduct, be detected by law enforcement, and have criminal charges filed against them. For participants who return quickly to criminal behavior, most will have only one or two new arrests within the first two years. As will be discussed later, this can create what is called a *skewed* or *non-normal distribution* in the data. For mathematical reasons, one is unlikely to detect statistically significant differences if the distribution is skewed. Therefore, recidivism during the first two years is usually most informative when analyzed in terms of percentages of participants who had at least one new arrest or conviction.

Date-stamping arrest and conviction data permits the evaluator to analyze in-program versus post-program recidivism separately, and to calculate the average length of time until the first new arrest or conviction has occurred. Using a statistical technique called a *survival analysis*, the evaluator can determine the degree to which the DTC significantly delayed the onset of new criminal activity. The time to re-arrest is listed as a recommended performance indicator because it has important implications for public safety and cost but is also often highly correlated or redundant with new arrest and conviction rates.

It is usually advisable to categorize new offenses by the level or severity of the offense (e.g., felony vs. misdemeanor vs. summary offenses) and by the type of crime that was committed (e.g., drug offenses, property or theft offenses, violent offenses,

technical violations, prostitution, and traffic offenses). Different categories of offenses often 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 offenses may not directly involve victims and are more likely to receive a less costly probation sentence.

As will be discussed later, most studies have examined recidivism for a period of between two and five years after entry into the DTC. Because many DTC programs are approximately twelve to eighteen months in length, measuring outcomes after two years allows adequate time to elapse for most participants to have completed or been discharged from the program. After three to five years, recidivism rates for persons charged with drug-related crimes are likely to reach a plateau (Gossop et al., 2005; Inciardi et al., 2004; Martin et al., 1999). This means that most participants who will recidivate are likely to have done so by that time. Therefore, estimates of recidivism are likely to be reasonably stable after a period of approximately three to five years.

Incarceration

Incarceration typically accounts for the greatest cost to society from persons charged with crimes related to substance use disorders. Depending on the country and the conditions of supervision, a day in jail or prison may cost between five and twenty times more than a day on probation or in community-based substance use treatment (Belenko et al., 2005; Zarkin et al., 2012).

New incarceration rates are calculated in a similar manner to new arrest rates and new conviction rates. The number of participants who were incarcerated for a new offense is divided by the number who entered the program minus the neutral discharges:

$$\textit{New incarceration rate} = \frac{\textit{\# incercarated}}{\textit{\# entered program-\# of neutral discharges}}$$

New incarceration rates are usually not very informative, however, because they are redundant with new arrest and conviction rates. In most instances, a participant will not be incarcerated unless he or she was first arrested for and convicted of a new offense or for repeatedly violating the terms of probation. Therefore, new incarceration rates generally do not provide much useful information that new arrest and conviction rates do not.

Nonetheless, the length of time that participants were incarcerated is far more informative than the incarceration rate, especially when evaluating the cost impacts of a DTC. The core performance indicator is the total number of days that participants were incarcerated. Date-stamping is therefore critical for examining the length of incarceration. This requires the evaluator to record the dates that participants entered and were subsequently released from custody. If a participant was incarcerated on more than one occasion, the days for each episode are summed to determine the total length of incarceration.

Days of incarceration = date of release from custody = date of entry into custody

Evaluators typically distinguish between incarceration that occurred while participants were enrolled in the DTC, and incarceration that occurred after discharge from the DTC. In-program incarceration often reflects brief jail sanctions that may be imposed by the judge as a sanction for misconduct in the program, whereas post-program incarceration typically reflects pre-trial detention or criminal sentences for new charges. In cost evaluations, in-program jail sanctions are usually counted as investment costs for the DTC, whereas post-program sentences or detention are counted as outcome costs.

In some countries, there may be an important distinction between jails and prisons. In those national contexts, jails are typically used for pre-trial detention or short sentences of less than one year, whereas prisons are typically reserved for sentences of longer than one year. Jails are often located in and administered by the same county or precinct where the offender resides or where the offense was committed, whereas prisons may be located a

distance away and may be administered by state, provincial, or federal government agencies. From a fiscal standpoint, a day in prison may be substantially more expensive than a day in jail, and may thus be valued differently in a cost analysis.

Timeliness of Services

When the earliest DTCs were developed in the late 1980s and early 1990s, two of their goals were to clear clogged caseloads in the courts and resolve drug-related cases more efficiently. Having a specialized team with representatives from the defense and prosecution was intended, in part, to achieve a speedier resolution of factual controversies and impose rehabilitative dispositions more quickly.

Like many other performance indicators mentioned in this manual, measuring the timeliness of services in a DTC requires careful attention to date-stamping. One measure, *intake efficiency*, is defined as the number of days from arrest to entry into the DTC program. Some DTCs serve individuals who are charged with violations of probation, in which case the date of the probation violation is counted in lieu of an arrest date. Studies in the United States have found that outcomes were significantly better for DTCs that reduced the delay from arrest to entry to less than fifty days, and ideally to less than twenty days (Carey et al., 2012). Evaluators from other OAS member states will need to establish their own performance benchmarks for their programs, but it is highly likely that speedier entry into the DTC program will produce significantly better outcomes. Intake efficiency is therefore calculated as:

Intake efficiency = Date of entry into DTC-Date of arrest or probation violation

Evaluators may also wish to calculate the time delay from entry into the DTC to the first treatment session, called *treatment intake efficiency*:

In most studies to date, once participants entered the DTCs, the programs were usually able to initiate treatment within only about one to two weeks. Therefore, the short time delay between entry and the first treatment session was not found to predict outcomes in

many instances. If other countries experience longer delays in initiating treatment for their DTC participants, this performance indicator might provide better prediction of outcomes. Because it has generally not been found to be predictive of outcomes thus far, it is listed as a discretionary performance indicator for evaluators to consider, and is calculated as follows:

Treatment intake efficiency = Date of first treatment session-Date of entry into DTC

Probation Supervision

DTCs combine substance use treatment with strict behavioral accountability. In some countries such as the United States, Canada, and the United Kingdom, probation officers or other community supervision officers play a critical role in supervising participant performance in the program. In those countries, the probation officer may monitor participants' compliance with treatment, conduct home or employment visits, perform drug and alcohol testing, and report compliance information to the judge. If a participant commits a serious infraction or absconds from the program, the probation officer may be empowered to take the individual into custody and bring him or her before the court for a response.

Other countries, such as Panama, do not have probation officers; rather, the judge and psychosocial team are primarily responsible for monitoring compliance in the program. For countries that do not have probation officers, performance indicators for probation supervision may not be a relevant consideration for DTC evaluations.

Few studies have examined the contributions of probation to DTC outcomes, and no performance benchmarks have yet been identified for probation supervision in DTC programs. For this reason, the performance indicators for probation supervision are listed as discretionary. It is hoped that evaluators in OAS member states will focus greater attention on this important issue and contribute to the international body of research on best practices for probation officers, where relevant, in DTC programs.

Performance indicators for probation supervision are measured in a similar manner to

those for court hearings and substance use treatment. The dosage of probation supervision is calculated by the number of in-office probation sessions that were attended plus the number of field visits that were conducted. A field visit might include inspecting a participant's home or verifying that the participant was present at his or her school or place of employment at the appropriate times. This variable is considered a program-level performance indicator because it reflects the amount of probation supervision that was provided to participants.

Dose of probation = # of probation sessions attended + # of field visits

The probation attendance rate represents the proportion of probation sessions that were attended, divided by the number of sessions that were scheduled minus any sessions that were cancelled or rescheduled in advance or with the permission of the probation officer. Field visits are not included in this calculation because they are typically unannounced, and participants usually do not choose whether to comply with them. This variable is a participant-level performance indicator because it reflects the degree to which participants were compliant with probation requirements. Also, by date-stamping probation information, the evaluator can measure the density of probation services that were scheduled or delivered during each month or phase of the program.

In many DTCs, probation supervision, like court hearings and other supervision requirements, are gradually reduced in intensity or frequency as participants make progress in treatment. By examining the density of supervision per month or phase of the program, evaluators can measure the degree to which probation supervision and other services were decreased in intensity over time. The probation attendance rate is calculated as follows:

$$\textit{Treatment intake efficiency} = \frac{\textit{\# of probation sessions attended}}{\textit{\# of sessions scheduled - \# of sessions cancelled or rescheduled}}$$

Rewards and Sanctions

Performance indicators for rewards and sanctions have met with limited success in DTC evaluations to date (Bonomo, 2012; Linquist et al., 2006). Most studies have reported the average numbers of rewards and sanctions that were administered, and perhaps categorized the rewards and sanctions by magnitude or severity.

This approach has not been informative for at least two reasons. First, the imposition of rewards and sanctions is highly correlated with other performance indicators, such as treatment attendance rates and drug test results. By design, participants in DTCs receive rewards for such behaviors as attending treatment sessions and providing drug-negative urine specimens and receive sanctions for missing treatment sessions and committing other violations. Therefore, the numbers of rewards and sanctions that are imposed are typically redundant with other performance indicators.

Second, behavioral research reveals that the critical factor for success is not how many rewards or sanctions were imposed, but rather whether they were applied with certainty and immediacy (Harrell & Roman, 2001; Hawkin & Kleiman, 2009). Certainty refers to the ratio of rewards to achievements, or the ratio of sanctions to infractions (Marlowe & Kirby, 1999). For example, if participants received a reward for every achievement in the program, or a sanction for every infraction, then the certainty ratio would be 1.0 (or 100 percent) and the program would be more likely to produce favorable outcomes.

$$\textit{Certainty of rewards} = \frac{\textit{\# of rewards administered}}{\textit{\# of achievements}}$$

$$\textit{Certainty of sanctions} = \frac{\textit{\# of sanctions imposed}}{\textit{\# of infractions committed}}$$

The certainty ratios for rewards and sanctions are listed as discretionary performance indicators because they have not been validated in DTC program evaluations, and because no performance benchmarks have been identified for use in DTCs or other criminal justice programs. In addition, they are likely to be considerably more difficult and expensive to measure than other performance indicators discussed thus far. Evaluators will need to keep track not only of the numbers of rewards and sanctions that are imposed, but also of the numbers of achievements and infractions that occurred. Because many different behaviors may be counted as achievements or infractions in a DTC, this process requires careful measurement of numerous variables. For many evaluators, this level of effort may be more complicated or costly than is feasible or desirable.

Effectiveness is also substantially influenced by the immediacy with which rewards and sanctions are imposed (Harrell & Roman, 2001; Hawkin & Kleiman, 2009; Marlowe & Kirby, 1999). The sooner rewards are administered after an achievement, and the sooner sanctions are imposed after an infraction, the better the effects are likely to be. Examining immediacy would require evaluators to measure the time delay between each achievement and the delivery of a reward, and between each infraction and the delivery of a sanction. This process would add date-stamping to the already complicated measurement that is required to examine certainty. Because this process is likely to be prohibitively costly and difficult for many evaluators, immediacy is not included in the list of performance indicators.

Some evidence suggests outcomes may be better for programs that provide a 4:1 ratio of rewards to sanctions for participants (Gendreau, 1996; Wodahl et al., 2011). Many DTC participants have long histories of receiving punishment for wrongdoings, but they are often unaccustomed to receiving positive reinforcement for behaving appropriately. Offering a greater proportion of rewards for productive behaviors than sanctions for infractions is hypothesized to increase the effectiveness of correctional rehabilitation programs.

Support for the 4:1 ratio must be viewed as preliminary, however, because it was

derived from *post hoc* (after the fact) correlations rather than controlled research studies. By design, sanctions are imposed for poor performance and rewards are provided for good performance. Therefore, a greater proportion of rewards over sanctions might not have caused better outcomes, but rather better outcomes might have elicited a greater proportion of rewards. For this reason, the balance of reinforcement (i.e., the ratio of rewards to sanctions) is listed as a discretionary performance indicator for evaluators to consider examining in their programs.

$$\textit{Balance of reinforcement} = \frac{\textit{\# of rewards administered}}{\textit{\# of sanctions imposed}}$$

Restorative Justice Interventions

Many DTCs include restorative justice interventions in their programs. Participants may be required, for example, to pay restitution to victims, participate in victim impact panels, or perform community service such as working in a soup kitchen or homeless shelter. Panama, for example, requires compensation or restitution for crime victims.

The most commonly reported performance indicators for restorative justice interventions include the number of hours of community service that were performed, and the degree to which participants satisfied their financial obligations (from no payment to partial payment to full payment). The financial obligations are typically categorized according to payment of fines and fees to the court, payment of restitution to victims, and payment of other personal financial obligations such as child support or alimony.

Although restorative justice interventions have been shown to improve outcomes in other criminal justice contexts (Latimer et al., 2005; Sullivan & Tifft, 2008), no study has examined their impact in DTCs and no performance benchmarks have been identified for their use with individuals charged with drug-related crimes. Evaluators from OAS member

states are invited to examine the role of restorative justice interventions in their DTC programs and share the information they learn with the global research community.

Employment

Outcomes in DTC programs are significantly better for participants who were employed at entry or who obtained employment during their enrollment in the program (Deschenes et al., 2009; Leukefeld et al., 2007). Better outcomes are also produced by DTCs that offer employment counseling and require participants to be employed or enrolled in an educational program as a condition of graduation from the program (Carey et al., 2012). A common complaint among DTC participants, especially racial and ethnic minority participants, is that the programs paid insufficient attention to their employment needs (Cresswell et al., 2001; Gallagher, 2012).

Employment outcomes are commonly reported in one of two ways in DTC evaluations. The *discharge employment rate* represents the proportion of participants who were employed at discharge, divided by the number who entered the program minus the neutral discharges. Being employed (or enrolled in an educational program) at discharge predicts significantly better post-program outcomes. (Being employed at entry also predicts better post-program outcomes; however, this variable is a participant characteristic or predictor variable rather than a performance indicator because it is not influenced by participants' enrollment in the DTC.)

$$\textbf{Discharge employment rate} = \frac{\textbf{\# employed at discharge}}{\textbf{\# entered program-\# of neutral discharges}}$$

Although employment at discharge is significantly predictive of outcomes, it does not indicate whether the DTC was responsible for improving participants' employment. High employment rates at discharge could simply indicate that the DTC targeted low-risk

individuals who were already more likely to have a job. This would say very little about how the DTC performed. For this reason, the discharge employment rate is listed as a discretionary performance indicator.

The *employment improvement rate* is calculated by subtracting the number of participants who were already employed at entry from those who were employed at discharge, and then dividing by the number who entered the program minus the neutral discharges. If more participants are employed at discharge than at entry, the value of this ratio will be positive and will represent the percentage of individuals who became employed while they were enrolled in the DTC program. This performance indicator is listed as recommended because it significantly predicts outcomes, but it often requires the use of self-report assessments of participants' employment status.

$$\textit{Employment improvement rate} = \frac{\textit{\# employed at discharge - \# employed at entry}}{\textit{\# entered program - \# of neutral discharges}}$$

Nonetheless, neither of the above performance indicators measures the *quality* of participants' employment. For example, they do not indicate whether some participants may have transitioned from part-time to full-time work, from seasonal to stable employment, and/or received a raise or promotion at their job.

Evaluators typically assess the nature or quality of participants' employment using self-report assessment tools. Examples of commonly used tools for this purpose include, but are not limited to, the Addiction Severity Index (ASI) and the Global Appraisal of Individual Needs (GAIN). The Appendix provides links to websites where the reader can obtain further information about these and other assessment tools. Items in these tools inquire about such issues as the number of days the participant was employed during the preceding month, the type of work that was performed (e.g., clerical, manual labor, or supervisory), and whether the employment was full-time, part-time, temporary, or seasonal. Although

these tools were developed in the United States, they have been translated into Spanish and other languages and have been validated in several countries, including South American countries.

Education

Many DTCs require or encourage participants to earn a high school diploma or equivalency degree if they do not have one already. Because it usually takes several months or years to accomplish this goal, most participants will have completed the DTC program before they could have earned an educational or equivalency degree. For this reason, it is usually not informative to report the percentage of participants who earned a degree by the time of discharge.

The performance indicator for the *educational improvement rate* is measured either as a change in participants' educational status or enrollment in an educational program. The difference between the number of participants who met either of these criteria at discharge and the number who met either criterion at entry is divided by the number who entered the program minus the neutral discharges. If more participants have an educational degree or are enrolled in an educational program at discharge than at entry, this ratio will be positive and will represent the percentage of individuals who improved or are in the process of improving their education.

$$\textbf{\textit{Educational improvement rate}} = \frac{(\#\text{ with diploma or equivalent at discharge} + \#\text{ enrolled in educational program at discharge}) - (\#\text{ with diploma or equivalent at entry} + \#\text{ enrolled in educational program at entry})}{\#\text{ entered program} - \#\text{ of neutral discharges}}$$

Studies are inconclusive concerning the influence of educational improvements on outcomes in DTC programs. Because support for the effects of educational efforts is unclear at present, educational improvement is listed as a discretionary performance indicator for evaluators to consider.

Housing

Significantly better outcomes are achieved by DTCs that require participants to be living in safe and drug-free housing as a condition of graduation from the program (Carey et al., 2012). Outcomes are also substantially better for DTCs that provide transitional housing when it is needed (McKee, 2010). Without safe and drug-free housing, it is highly unlikely that participants will maintain sobriety or desist from criminal activity after leaving the program. Because drug-free housing is strongly and consistently associated with successful outcomes, improvements in housing are listed as a core performance indicator.

The performance indicators for housing are calculated in a similar fashion to those for employment and education. The *discharge housing rate* represents the number of participants who were living in safe and sober housing at discharge, divided by the number who entered the program minus the neutral discharges.

$$\textbf{Discharge housing rate} = \frac{\textbf{\# in stable housing at discharge}}{\textbf{\# entered program-\# of neutral discharges}}$$

The *housing improvement rate* represents the number of participants who were living in safe and drug-free housing at discharge minus those who were already living in safe and sober housing at entry, divided by the number who entered the program minus the neutral discharges. This performance indicator reflects changes in housing that occurred while participants were enrolled in the DTC program. If more participants are living in stable housing at discharge than at entry, the value of this ratio will be positive and will represent the percentage of participants whose housing improved while they were enrolled in the DTC.

$$\textbf{Housing improvement rate} = \frac{\textbf{\# in stable housing at discharge - \# in stable housing at entry}}{\textbf{\# entered program - \# of neutral discharges}}$$

These performance indicators do not measure the quality of the housing arrangements. Evaluators are encouraged to categorize the quality of housing in terms of, for example, whether the participants own or rent their residence, are living with family members or friends, or are living in a recovery house or residential treatment facility. Instruments such as the ASI and GAIN (listed in the Appendix) evaluate the quality of participants' housing in the context of substance abuse treatment.

Emotional Health

Outcomes are significantly better for DTCs that offer psychiatric or mental health treatment for participants who need these services (Carey et al., 2012). Continued emotional problems at discharge predict significantly poorer post-program outcomes in DTC evaluations (Mendoza et al., 2013; Peters et al., 2012).

In most studies, evaluators have used self-report questionnaires or structured interviews to determine which participants needed mental health services and whether their symptoms improved during their enrollment in the DTC. Many of these instruments assess problems in multiple areas of participants' lives, including medical, psychiatric, family, and social problems. Therefore, the same tools can often be used to develop performance indicators for several of the domains discussed here and below. The Appendix lists several clinical assessment instruments that have been used to assess emotional problems and other psychosocial problems in DTC evaluations and have been translated into Spanish or validated in multiple countries.

Discharge emotional problems are calculated by dividing the number of participants who were experiencing significant emotional problems at discharge by the number who entered the program minus the neutral discharges. This ratio represents the proportion of participants who did not resolve their emotional problems prior to discharge, and therefore may be at increased risk for relapse to substance use or criminal recidivism.

$$\textbf{Discharge emotional problems} = \frac{\textbf{\# with emotional problems at discharge}}{\textbf{\# entered program - \# of neutral discharges}}$$

The *emotional improvement rate* indicates the degree to which the DTC was responsible for reducing participants' emotional problems. It is calculated by dividing (A) the number of participants who were experiencing significant emotional problems at entry minus those experiencing significant emotional problems at discharge, by (B) the number who entered the program minus the neutral discharges. If fewer participants are experiencing emotional problems at discharge than at entry, the value of the ratio will be positive and will represent the percentage of participants whose emotional problems resolved while they were in the DTC program. This performance indicator is listed as recommended because it significantly predicts outcomes but often requires the use of repeated self-report assessments of participants.

$$\textbf{Emotional improvement rate} = \frac{\textbf{\# with emotional problems at entry - \# with emotional problems at discharge}}{\textbf{\# entered program - \# of neutral discharges}}$$

A substantial proportion of DTC participants may leave the program prematurely and without warning to staff members. This can complicate efforts to assess participants' emotional status (or their status on other psychosocial variables) at the point of discharge. Counselors in DTC programs are therefore encouraged to periodically reassess participants' emotional health and status on other psychosocial variables. Many assessment tools, including the ASI, GAIN and others listed in the Appendix, can be used to reassess participants' functioning at intervals of every three to six months. Performing periodic reassessments helps to ensure that information about participants' emotional functioning will be available at or near the time of discharge and is also a critical component of good clinical practice. Effective clinicians periodically reevaluate their clients to determine whether treatment is working or whether adjustments may be needed to the treatment plan.

Medical and Dental Health

Preliminary evidence suggests outcomes may be better for DTCs that offer medical or dental treatment for participants who need these services (Carey et al., 2012). As with emotional problems, in most studies evaluators have relied on self-report questionnaires or interviews to determine participants' need for medical or dental services. No studies have examined the degree to which continued medical or dental problems at discharge predict long-term outcomes. Therefore, performance indicators related to medical and dental services are listed as discretionary variables for evaluators to consider.

The performance indicator for *discharge medical or dental problems* is calculated by dividing the number of participants who were experiencing significant medical or dental problems at discharge by the number who entered the program minus the neutral discharges. This ratio represents the proportion of participants who did not resolve their medical or dental problems prior to discharge, and therefore may be at risk for relapse to substance abuse or criminal recidivism in the future.

$$\textbf{Discharge medical or dental problems} = \frac{\textbf{\# with medical/dental problems at discharge}}{\textbf{\# entered program-\# of neutral discharges}}$$

The performance indicator for *medical or dental improvement* is calculated by dividing the number of participants who were experiencing medical or dental problems at entry minus those experiencing medical or dental problems at discharge, by the number who entered the program minus the neutral discharges. If fewer participants are experiencing medical or dental problems at discharge than at entry, the value of this ratio will be positive and will represent the percentage of participants whose medical or dental problems were resolved while they were enrolled in the DTC program.

$$\textbf{Medical/dental improvement} = \frac{\textbf{\# with medical/dental problems at entry - \# with medical/dental problems at discharge}}{\textbf{\# entered program - \# of neutral discharges}}$$

Family Relationships

Outcomes are significantly better in DTCs that offer family counseling for participants who need these services (Carey et al., 2012). A national study of twenty-three DTCs in the United States found that reducing family conflicts and improving family support were commonly derived benefits of DTC programs (Rossman et al., 2011). Evaluators typically administer self-report questionnaires or interviews such as the ASI or GAIN to determine whether participants need family-based services and to gauge whether family functioning has improved.

Discharge family problems are calculated by dividing the number of participants who were experiencing significant family problems at discharge by the number who entered the program minus the neutral discharges. This ratio represents the proportion of participants who did not resolve their family problems prior to discharge, and therefore may be at increased risk for relapse to substance abuse or criminal recidivism.

$$\textbf{Discharge family problems} = \frac{\textbf{\# with family problems at discharge}}{\textbf{\# entered program - \# of neutral discharges}}$$

The *family improvement rate* is calculated by dividing the number of participants who were experiencing family problems at entry minus those experiencing family problems at discharge, by the number who entered the program minus the neutral discharges. If fewer participants are experiencing family problems at discharge than at entry, the value of the ratio will be positive and will represent the percentage of participants whose family problems resolved while they were in the DTC program. This performance indicator is listed as recommended because it significantly predicts outcomes but requires the use of self-report assessments.

$$\textbf{Family improvement rate} = \frac{\textbf{\# with family problems at entry - \# with family problems at discharge}}{\textbf{\# entered program - \# of neutral discharges}}$$

As was noted previously, a substantial proportion of DTC participants may leave the program prematurely and without warning. Therefore, counselors are encouraged to periodically reassess participants' family functioning at intervals of approximately every three to six months. Performing periodic reassessments will ensure that information about participants' family functioning is available at or near discharge and is also a feature of good clinical practice.

Social Relationships

Social problems refer to serious or repeated interpersonal conflicts with persons other than family members, as well as social isolation or alienation from other people. Many DTC programs provide interventions designed to reduce interpersonal conflicts and increase participants' involvement in healthy social interactions. However, no studies have examined the impact of continued social or interpersonal problems on long-term outcomes from DTC programs. Therefore, performance indicators related to social problems are included as discretionary variables for evaluators to consider.

The performance indicator for *discharge social problems* is calculated by dividing the number of participants who were experiencing significant social problems at discharge, by the number who entered the program minus the neutral discharges. This ratio represents the proportion of participants who did not resolve their social problems prior to discharge, and therefore may be at increased risk for relapse to substance abuse or criminal recidivism in the future.

$$\textit{Discharge social problems} = \frac{\textit{\# with social problems at discharge}}{\textit{\# entered program - \# of neutral discharges}}$$

The performance indicator for the *social improvement rate* is the number of participants who were experiencing social problems at entry minus those experiencing social problems at discharge, divided by the number who entered the program minus the neutral discharges.

If fewer participants are experiencing social problems at discharge than at entry, the ratio will be positive and will represent the percentage of participants whose social problems resolved while they were in the program.

$$\textit{Social improvement rate} = \frac{\textit{\# with social problems at entry - \# with social problems at discharge}}{\textit{\# entered program - \# of neutral discharges}}$$

Birth of Drug-Free Babies

Consumption of alcohol or other drugs during pregnancy, especially cocaine or stimulants, is associated with serious birth complications and physiological and developmental deficits for the newborn (Cooper, 2004). The added costs to society of caring for a drug-exposed baby can be very high.

Because few DTC participants or their mates deliver babies during their enrollment in the program, it is difficult to measure statistically significant effects on this variable. Nevertheless, the sentiments of policymakers, members of the public, and other stakeholders are understandably swayed by this outcome measure. Saving one innocent newborn from a lifetime of misery might be worth the total costs of a DTC to many people. Because births are infrequent events in DTC programs, it is usually not informative to calculate discharge rates or improvement rates. Most evaluations simply report the number of drug-free newborns that were delivered to or fathered by DTC participants. For DTCs that have been in existence for many years, the results may be dozens of innocent lives saved.

$$\textit{Treatment intake efficiency} = \textit{Date of first treatment session} - \textit{Date of entry into DTC}$$

ATTITUDINAL VARIABLES

The performance indicators discussed thus far have focused on objective events or behaviors, such as drug test results or schedules of court hearings. It has also been hypothesized that DTCs may improve outcomes by altering the subjective attitudes of participants or staff members. For example, DTCs might enhance participants' intrinsic motivation for change or their respect for the justice system. In turn, these positive attitudinal changes might influence them to reduce their drug use or criminal activity. Similarly, staff members in DTCs might develop more optimistic attitudes towards rehabilitation, and this in turn might influence them to interact more productively with participants.

As will be discussed later, testing such hypotheses requires evaluators to perform what are called *mediation analyses*. Changes in participants' or staff members' attitudes are sometimes referred to as *mediator variables* because they occur in the middle between the delivery of DTC services and subsequent changes in participants' behaviors and are hypothesized to be a necessary condition for those behavioral changes to occur. Unlike moderator analyses, which as discussed earlier indicate *who* DTCs benefit, mediator analyses are intended to indicate *how* they function.

Participant Attitudes

Research in this area is relatively new; however, a few studies have found that participants' scores on attitudinal measures did appear to influence longer-term improvements in sobriety or desistence from crime. One study, for example, found that although many participants initially entered a DTC to avoid going to prison, those who gradually developed intrinsic motivation to improve their lives had significantly better long-term outcomes (Kirk, 2012). In other studies, significantly better outcomes were achieved by participants who perceived the DTC as applying fair procedures and having benevolent aims, character-

istics referred to as *procedural justice* or *procedural fairness* (Dane, 2012; Frazer, 2006; Gottfredson et al., 2007; Zweig et al., 2012). Better outcomes have also been achieved by participants who perceived a rational connection between their own behaviors and the imposition of sanctions and incentives in the DTC, a characteristic referred to as *perceived deterrence* (Marlowe et al., 2005).

The following attitudinal variables on the part of participants have been hypothesized to mediate outcomes in DTCs or have been examined in DTC program evaluations. It is premature to conclude whether these variables reliably impact long-term outcomes in DTCs, but evaluators in OAS member states may wish to examine these variables in their own research. The Appendix provides information on how to obtain assessment tools that measure these variables and that have been translated into Spanish or validated in multiple countries.

Procedural Justice or Procedural Fairness: A substantial body of research has found that criminal defendants and other litigants were more likely to react favorably to an adverse judgment or punitive sanction if they believed fair procedures were followed in reaching the decision. The best outcomes were achieved when the defendants were (1) given a reasonable opportunity to explain their side of the dispute; (2) treated in an equivalent manner to similar people in similar circumstances; and (3) treated with respect and dignity throughout the process (Burke & Leben, 2007; Tyler, 2007). Some commentators in Australia (King, 2009), Scotland (McIvor, 2009) and the United States (Burke, 2012; Wiener et al., 2010) have hypothesized that enhancing participants' perceptions of procedural justice may be the primary mechanism by which DTCs reduce crime and substance abuse.

Motivation for Change: DTCs were designed to treat persons suffering from addiction who have low motivation to change their substance use or criminal activity. The power of the court is used to maintain participants' adherence to treatment despite a potential desire to leave treatment prematurely. Eventually, however, participants must graduate from the DTC. If they have not developed intrinsic motivation by that time, the odds of long-term success are likely to be slim to none. Therefore, evaluators should periodically assess par-

ticipants' motivation for change at approximately three- to six-month intervals to determine whether it is improving over time. As was noted above, at least one study reported better outcomes for DTC participants whose internal motivation improved over time (Kirk, 2012).

Perceived deterrence: The criminal justice system usually does not apply behavioral contingencies such as sanctions and incentives effectively. Punishment may be administered after long delays or multiple infractions, and may be too severe or too lenient in magnitude to achieve effective results. One goal of a DTC is to administer rewards and sanctions in a systematic, rapid, and even-tempered manner that is consistent with scientific principles of effective behavior modification. Studies suggest participants in DTCs may recognize the rationality by which rewards and sanctions are administered and may respond favorably as a result (Goldkamp et al., 2002; Harrell & Roman, 2001; Marlowe et al., 2005).

Satisfaction with the Program or Treatment: Participants' satisfaction with substance use treatment, including the cultural sensitivity of treatment staff, is a reliable predictor of outcomes in community-based treatment programs in several countries (Richardson et al., 2011). Studies have almost uniformly reported high levels of satisfaction with the services provided in DTCs. For statistical reasons, variables tend to be less predictive of outcomes when most participants score the same or nearly the same on the assessment tools. Perhaps for this reason, satisfaction with the program has frequently not been found to be predictive of outcomes in DTC evaluations. If participants in other OAS member states have a wider range of reactions to DTCs, this variable might have better utility for predicting outcomes.

Therapeutic Alliance: The quality of the therapist-patient alliance has been found to be a reliable predictor of psychotherapy outcomes in several countries including Canada, Italy, Australia and the United States (Ardito & Rabellino, 2011; Summers & Barber, 2003). No study has examined the impact of the therapeutic alliance on outcomes in DTCs, but this variable (particularly the alliance between the participant and judge) has been hypothesized to influence outcomes in DTC programs as well. Research is needed to test this hypothesis in a range of DTC programs and in different countries.

Staff Attitudes

Considerably less research has examined staff members' attitudes as compared to participants' attitudes in DTC programs. Evidence suggests the more training staff members receive and the longer they work in DTC programs, the more likely they are to endorse optimistic attitudes about offender rehabilitation and to favor rehabilitative goals over punitive goals (Van Wormer, 2010). Changes in staff member attitudes have not, however, been correlated with outcomes in DTC programs. Additional research is needed to understand the degree to which staff member attitudes may impact outcomes in DTC's and how to influence those attitudes to bring about more effective behavioral change.

OBSERVER RATING SCALES

Observer rating scales may also contribute valuable information to DTC program evaluations. Observer rating scales may be used to assess objective events occurring in DTCs, such as the number of times the judge spoke directly to participants during court sessions. More commonly, however, they are used to assess subjective traits or characteristics of staff members or participants, such as the degree to which the judge appeared to be "competent" or "knowledgeable," or participants appeared to be "attentive" or "compliant."

In a few studies, independent raters observed court hearings in DTCs and rated the behaviors of the judge or participants. In one study of twenty-three DTCs, significantly greater reductions in crime and substance use were produced by judges who were rated as being more respectful, fair, attentive, enthusiastic, consistent, and caring in their interactions with participants (Zweig et al., 2012). Another study found that participants in a DTC were significantly less likely to be arrested for a new crime if they were rated as being more prepared and behaving appropriately during court hearings (Reingle et al., 2012). Still other studies have found that participants and staff members tended to rate court interactions similarly (Farole & Cissner, 2007; Satel, 1998). For example, they commonly agreed on whether interactions appeared to be productive or unproductive. The Appendix provides information about observer rating scales that have been used successfully in DTC program evaluations.

Observer rating scales typically require substantially more effort on the part of evaluators than self-report questionnaires. Raters must be trained carefully on how to score the instruments and must demonstrate high levels of inter-rater reliability or agreement with other raters. If two raters reach different conclusions about the same behavior or characteristics of individuals they are observing, then the information will not be useful to an evaluator. It can also be very costly to pay independent observers to attend court hearings or other sessions and score rating instruments. As a result, observer rating scales are most likely to be used in well-funded studies that are conducted by professional scientists or university-based researchers.

COMPARISON GROUPS

Performance indicators cannot be interpreted in a vacuum. To answer most research questions, the performance of a DTC must be compared to that of an equivalent and unbiased comparison group. Comparing what happened in the DTC to what would most likely have happened if the DTC did not exist is referred to as examining the *counterfactual hypothesis* or the possibility that the DTC was ineffective.

Some comparison conditions are reasonably unbiased and can yield a fair and accurate assessment of what would most likely have occurred without the DTC. Others, however, may be systematically biased or skewed in such a manner as to make the DTC look better or worse than it deserves. This may lead to an unwarranted rejection of the counterfactual hypothesis and the erroneous conclusion that the DTC was effective when, in fact, it was not.

Random Assignment

From a scientific standpoint, the best strategy is to randomly assign eligible individuals either to the DTC program or to an alternative disposition, such as probation. Random assignment provides the greatest assurance that the groups started out with an equal chance of success. Therefore, if outcomes are better for the DTC, these can be confidently attributed to the effects of the DTC program as opposed to differences in the severity of participants' problems before they entered the programs.

Even when an evaluator employs random assignment, there is still a possibility (although a greatly diminished possibility) that the groups differed on important dimensions from the outset. This requires the evaluator to perform a confirmation of the randomization procedure. As will be discussed later, the evaluator will need to check for possible pre-existing differences between the groups that could have affected the results. If such differences are found to exist, there are statistical procedures the evaluator might be able to employ to

adjust for those differences and obtain scientifically defensible results. For example, DTC participants might have had less severe drug problems to begin with than comparison participants. If so, then better outcomes for the DTC might simply have been the result of treating higher functioning individuals. If random assignment fails to avoid this problem, the evaluator will need to adjust statistically for differences in drug problems.

In many instances, random assignment is not feasible for legal, ethical, or practical reasons. Some staff members may have ethical concerns about denying services to otherwise eligible individuals. In addition, some DTCs may have difficulty filling their slots and may not want to turn away eligible individuals.

Nonetheless, random assignment is generally *not* considered to be unethical if appropriate safeguards are instituted (National Research Council, 2001). Random assignment has been used successfully in several research studies of DTCs in Australia (Jones et al., 2011) and the United States (Breckenridge et al., 2000; Gottfredson et al., 2003; Harrell et al., 1999; MacDonald et al., 2007; Turner et al., 1999). Safeguards may include requiring participants to provide informed consent to be randomly assigned, or having an independent ethics review board oversee the safety and fairness of the study. In some countries, local colleges or universities have ethics review boards with names such as Institutional Review Board (IRB) or Data and Safety Monitoring Board (DSMB), which have the authority and expertise to provide such ethical oversight for research studies.

Ethical scholars generally agree that random assignment is acceptable, and indeed should be required, if a treatment program has not yet been proven effective (Edwards et al., 1998; National Research Council, 2001). In many countries, for example, new medications must be evaluated through random assignment before being approved for public use. It may be unethical or unlawful *not* to use random assignment in final-phase medication trials. Random assignment becomes objectionable only after a treatment has been proven to be effective, and then it may not be ethical to deny that effective treatment to deserving individuals. In some countries, DTCs have been proven to be effective, and therefore it may

raise ethical concerns to use random assignment to decide who gets into those programs. In countries that have not yet established the effectiveness of their DTCs for their citizens, it is therefore not necessarily unethical to use random assignment

Additionally, random assignment poses far fewer ethical challenges if a DTC has insufficient capacity to treat many individuals who would otherwise be eligible for its services. If many deserving people must be turned away anyway, then it would perhaps 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 might provide the best protection against unfair discrimination or unconscious bias.

Random assignment may also be less objectionable if it is conducted within a DTC after participants have already been accepted. For example, studies have randomly assigned participants within DTCs to receive different schedules of court hearings (Festinger et al., 2002; Marlowe et al., 2006, 2007), different rewards for accomplishments (Marlowe et al., 2008a), or specialized treatment services (Messina et al., 2012). Because all participants were already admitted to the DTCs, each participant received at least some of the additional services and opportunities the programs had to offer. Any differences in the services they received were a matter of degree rather than being all or nothing. This arrangement may be also more ethically acceptable to staff members and other stakeholders of DTC programs if participants provide voluntary and informed consent to being randomly assigned.

Despite these safeguards, it might nevertheless be impractical to engage in random assignment. If random assignment is not possible, then several alternative strategies are available to select acceptable comparison conditions.

Quasi-Experimental Comparison Groups

The next best approach after random assignment is to use a quasi-experimental comparison group. This term refers to individuals who would have been eligible for the DTC but did not enter the program *for reasons that are unlikely to have affected their outcomes*. An

excellent example is individuals who were eligible for and willing to enter the DTC but were denied access simply because there were no empty slots available. This is referred to as a *wait-list comparison group*. The mere happenstance that the program was full is unlikely to have led to the systematic exclusion of individuals who had more severe problems or poorer prognoses than the average participant, and therefore is unlikely to have biased the results.

Less optimal, but still potentially acceptable, quasi-experimental comparison groups include individuals who would have been eligible for the DTC but were arrested in the year or so before the DTC was established, or who were arrested in an immediately adjacent county that does not have a DTC. Because these individuals were arrested at an earlier point in time or in a different geographic region than the DTC participants, there might still be systematic differences between the groups that could have biased the outcomes. For example, socioeconomic conditions or the types of drugs that are commonly used might differ between neighboring communities. However, the likelihood of this occurring is usually not substantial, and these may be the only practical comparison conditions that can be used for purposes of many DTC program evaluations.

When using a quasi-experimental comparison group, it is essential for the evaluator to check for pre-existing differences between the groups that could have affected the results. For example, the comparison individuals might have had more serious criminal histories to begin with. This, in turn, might have put them at greater risk for criminal recidivism. If so, then superior outcomes for the DTC group might not be attributable to the effects of the DTC program, but rather to the fact that it treated a less severe population. As will be discussed later, there are various statistical procedures an evaluator can employ to adjust for such differences and obtain scientifically defensible results.

Matched Comparison Groups

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 population of individuals involved in the justice system. 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 a DTC or are dissimilar to DTC participants on characteristics that are likely to have influenced their outcomes. For example, some probationers might not have had drug problems, or might have been charged with offenses that would exclude them from participation in a DTC. The evaluator must, therefore, select a subset of individuals from the probation pool that are comparable to the DTC participants on characteristics that are known to affect outcomes. For example, the evaluator might pair each DTC participant with a probationer who has the same criminal history, demographic characteristics and substance use diagnosis. Because the evaluator will choose only those probationers who are comparable to DTC participants on multiple characteristics, it is necessary to start with a large sample of potential candidates from which to select the comparison sample.

The success of any matching strategy will depend on whether the evaluator has adequate information about the comparison candidates to make valid matches. If data are not available on such important variables as the probationers' criminal histories or substance use problems, it will not be possible to place confidence in the validity of the matches. It is not sufficient to simply match the groups on variables that are easy to measure and readily available, such as gender, ethnicity or race, because the groups might have differed on other important dimensions that were not examined.

Propensity Score Analyses

An evaluator may also use a statistical procedure called a *propensity score analysis* to mathematically adjust for differences between the groups. This advanced procedure calculates the probability that an individual with a given set of characteristics would be in the DTC group as opposed to the comparison group—in other words, the relative similarity

of the individual to one group as opposed to the other. The analysis then statistically adjusts for this relative similarity when comparing outcomes between the groups. Advanced statistical expertise is ordinarily required to implement and interpret this 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, it will not be possible to place confidence in the adjustments. Again, it is not sufficient to merely adjust the scores based on easily measured variables such as gender or race, because the groups might have differed on other important characteristics that were never examined.

Invalid Comparison Groups

There are several comparison groups that have been used in DTC program evaluations which are likely yielded biased results. In most instances, it is *not* justified to compare outcomes from a DTC to those of persons who refused to enter the DTC, were denied access to the DTC due to their clinical or criminal histories, dropped out of the DTC, or were discharged unsuccessfully from the DTC. The probability is unacceptably high that such individuals had relatively poorer prognoses or more severe problems to begin with. For example, there is a high likelihood that they may have had more serious criminal or substance use histories, lower motivation for change, or lesser social supports. As a result, comparing them to the DTC sample is apt to unfairly favor the DTC program. Given the high likelihood that these groups were seriously disadvantaged from the outset, statistical adjustments cannot be relied upon to overcome the biased differences.

INTENT-TO-TREAT ANALYSES

A common error in DTC evaluations is to examine outcomes only for participants who successfully graduated from the program. The logic for performing such an analysis is understandable. Evaluators are often most interested in learning about outcomes of participants who received all services the program has to offer. If individuals who dropped out or were terminated prematurely are included in the analyses, then the outcomes will be influenced by those who did not actually receive all intended services.

Although this reasoning might seem logical, it is scientifically flawed. Outcomes should be examined for *all* individuals who participated in the DTC, regardless of whether they successfully graduated or were unsuccessfully terminated from the program. This is referred to as an *intent-to-treat* analysis because it examines outcomes on all individuals whom the program initially set out to treat. It is not appropriate to report outcomes only for graduates because this unfairly and falsely inflates the apparent success of the program. Individuals who graduated from the DTC are likely, for example, to have had less severe drug or alcohol problems to begin with, less severe criminal propensities, higher motivation for change, or better social support systems than the average DTC participant. 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 the DTC. The critical question is how the program fared for all participants who met the program's eligibility criteria and were accepted into the program.

This issue is particularly important if outcomes are contrasted against those of a comparison group, such as probationers. Selecting the most successful DTC cases and comparing their outcomes to all probationers is unfair. It is akin to selecting only the A+ students from one classroom, comparing their test scores to all students in a second classroom, and then concluding that the first class had a better teacher. This would clearly be a biased and unfair comparison.

This is not to suggest that outcomes for graduates are of no interest. Programs may, indeed, want to know what happens to individuals who received all services offered by the

DTC. However, this should be a *secondary analysis* that is conducted after the intent-to-treat analysis has been completed. If it is first determined that the program achieved positive 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. However, if the intent-to-treat analysis is not significant, it is generally not acceptable to move on to evaluate outcomes for graduates alone.

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

Cohorts

It is usually not justifiable to evaluate a program at a single point in time and assume the results are representative of what can be expected from the program in the future. Conditions often change over time and may significantly influence the effectiveness of a DTC program. For example, a DTC might experience substantial staff turnover or a decrease in funding which could seriously detract from the quality of the services it can offer. Conversely, staff members may improve their skills over time through experience and continuing education, which could significantly improve the program's performance.

It is important for evaluations to be conducted periodically to measure changes in performance over time. Commonly, this is accomplished by evaluating cohorts of DTC participants. As was mentioned earlier, a *cohort* is defined as a group of individuals who entered the program during the same specified time period, typically an interval of 12 months (Heck, 2006; Rubio et al., 2008). For example, participants who entered the DTC during the calendar years 2011, 2012, and 2013 might be defined as three cohorts, and outcome analyses could be conducted separately for each cohort. This would provide a series of "snapshots" indicating how well the program performed over consecutive years.

FOLLOW-UP WINDOWS

There is no one correct follow-up window or time period for evaluating the performance of a DTC. The most appropriate length of the follow-up window will depend on what performance indicator is being assessed. For example, it may be informative to measure participants' attendance in treatment during the first several months of a DTC program. It is important to know whether the DTC is successful at retaining participants in treatment and providing an adequate dosage of treatment. However, most DTC programs gradually decrease the dosage of treatment as participants move through the phases of the program. As a result, treatment attendance may become a relatively less important performance indicator after several months or a year. It may become more important after several months to know whether participants' substance use declined or their employment rate increased.

It is usually not very informative to measure new arrest rates or new conviction rates during the first few months of a DTC program. It typically takes several months for a participant to re-engage in criminal behavior, be detected by the authorities, and be formally charged or convicted in a legal proceeding. As a result, new arrest and conviction rates are likely to be low during the first several months of treatment. (This does not include technical violations, which may be more common.) This issue can be problematic if new arrests or convictions are contrasted against those of a comparison group, such as probationers. If recidivism is low in both groups, it will be difficult to detect statistically significant differences between the groups. This is true even if the DTC is, in fact, a superior program. For example, if ten percent of the probationers reoffended during the first six months and five percent of the DTC participants reoffended, for mathematical reasons this difference would probably not be statistically significant unless there were numerous participants in the study. However, it might be quite meaningful from a clinical or public policy perspective.

Reducing recidivism by five percentage points can provide important public safety and cost benefits for local communities.

In many studies, the best course of action may be to wait at least twelve months, and preferably twenty-four months, from entry before reporting recidivism outcomes. This should allow adequate time to elapse for reoffending rates to diverge sufficiently between the groups to detect statistical significance.

As was mentioned earlier, if it is feasible, recidivism should be tracked for at least three years post-entry and ideally up to five years post-entry. Research suggests most new offenses for drug and alcohol-involved persons occur within three to five years after treatment (Gossop et al., 2005; Inciardi et al., 2004; Martin et al., 1999). Therefore, following participants for three to five years should ensure that most recidivism events are accounted for in the evaluation results. Recidivism analyses may still be informative after the first or second year, and those interim analyses should certainly be reported as the data become available. However, following participants for three to five years is likely to elicit stable estimates of criminal recidivism.

Starting the Clock

In many evaluations, the follow-up window is started from the time participants first entered the DTC program or comparison condition. For example, outcomes for each participant might be measured over a period of twenty-four months starting from the date of entry into the DTC or probation. One reason for this practice is that it gives all participants an equivalent follow-up window. Each participant would have twenty-four months in which to engage in behaviors that are relevant to evaluation outcomes, such as committing a new offense or finding a job. Although it is not always clear when "entry" has occurred, most DTCs and probation programs have a formal entry hearing or comparable proceeding at which the participant enters a plea or is adjudicated and is officially enrolled in the program. The date of this hearing or proceeding is typically counted as the entry date because

it marks the point in time when the programs gained authority over participants to impose conditions of treatment and supervision.

Consider, instead, what would happen if outcomes were measured for twelve months from the date of *discharge* from the program. Some participants might have been discharged after twelve months of treatment whereas others might have been discharged after eighteen months of treatment. As a result, outcomes for the former participants would be measured for twenty-four months from entry and outcomes for the latter participants would be measured for thirty months from entry. The latter individuals would have six additional months in which to commit new offenses or engage in other outcome behaviors.

Time at Risk

Regardless of the length of the follow-up window, it is important for the follow-up window to be comparable for all participants. This ensures that all participants have the same *time at risk*, meaning they have the same opportunity to engage in substance use, crime, and other behaviors of interest to the evaluator. If, for example, an evaluator measured criminal recidivism over twelve months for the DTC participants but measured recidivism over twenty-four months for the comparison group, this would give an unfair advantage to the DTC participants. The comparison group would have twelve additional months in which to commit new crimes.

If the time at risk differs significantly between the groups, the evaluator might be able to compensate for this problem by statistically adjusting for the time at risk in the outcome analyses. For example, the evaluator might enter the time at risk as a covariate in the statistical analyses. A *covariate* is a variable that is entered first into a statistical analysis, and then the independent effect of the variable of interest (in this case, treatment in a DTC) is evaluated after first taking the effect of the covariate into consideration. This procedure would indicate whether DTC participants had significantly better outcomes than probationers after first accounting for the influence of their time at risk. The use of covariates is

not always successful, however, and the best course of action is to begin with equivalent follow-up windows whenever possible.

Time at Liberty

The issue of *time at liberty* is comparable to time at risk in that both relate to participants' opportunities to reoffend or engage in other outcome behaviors. The difference is that time at liberty refers 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 may severely curtail participants' ability to use drugs, commit new offenses, obtain a job, or engage in other behaviors of interest to the evaluator.

A common error in DTC evaluations is to neglect the issue of time at liberty (and time at risk) when contrasting outcomes to a comparison group. In some jurisdictions, for example, individuals who do not enter DTC may be more likely to receive a jail sentence. If they were jailed for a portion of the follow-up period, they might have had fewer opportunities to reoffend or use drugs than the DTC participants who were treated in the community. The evaluator might wrongfully conclude that the DTC caused participants to reoffend or use drugs more often than the comparison individuals, when in fact they simply had more time at liberty to do so.

Under such circumstances, the evaluator might need to statistically adjust for the time at liberty in the outcome analyses. For example, the evaluator might need to enter the time at liberty as a covariate in the statistical models. This would reveal whether the DTC participants had significantly better outcomes after first taking into account the influence of their time at liberty.

Evaluators are not always advised, however, to adjust for time at liberty. In cost analyses, for example, the time that participants spent in jail or a residential treatment facility might be an important outcome measure in its own right, and might be valued differently than

other variables from a fiscal standpoint. Deciding on whether to statistically adjust for participants' time at liberty requires careful consideration of the aims of each analysis.

Adjusting for participants' time at risk and time at liberty obviously requires evaluators to pay careful attention to the issue of date-stamping. It is essential to record the dates on which participants entered and were subsequently released from jail, prison, residential treatment facilities, home detention, and other restrictive settings. If these dates are not recorded, then it will not be possible to control for participants' time at liberty or time at risk as covariates in the outcome analyses, and the results of the evaluation could be seriously compromised.

Additionally, this once again highlights how valuable it is for DTC programs to take the time to set up quality data collection, entry, and management systems at their outset. Proper, consistent data practices in these areas will help provide future evaluators with reliable, useful data—and in turn make future evaluations both less expensive and more reliable.

AVOIDING COMMON ANALYTIC MISTAKES

This chapter discusses common analytical errors that have been made in many DTC program evaluations, and which have laid waste to hundreds of hours of work and thousands of dollars in evaluation fees. Guarding against these common errors can avoid substantial distress stemming from indefensible evaluation results.

Target Population

It is common practice for evaluators to correlate participant characteristics—such as age, gender, race, nationality, or number of prior convictions—with outcomes in DTC programs. Determining which types of individuals performed best in the program may help to identify the optimal target population for the DTC and may indicate which participants will require enhanced services in order to succeed. As was discussed earlier, participant characteristics that correlate reliably with outcomes are called predictor variables or risk factors because they predict results or put participants at risk for relatively poorer outcomes.

As was also noted earlier, however, there is a critical distinction between predictor variables and moderator variables. Predictor variables simply correlate with outcomes. For example, first-time offenders often have better outcomes in DTCs than individuals with prior criminal convictions. Therefore, criminal history is a predictor variable for DTC programs. Moderator variables, in contrast, indicate which types of participants performed better in the DTC as compared to an alternative program, such as probation. Individuals with prior convictions have poorer outcomes in *all* correctional rehabilitation programs, not just in DTCs. If every program chose to target first-time offenders because they tend to have better outcomes, there would be no programs available for the individuals who need treatment the most.

To determine whether participants' criminal history is a moderator variable, the evaluator must determine whether differences in outcomes between the DTC and com-

parison program varied in magnitude by participants' criminal history. For example, assume that ten percent of first-time offenders committed a new crime after being treated in a DTC, and fifteen percent of first-time offenders committed a new crime after being sentenced to probation. Assume further that twenty percent of offenders with prior criminal convictions committed a new crime after being treated in a DTC and thirty-five percent of offenders with prior convictions committed a new crime after being sentenced to probation. The difference in recidivism for the first-time offenders is only five percentage points (10 percent vs. 15 percent) but the difference in recidivism for offenders with prior convictions is fifteen percentage points (20 percent vs. 35 percent). In other words, the magnitude of the effect of the DTC is smaller for the first-time offenders than for the individuals with prior convictions. This is true even though recidivism was lower for the first-time offenders in the DTC than for offenders with prior convictions in the DTC (10 percent vs. 20 percent). Under these circumstances, participants' criminal history would be a moderator variable for the effects of the DTC.

Failure to appreciate the distinction between predictor variables and moderator variables has led many DTCs to treat the wrong population. For example, studies in the United States have revealed that DTCs reduced crime approximately twice as much and were approximately fifty percent more cost-effective when they treated persons with prior convictions or other risk factors for reoffending (Bhati et al., 2008; Carey et al., 2012; Lowenkamp et al., 2005). Yet many DTCs in the United States continue to serve first-time and low-risk offenders because these individuals are more likely to graduate and less likely to commit new crimes than repeat offenders. Because first-time and low-risk offenders typically perform just as well on probation or pre-trial diversion as they do in a DTC, the contributions of these DTCs to public safety and public health tend to be minimal (Sevigny et al., 2013).

To avoid making this same mistake, evaluators in OAS member states should perform what are called *moderator analyses* or *interaction analyses* rather than simply correlating participant characteristics with outcomes. These analyses reveal three findings that must be interpreted together. They reveal (1) which program produced better outcomes (e.g.,

DTC vs. probation); (2) which types of participants had better outcomes (e.g., first-time vs. repeat offenders) and; most importantly, (3) which types of participants performed better in which type of program. This third piece of information indicates the best way to match individuals to the most effective programs to meet their needs and indicates how a country or jurisdiction should marshal its resources to be most efficient and cost-effective. If the results of moderator analyses are the same in other OAS member states as they have been in Canada and the United States, they are likely to indicate that, all else being equal, first-time addicted offenders should typically be placed on probation or pretrial diversion, and repeat addicted offenders should be placed in a DTC.

Disadvantaged Groups

DTCs commonly serve persons who have historically experienced discrimination or reduced social opportunities due to their race, ethnicity, nationality, gender, physical or mental disability, religion, or socioeconomic status. It is important to determine whether members of these disadvantaged groups receive the same services and succeed at the same rate as other participants in DTC programs.

Many evaluators have attempted to address this issue by correlating variables such as race or ethnicity with outcomes. For example, studies in Australia (Rysavy et al., 2011) and the United States (Finigan, 2009; Marlowe, 2013) found that racial minorities and members of indigenous populations had significantly poorer outcomes in many DTCs than other participants. These findings have led some commentators to conclude—prematurely—that DTCs discriminate against persons of color (O'Hear, 2009).

It is therefore not sufficient to simply correlate outcomes with participant characteristics such as race or ethnicity. Other variables might be correlated with race or ethnicity and might be truly responsible for the differences in outcomes.

In fact, evidence suggests disparities for minorities in DTCs are *not* a function of race or ethnicity *per se*, but rather are explained by broader societal burdens which are often

borne disproportionately by racial and ethnic minorities, such as lesser educational or employment opportunities or a greater infiltration of crack cocaine and other seriously addictive drugs into some minority communities. When evaluators have adjusted statistically for these confounding variables in the analyses, the impact of race and ethnicity disappeared altogether (Dannerbeck et al., 2006).

This finding requires evaluators to account statistically for the influence of variables that may be correlated with race, ethnicity, or other characteristics of disadvantaged groups—especially participants' socioeconomic status and primary drug(s) used—and then determine whether race or ethnicity continues to predict poorer outcomes after these covariates have been factored out. This process can be accomplished in several ways, but the essential procedure is to follow three basic steps: First, the evaluator must determine whether variables such as socioeconomic status or drugs used are significantly correlated with race, ethnicity or other characteristics of disadvantaged groups. If the answer to this first question is yes, then the evaluator must determine whether these variables are also correlated with outcomes, such as graduation rates or new arrest rates. If the answer to the second question is also yes, then the third step is to enter these variables as covariates into analyses comparing outcomes between members of disadvantaged groups and other participants. If the disadvantaged groups continue to have significantly poorer outcomes after this third step has been followed, then and only then would it be justified to conclude that there are disparate impacts for those disadvantaged groups in DTCs.

Adjusting for Baseline Differences

As was discussed previously, most evaluations compare outcomes between DTC participants and those of a comparison group such as probationers. Prior to making these comparisons, it is necessary to rule out competing explanations that might account for differences in outcomes other than the effects of the DTC program. For example, the probationers might have had more severe criminal histories to begin with. If so, then superior

outcomes for the DTC participants might have had nothing whatsoever to do with the effects of the DTC program, but rather might simply have reflected the fact that the DTC treated a population that was already more likely to arrive at more favorable outcomes. Indeed, even when an evaluator employs random assignment, there is still a possibility (although a greatly diminished possibility) that the groups may differ on important/material dimensions.

If the groups did materially differ at the time of entry, all is not necessarily lost. There are procedures an evaluator can employ to adjust for the differences and obtain potentially defensible results. This typically involves a three-step process. First, the groups are compared on risk factors that are known to influence outcomes among drug-involved individuals charged with criminal offenses. These risk factors, such as age and prior criminal convictions, are listed earlier in this manual. If the groups are found to differ on any of these risk factors, the evaluator must next determine whether, in fact, these risk factors predicted poorer outcomes in the current sample. For example, although a younger age often predicts poorer outcomes among drug-involved offenders, it might not have done so in the present study. If it does not predict outcomes in the current study, the evaluator does not need to be concerned about this variable going forward.

Third, any variables that *both* (A) differ between the groups *and* (B) predict outcomes must be controlled for statistically in the outcome analyses. For example, if the groups differ by age, and the age of a participant predicts outcomes, then age must be entered as a covariate in the outcome analyses. How this is accomplished will depend on what statistical test is being performed. For example, the evaluator might use an analysis of covariance (ANCOVA) or a hierarchical regression analysis, which can adjust for covariates that might be confounding the results.

Unfortunately, in many evaluations the DTC might not have collected data concerning all risk factors that should be considered. For example, the evaluator might not have any way of knowing whether the participants had prior criminal convictions or had been in substance use treatment previously. Without this important information, there is no way of knowing

whether the groups differed on these characteristics when they entered the programs and whether such differences might have influenced the outcomes that were produced. If this critical information is lost or unavailable, the results of the evaluation cannot be relied upon with a reasonable degree of scientific confidence. (Once more, the value of DTCs' initial design and implementation of robust data collection, entry, and management methods is manifest.)

Impact of Specific Services or Interventions

Many DTCs will want to know whether specific services they provide or consequences they impose improve, hinder, or have no effect on outcomes. For example, a DTC practitioner might want to know whether administering punitive sanctions or referring participants to a new type of counseling group improved their outcomes and was worth the added effort and expense.

Ideally, such analyses should not be conducted *post hoc*, or after the fact. Assume, for example, that an evaluator correlated the number of sanctions that participants received with their outcomes in a DTC. It is determined from this analysis that participants who received more sanctions were less likely to graduate successfully from the program. This could lead the evaluator to the erroneous conclusion that sanctions made outcomes worse. But such a conclusion is likely to have confused cause with effect. It is more likely that poor performance led the staff to apply more sanctions. If so, then more sanctions did not cause worse outcomes, but rather worse outcomes elicited more sanctions.

Similarly, participants who have more severe addiction problems or who are not responding to standard treatment are more likely to be referred for more intensive or specialized treatment. Simply correlating treatment with outcomes could lead to the erroneous conclusion that more treatment led to poorer outcomes, when in fact poor performance often leads to a referral for more treatment. For example, participants with more severe addictions are more likely to be referred for residential treatment than outpatient treatment. If their long-term outcomes are worse than for participants who were referred

to outpatient treatment, the evaluator might reach the wrong conclusion that residential treatment is ineffective or harmful, when in fact residential programs simply treated a more serious population than outpatient programs.

The best method is to address these questions *a priori*, or beforehand. If possible, participants should be randomly assigned in advance to different types or dosages of services. For example, on a random basis some participants might be required to attend fifty hours of treatment whereas others might be required to attend twenty hours of treatment. Using this research design, any differences in outcomes may be reasonably attributed to the dosage of treatment and not merely to systematic referrals by staff.

An alternative approach is to conduct a *mediation analysis*. As mentioned earlier, a mediator variable is a third variable that affects the relationship between a predictor variable (such as enrollment in a DTC) and an outcome variable (such as recidivism or graduation rates). The mediator variable can be thought of as an intervening variable that explains the relationship between the predictor variable and the outcome variable. In order to confirm that there is a mediation effect, several statistical relationships must be established (Baron & Kenny, 1986; MacKinnon, 2008): (1) the predictor variable must be significantly correlated with the outcome variable, (2) the predictor variable must be significantly correlated with the mediator variable, and (3) the mediator variable must be significantly correlated with the outcome variable. If these three conditions are all met, then it must be further established that (4) the relationship between the predictor variable and the outcome variable is eliminated or significantly reduced in magnitude after controlling for the mediator variable as a covariate.

To illustrate, assume that an evaluator wanted to know whether DTCs were responsible for better outcomes than probation, because the former provide more hours of treatment to participants. To do so, the evaluator would need to establish that (1) being assigned to the DTC was correlated with better outcomes than being assigned to probation, (2) being assigned to the DTC was correlated with receiving more hours of treatment than probation,

and (3) receiving more hours of treatment was correlated with better outcomes. Then, the evaluator would need to establish that the effect of the DTC was significantly reduced after accounting for the amount of treatment participants received. If all three conditions are met, then the evaluator can reasonably conclude the DTC worked by increasing participants' exposure to treatment, which in turn led to better outcomes. In other words, increasing exposure to treatment partially mediated or partially explained the positive effects of the DTC.

Common statistical procedures that may be used to test for mediation include, but are not limited to, ordinary least squares (OLS) multiple regression analyses, structural equation modeling (SEM), logistic regressions (for binary mediator or outcome variables), and multilevel modeling (for clustered or nested data). In many instances, expert statistical consultation will be required to conduct these sophisticated mediation analyses correctly.

Infrequent Events

Some events occur infrequently in DTC programs. For example, if outcomes are examined during the first six months, arrests for new offenses might not occur at high rates. It usually takes a longer time than six months for participants to re-engage in criminal conduct, be detected by law enforcement, and have criminal charges filed against them. Among those participants who do quickly return to criminal behavior, one would not expect to see more than about one or two new arrests during the first six months.

A common mistake in DTC evaluations is to employ statistical tests such as t-tests or analyses of variance (ANOVA) which examine the average number of infrequent events, such as the average number of new arrests. Because the number of new arrests is likely to be zero for many participants, the average number of arrests is likely to be less than one or two. This creates what is called a *skewed distribution* or *non-normal distribution* of scores. For mathematical reasons, one is unlikely to detect statistically significant differences if the distribution is skewed in this manner. Although there might be a clinically meaningful difference between the groups, it is unlikely to be detected by the statistical test.

For infrequent events, it may be advisable to transform the data into a binary or dichotomous variable (e.g., any new arrest vs. no new arrest) and to use what are called *nonparametric* statistical tests. For example, rather than analyzing the average number of new arrests, the evaluator might analyze the percentage of participants who had at least one new arrest. Nonparametric tests can help to avoid the problems associated with skewed distributions and can lead to more sensitive and powerful findings. Examples of commonly used nonparametric tests include, but are not limited to, chi-square tests, Fishers Exact Probability Tests, Mann U tests, Wilcoxon tests, and logistic regression analyses.

Alternatively, evaluators may perform what is called a *log-linear transformation* of the data to compensate for the skewed or non-normal distribution. This procedure has the mathematical effect of smoothing out the skewed distribution or making it function as if it were a normal distribution. Expert statistical consultation will often be required to know when and how to employ an advanced compensatory procedure such as a log-linear transformation.

COST ANALYSES

Mike Finigan and Shannon Carey

DTC programs use resources from multiple systems and agencies (e.g., the court, the criminal justice system, and the substance abuse treatment system). Such programs thus need to be designed and implemented in a manner that maximizes benefits and minimizes costs to participants, the various public systems/agencies involved, and to society in general.

Policymakers and the public therefore have an interest in calculating and comparing these benefits and costs. One way to provide a familiar and useful standard for such calculations and comparisons is monetizing (putting dollar amounts to) the various benefits and costs through so-called cost-efficiency analyses. Comprehensive DTC cost-efficiency analyses require the following elements:

1) A careful process evaluation to identify times and places where costs and potential benefits are incurred;
2) Collection of quantitative data on program activities (e.g., from administrative databases such as criminal justice databases that contain dates of arrests and charges);
3) An analysis of the DTC program's benefits and costs; and
4) An analysis of an appropriate comparison group in an alternative to the DTC, to juxtapose with the analysis of the DTC program.

Naturally, putting theory into practice is often more complex than it may first appear. Potential challenges inherent in designing and carrying out cost-efficiency analyses include:

- Assigning dollar amounts to the value of certain DTC benefits can be difficult;
- Cost-efficiency analysis involves making certain assumptions (e.g., that the goals and outcomes sought by different programs are the same; that the dollar value of a cost or benefit in one location is the same as its value in another) that must be tested to ensure the integrity of the analysis; and
- Monetary benefits that a DTC may generate over an alternative approach are sometime accrued by agencies other than the DTC itself (e.g., state corrections agencies).

There are four major types of cost analysis that have been applied to DTCs as well as substance abuse treatment programs in general. They range from simple/inexpensive but less accurate approaches to those that are complex/expensive but more accurate. As such, each implies certain trade-offs that policymakers, evaluators, and DTC program staff will need to take into consideration.

Type 1: Program-cost analysis. Program-cost analyses involve calculating all of the expenses required to implement and operate a given program. They are the most basic form of cost analysis, and all other types build upon this type.

A basic program-cost analysis is used to show what resources are needed to fund a particular program. The program's budget(s) and an accounting of all other resources put into the program (in-kind resources, etc.) are required to properly assess a program's full cost. Consequently, the program costs that a basic program-cost analysis measures should include both the more visible and easily calculated direct costs, such as salaries and benefits for DTC staff, but also the program's indirect costs, such as those for support services, training, office space, and administrative overhead generally.

Due to their relative simplicity and inexpensive nature, program-cost analyses can be useful for programs that have few evaluation resources and/or need only to carry out a limited assessment of costs.

Type 2: Cost-allocation analysis. Cost allocation analyses determine the unit cost (or cost per unit) of a particular service in the program (e.g., the cost per court session, or the cost per screening). This provides the basic information required for more sophisticated cost-effectiveness analysis or cost-benefit analysis, the third and fourth types of analyses explained here. Moreover, cost-allocation analyses are vital to study DTC programs in which multiple agencies contribute resources and services.

Accordingly cost-allocation analyses require not only setting up accounting or budgeting systems in a way that allows program staff to allocate costs to the various services provided,

but also collecting information about program participants and services delivered.

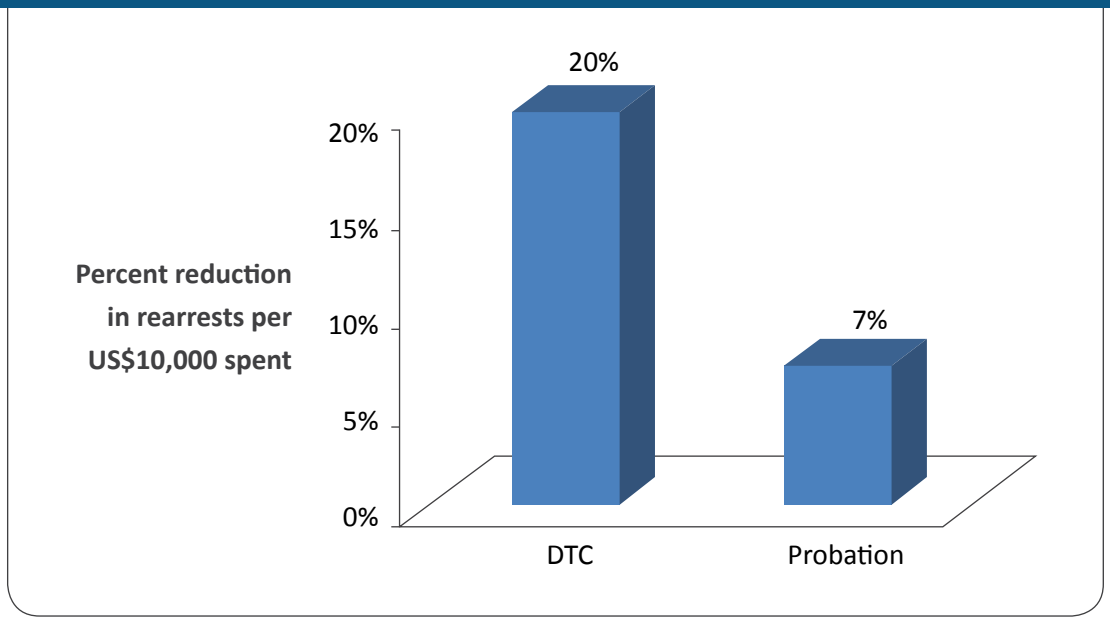
While cost-allocation analyses permit calculation of program and unit costs, they cannot by themselves determine whether the program is an efficient use of resources. Such a determination requires a comparison of the DTC program to other alternatives. For example, a similar program may produce equivalent results to the DTC program at a lower cost, thus making it more efficient in terms of costs. In order to reach conclusions about efficiency, a cost-effectiveness analysis is required, which is the next type of analysis considered.

Type 3: Cost-effectiveness analysis. Cost-effectiveness analyses evaluate alternative programs' costs and effects with regard to producing a desired outcome. In a cost-effectiveness analysis, the costs of different programs are assessed and then compared in terms of similar outcomes.

Nonetheless, keep in mind that program outcomes are not expressed in monetary terms, but rather in some other impact, such as a reduction in recidivism. In the context of a cost-effectiveness analysis, a reduction in recidivism is not express in monetary terms, but is instead expressed in a non-monetary value, such as the percent reduction in rearrests. In that context, the cost-effectiveness analysis would express results in terms the percent reduction in rearrests per a certain amount of money spent.

The following (see Figure 3) is an example of a cost-effectiveness analysis wherein a DTC is compared to a standard probation program cost, in terms of each program's effectiveness in producing the desired outcome (reduction in rearrests) per every US$10,000 in costs:

Figure 3. Compares the Impact of Drug Treatment Court and Standard Probation per US$10,000 spent



In this example, Figure 3 shows that DTC for drug offenders reduces the percentage of rearrests substantially more per each US$10,000 spent than the standard probation program. Accordingly, while cost-effectiveness analyses are more complex and require a full outcome evaluation for each program being compared, they are also a valuable tool for policymakers that wish to determine the cost efficiency of alternative approaches.

Type 4: Cost-benefit analysis. Lastly, a cost-benefit analysis (CBA) calculates the monetary cost of programs as well as the monetary cost of the outcomes to the system. This results in a cost-to-benefit ratio that monetizes both costs and outcomes—that is, a ratio that indicates that there are X dollars in benefits for every dollar invested in the program. A program with benefits that outweigh its costs is said to have *net benefits*.

CBAs provide the most complete picture of the value of a complex program such as DTC, and represent the only approach that allows an assessment of cost-benefit ratio. While this is the most comprehensive type of cost analysis, it also takes more time, effort, and

resources to conduct. CBAs require a full outcome evaluation, data on outcomes for both program participants and a comparison group, and cost data that can be associated with those outcome data. Nonetheless, given that policymakers are interested in seeing the monetary benefits of a new program before deciding to provide financing, CBAs can be extremely useful.

To provide a real-world example, one of the earliest CBAs of U.S. DTCs found the following: An estimated US$1,002,979 was spent per cohort of clients who participated in the DTC program each year. On the other hand, the DTC program resulted in US$2,476,795 of avoided costs (or savings). Therefore, the study calculated that every taxpayer dollar spent on those cohorts produced US$2.50 of savings to the taxpayer. Furthermore, if the broader costs (including victimization and theft costs) were estimated and included in the analysis, the ratio of benefit to the taxpayer was ten-to-one—that is, the program saved ten dollars for every dollar it spent. (Carey & Finigan, 2004).

Cost Assessment Techniques

The way costs are assessed also varies depending on the policy questions desired and the budget for cost evaluation. A number of assessment techniques exists, as explained below.

Funded budgets

Some cost analyses associated with the implementation of DTCs have assumed that a reasonable way of assessing the cost of DTC is to focus solely on the funds spent creating the DTC program. (In the United States, these are often federal grant funds spent on the project.) This approach, however, fails to account for resources that the existing criminal justice and treatment systems, which are not funded by DTC grants or other direct DTC appropriations, must contribute to the DTC program. Thus, the approach of using federal grant money or direct federal DTC appropriations to assess investment costs seriously underestimates the investment of state and local jurisdictional and agency resources for DTCs.

Marginal costs

The marginal cost represents the amount that the total program cost changes when a unit of output (also referred to as "workload") changes. In a cost-benefit analysis, "marginal" thus does not mean small or insignificant, but rather the costs that are at the margin of an existing level of operations. In other words, the term describes the cost or benefit that will be realized because of changes in units of activity.

In the context of the criminal justice system, the marginal cost is the amount of change in an agency's total operating costs when output (such as arrests, court filings, or jail days) changes because of changes to policies or programs. For example, costs for a prison—such as building maintenance, utilities, and overhead expense—are fixed and generally do not change whether or not a prison bed is filled. Variable costs such as food supplies, however, would change if there were fewer or more offenders in the prison.

The marginal cost approach argues that marginal costs should only include variable costs (e.g., overtime, supplies, food) and not fixed costs (e.g., rents, utilities), because fixed costs do not materially change with the addition of one more unit of workload (Henrichson & Galgano, 2013). Therefore, when using a marginal cost approach, only these variable costs should be counted in measuring increases or decreases in cost (or measuring monetary savings).

There are, however, several reasons that this marginal cost approach may not work in complex inter-organizational settings such as the DTC operating environment. First, it does not account for organizational or other structural changes that occur in a criminal justice system when a DTC is implemented. For instance, because the marginal cost approach assumes that the standard operation of the court remains the same, it fails to account for the fact that court calendars may change as judges increase or decrease caseloads to accommodate the DTC judge's caseload. Due to the same kind of assumption about the standard probation approach remaining the same when a DTC is added to the system, the marginal cost approach also fails to assess the systemic effect involved in caseload re-

assignments for probation officers resulting from the assignment of DTC caseloads. These caseloads typically have more intensive supervision requirements associated with the DTC program, a practice that affects both fixed and marginal costs. In addition, this approach often underestimates the costs of day-to-day operations, so-called "business-as-usual" costs that DTCs often offset or reduce. In other words, the marginal cost approach assumes that the DTC environment would not improve the cost of processing "business-as-usual" cases, and thus offset some of the new costs and/or eliminating some of these "business-as-usual" costs. It is clear that in some cases, the implementation of DTCs actually reduces the costs of processing eligible clients as compared to "business-as-usual." In a study of nine DTCs in California, one court found that processing clients through DTC actually saved the criminal justice system US$1,500 per client when both fixed and marginal costs such as rent and utilities were included in the analysis (Carey, Finigan, Crumpton, & Waller, 2006). Marginal cost analysis may have identified the "new" costs from the introduction of the DTC program and failed to assess the savings in the traditional system that accrued from this new method of processing cases. The same result was found in a DTC in Oregon (Carey & Finigan, 2004). It is also clear from several other sites that simply assessing the variable or semi-fixed costs of all DTC components that are new to the system would inaccurately represent the net cost of the DTC.

Broader societal outcomes

Some cost approaches can measure the impact of DTCs beyond their immediate impact on the criminal justice, court and substance abuse systems—something that some policymakers desire. A notable example of this approach is the recent analysis by the National Institute of Justice (the Multisite Adult Drug Court Evaluation, MADCE), which included social productivity and broader service use in its analysis (Roman, 2013) (see Table 2).

Table 2. Outcomes Measured by MADCE

Outcome Category	Outcome Sub-Category	Impacts
1. Social productivity	A. Employment B. Education C. Services and support provided	• Earnings • Schooling • Child support payments, community service
2. Criminal justice system	A. Monitoring B. Police C. Courts D. Corrections E. Drug Treatment Court	• Probation officer meetings, drug tests, electronic monitoring • Arrests • Hearings • Jail and prison (sanctions or otherwise) • Case management, administrative costs
3. Crime and victimization		• Crimes committed
4. Service use	A. Drug treatment B. Medical treatment C. Mental health treatment D. Other	• Emergency room, detoxification, residential care, outpatient, methadone • Hospital stays unrelated to drugs • Stays in mental health facilities unrelated to drugs • Halfway houses, public housing, homeless shelters

The MADCE researchers used a "bottom-up" approach, which comparing the costs and avoided costs for each participant in terms of social productivity, criminal justice system activities, crime and victimization, and service use. The resulting measure could be a positive or negative value for each person in the study (Roman, 2013). The MADCE found that in some broader areas such as service use and financial support the drug court participants cost more, yet the savings within the criminal justice system more than offset these costs. The overall result was a cost savings (although the savings in this particular study were not significant).

This broad approach produces excellent society-wide results but is difficult to implement because getting good cost data on impacted expenses requires extremely detailed measurement, as described earlier in the section on "Sources of Information." These include finding data on such questions as use of detoxification services, emergency room visits, lost wages or wage increases, tax payments, welfare payments, and disability payments (as described in Table 2). Data sources for these data elements are often quite difficult to find, or don't exist at all.

Transactional and Institutional Cost Analysis (TICA)

A different cost-benefit approach that uses both marginal costs and fixed costs comes from transactional economics, and is called Transactional and Institutional Cost Analysis (TICA) (Crumpton, Carey, & Finigan, 2004). The TICA approach views an individual's interaction with publicly funded agencies as a set of *transactions* in which the individual utilizes resources contributed from multiple agencies. Transactions are those points within a system where resources are consumed and/or change hands.

In the case of DTCs, when a DTC participant appears in court or has a drug test, resources such as judge time, defense attorney time, court facilities, and urine cups are consumed. Court appearances and drug tests can, in this context, be analyzed as transactions. In addition, the TICA approach recognizes that these transactions take place within multiple organizations and institutions that work together to create the program of interest. These organizations and institutions contribute to the cost of each transaction that occurs for program participants. Because TICA measures the costs of both transactions and the specific organizations that are contributing resources, TICA is a particularly appropriate approach to conducting costs assessments in an environment such as a DTC, which involves complex interactions among the multiple organizations involved in the DTC process.

Cost to the taxpayer

In order to produce information useful to policymakers, in the United States a "cost-to-taxpayer" approach is used. This focus on the taxpayer (or in the aggregate, the government) helps define which cost data should be collected (costs and avoided costs involving public funds) and which cost data should be omitted from the analyses (e.g., costs to the individual participating in the program). Although this "taxpayer" approach may not be applicable in some countries, a similar approach could be taken by focusing on a specific funding source (or a subset of funding sources) that are of particular relevance or import in the operation of the program. Focusing on a smaller subset of funding sources can help narrow down and better define the breadth of data that must be collected, which may make a cost analysis more feasible to accomplish with a limited budget, or a limited time.

Opportunity resources

Finally, the TICA approach looks at costs as "opportunity resources." The concept is similar to that of opportunity *cost* from the economic literature, which suggests that system resources are available to be used in other contexts if they are not spent on a particular transaction. The term opportunity *resource* describes these resources that are now available for different use.

For example, if substance abuse treatment reduces the number of times that a client is subsequently incarcerated, the local sheriff may see no change in his or her budget. However, an opportunity resource will be available to the sheriff in the form of a jail bed that can now be filled by another person who, perhaps, possesses a more serious criminal justice record than does the individual who has received treatment and successfully avoided subsequent incarceration.

Notably, the TICA approach does not use a marginal cost approach. All costs reported are *fully loaded*,[3] total costs. Thus, the TICA approach does not assume that fixed costs, such

3 Fully-loaded costs are those that include direct costs (such as salaries), as well as indirect costs such as bene-

as for building maintenance, utilities, salaries, or other overhead expenses, are off-limits. Those costs are still borne by funders and can change over time or be deployed for different purposes.

The TICA approach also does not make assumptions as to whether or not agencies will change their budgets in ways that enable savings to be literally realized. All "savings" calculated using the TICA approach represent "opportunity resources" that are newly available to government or agency actors to deploy as they choose. Rather than assume that direct dollars will be immediately saved (e.g., through reductions in the following year's budget), the TICA approach fully quantifies all resources (marginal or fixed)—such as a jail bed, treatment slot, or time spent in a court hearing—that are newly made available for a different offender or a different purpose.

Table 3 provides the six key steps in the TICA approach.

fits and overhead.

Table 3. The Six Steps of TICA

	Description	Tasks
Step 1:	Determine flow/process (i.e., how program participants move through the system)	Site visits/direct observations of program practice Interviews with key informants (agency and program staff) using a drug court typology and cost guide (See guide on www.npcresearch.com)
Step 2:	Identify the transactions that occur within this flow (i.e., where clients interact with the system). See Tables 4 and 5 for a list of program transactions.	Analysis of process information gained in Step 1
Step 3:	Identify the agencies involved in each transaction (e.g., court, treatment, police)	Analysis of process information gained in Step 1 Direct observation of program transactions
Step 4:	Determine the resources used by each agency for each transaction (e.g., amount of judge time per transaction, amount of attorney time per transaction, number of transaction events)	Interviews with key program informants using program typology and cost guide Direct observation of program transactions Administrative data collection of number of transactions (e.g., number of court appearances, number of treatment sessions, number of drug tests)
Step 5:	Determine the cost of the resources used by each agency for each transaction (e.g., staff salaries, overhead rates, supply costs)	Interviews with budget and finance officers Document review of agency budgets and other financial paperwork
Step 6:	Calculate cost results (e.g., cost per transaction, total cost of the program per participant)	Indirect support and overhead costs (as a percentage of direct costs) are added to the direct costs of each transaction to determine the cost per transaction The transaction cost is multiplied by the average number of transactions to determine the total average cost per transaction type These total average costs per transaction type are added to determine the program and outcome costs.

As part of a typical process evaluation, Step 1 (understanding the flow, or process, of the program), Step 2 (identifying the transactions in the process) and Step 3 (identifying the agencies involved) are performed (see the section of this manual on process evaluation). A thorough process evaluation should provide you with an understanding of the program process, including where transactions are occurring (e.g., places where program staff time or agency resources are being used) and what agencies are involved in each transaction. For example, when performing a site visit to the program, an evaluator can observe a DTC hearing (i.e., a transaction), the evaluator can observe how the staff and participants interact (the program flow), and which agencies are in attendance at the hearing.

Step 4 (determining the resources contributed by each agency) occurs through interviews and observation, as well as gathering quantitative data on numbers of transactions for each participant. Continuing with the example of the court hearing, the evaluator can use a stopwatch (or a clock) to time how long the staff from each agency spends in the court room. In addition, each of the staff members should be interviewed to ask them how long they spend preparing for the court hearing and how much time they spend on other DTC-related transactions. Finally, data on the number of court hearings attended by each participant in the program should be collected. If the program is collecting dates of court hearings in a program database, the number of court hearings attended is easily counted. If the program is not collecting this information, the evaluator can estimate the number of court hearings based on program policy on the frequency of court hearings required for participants (although this estimate will likely be less accurate).

Step 5 (determining the cost of the resources) is calculated by determining staff salaries (how much is each staff member paid for his or her time) and reviewing agency budgets for expenditures related to DTC activities, as well as to determine staff benefit rates and any overhead rates the agency may have. These rates are generally expressed as a percentage of staff salaries.

Step 6 (calculating the cost results) mostly involves simple multiplication and addition. For a court hearing (one of the most complicated transactions, as it involves every team member and agency that contributes resources), the salary of each staff member is multiplied by the percentage of time spent both preparing for and attending the court session. The cost for each staff member is then multiplied by the overhead rate and benefit rate for that agency. The "fully loaded" costs for each staff member are then added together for the sum total cost of the court appearance transaction. This total cost can be divided by the number of participants that attend a court session to determine the cost of a DTC appearance per participant. Following is a sample calculation.

If a judge works 40 hours per week, and spends 10 hours per week of that time preparing for DTC status review hearings (e.g., gathering information from the DTC team, participating in pre-court staffing meetings) and participating in the DTC hearings themselves, then the judge spends 25 percent of his or her time on preparing for or participating in DTC hearings. If the judge earns US$100,000 per year, then US$25,000 (US$100,000 X .25) goes toward status review hearings each year. Dividing US$25,000 by the 52 weeks in a year results in a cost of US$480.77 per week. If a DTC program sees 50 participants in status review hearings each week, then dividing US$480.77 by 50 results in a cost of US$9.62 per participant per hearing for the judge's time.

To achieve the "fully loaded" cost, it is necessary to add the cost of any fringe benefits earned by the judge (e.g., the judge may earn an additional 20 percent of his or her salary in vacation time, retirement payments, and health care benefits) and add the cost of any overhead paid by the court (e.g., a five percent overhead rate to heat and light the courthouse where the judge works, etc.). So, the fully loaded cost for the judge to participate in a single participant's court hearing would add this additional 20 percent and five percent to the US$9.62, resulting in a fully loaded unit cost of a court hearing for the judge of US$12.03 per DTC participant.

This exercise is repeated for every DTC team member that participates in status review hearings, and the fully loaded costs for all team members are summed to determine the total unit cost per court hearing per participant. Finally, to find the cost of status review hearings per participant throughout the length of the program, this unit cost is multiplied by the number of hearings attended by each participant (see Table 4).

A major advantage of this approach is that it offers a thorough assessment of both the investment costs of DTC for each transaction by each participating agency or organization in the system as well as the avoided costs due to the benefit of the outcomes to each agency or organization. For example, Tables 4 through 6 below demonstrate the kinds of costs and savings DTCs have shown using the TICA approach.

Investment costs

Table 4 provides the unit cost per transaction (e.g., each court appearance for a DTC participant costs US$83.62) and the average number of each type of transaction event per DTC participant (e.g., on average, participants in this program had a little over 23 appearances before the judge). The total cost per transaction is the unit cost (US$83.62) times the number of transaction events (23 court appearances), which equals a total transaction cost for court appearances of US$1,945 per participant. The cost of the DTC program is then the sum of the total costs of all transactions.

Table 4. Program Costs per Participant

Transaction	Unit cost	Avg. # of transactions/events per DC participants Per person	Avg. Cost per DC participant Per person
Drug Treatment Court Appearances	US$83.62	US$23.26	US$1,945
Case Management Days	5.83	494.37	2,882
Outpatient and Residential Treatment Months	746.74	12.00	8,961
UA Drug Tests	7.00	53.28	373
Jail Sanction Days	46.85	12.48	585
Jail Days While Awaiting Residential Treatment[4]	46.85	20.42	957
TOTAL			**US$15,703**

The unit cost multiplied by the number of events per person results in the cost for each transaction during the course of the program. In this example, when the costs of the transactions are summed, the result is a total program cost per participant of US$15,703.

Similarly, these same steps (TICA Steps 1-6) can be performed for a comparison group. (This is a comparison group that would be used in the outcome evaluation and this is why a cost-benefit analysis requires a valid outcome study). For example, using process evaluation methods, an evaluator can determine the process that occurs for a DTC-eligible court case for individuals who did not participate in DTC, determine the transactions that occur through that process, what agencies are involved, etc. Then, the difference between the cost of traditional case processing and the cost of participating in DTC (e.g., the cost of DTC minus the cost of traditional case processing) results in the cost of running a DTC over and

4 When a residential treatment bed is not available, this program's participants are sent to jail until a bed opens up.

above the cost of traditional case processing (i.e., the cost the system would have paid for that case anyway, if there had been no DTC). It is important to note that this requires good administrative data collection on the comparison group as well as on DTC participants, which does not always occur.

Outcome costs

To determine the outcome costs and savings, Steps 1 through 6 of the TICA approach are performed for each outcome transaction of interest (e.g., the cost of a new arrest, a new court case, days spent incarcerated). The cost per outcome transaction is summed across all transactions for both the DTC participants and for a matched comparison group. For example, Table 5 shows the outcome transactions measured in the same DTC as Table 4. The outcome transactions included in the study were arrests, court cases, days on probation, days in jail, and days in prison. These outcomes were measured during a three-year period from the time of DTC entry for each participant in the study. Table 5 provides the unit cost per transaction (e.g., the cost of a single arrest is US$129.47), and the total cost of each transaction per DTC participants and per comparison group member (e.g., the average number of arrests per person in 3 years was multiplied by the unit cost of US$129.47 for a total cost of US$66 per participant, revealing that there was an average of less than one rearrest per participant during that period). The right-hand column provides the difference in cost between the two groups (the cost of the comparison group minus the cost of the DTC group). When this difference is positive, this is considered a cost "savings." The sum of the savings across transactions is the total savings per participant due to participation in the DTC program. In Table 5, the total savings is shown to be US$23,957 per DTC participant.

Table 5. Recidivism (Outcome) Costs per Participant Over 3 Years

Transaction	Unit cost	Drug court participants Per person (n=103)	Comparison group Per person (n=358)	Cost difference Comparison minus drug court
Rearrests	US$129.47	US$66	US$388	US$322
Criminal District Court Cases	1,448.02	738	4,344	3,606
Probation Days	3.15	559	1,053	494
Jail Days	46.85	1,849	5,100	3,251
Prison Days	49.93	4,038	20,322	16,284
TOTAL		**US$7,250**	**US431,207**	**US$23,957**

Cost-benefit ratio

This table indicates that the cost for those processed through the DTC per person was US$23,957 less than standard court processing. It also allows calculation of a cost-to-benefit ratio. In the case of this DTC, when the cost of the program investment per participant (US$15,703) is assessed against the outcome savings per participant (US$23,957), (i.e. US$23,957 divided by US$15,703) the cost-benefit ratio comes to 1:1.5. That is, for every dollar spent on the program, there is a return of US$1.50.

In this example, the cost-benefit ratio may actually be an underestimate, since the cost of traditional court for the comparison group was not included in this example. When the investment in a traditional court case is included, the total investment cost used for this ratio would then be for every dollar spent *over and above* the cost of traditional court. So, if the cost of traditional court was US$10,000 per case, the "net" investment would be US$15,703 minus US$10,000 for an investment of US$5,703 per participant in the program over and

above the cost of traditional court. The cost-benefit ratio would then be 1:4.2—that is, for every dollar spent *over and above* the cost of traditional court, there is a return of US$4.20.

In addition, because the TICA approach measures the resources (e.g., each staff member) contributed by each agency involved in the program, it is possible to determine the investment and savings for each agency. Table 6 shows an example of the cost or savings incurred by each agency due to the cost of outcomes for DTC participants compared to a traditional court system.

Table 6. Recidivism (Outcome) Costs per Participant by Agency over three years

Agency	Drug Court outcome costs per participant	Comparison group outcome costs per individual	Difference/ savings per individual
Criminal District Court	US$161	US$375	US$214
District Attorney	194	453	259
Appointed Defense Attorney	383	895	512
Department of Criminal Justice	4,000	4,240	240
Law Enforcement	1,915	5,249	3,334
Adult Probation	559	1,037	478
TOTAL	**US$7,212**	**US$12,249**	**US$5,037**

Table 6 demonstrates that every agency involved in the program realizes some savings due to an offender's participation in the DTC program. In this example, law enforcement realizes the most savings due to decreased incarceration costs.

Conclusion – Cost Analyses

The type of cost analysis approach depends to a large extent on the data available and the policy questions being addressed, as well as the budget available to perform the cost analysis.

- ***Program-cost analysis*** can be useful when a program has few evaluation resources and needs only a minimal assessment of program costs.
- ***Cost-allocation analysis*** is vital in situations such as DTCs where multiple agencies contribute resources and services. Nonetheless, while cost-allocation analysis allows you to calculate program and unit costs, it cannot determine whether the program is an efficient use of resources.
- ***Cost-effectiveness analysis*** is useful when one has clear policy goals to produce the same outcomes, and outcome data for all the options, but does not need to place a monetary value on the outcomes.
- ***Cost-benefit analysis*** (such as the TICA approach) is useful when one needs to determine whether a program actually saves money and wants to compare the investment costs to the savings (or avoided costs) that result from positive outcomes.

SELECTING STATISTICAL CONSULTANTS

Apart from relatively simple types of analyses, it is often advisable to rely on experienced evaluators to perform appropriate statistical tests, avoid common analytic mistakes, and interpret the findings. If the data have been recorded properly and stored in the right format, it should not take long for an evaluator to analyze the results. The big question, however, is how DTC staff members—who have typically been trained as lawyers, clinicians, or criminal justice professionals—can identify competent evaluation consultants.

The first step in selecting the right team of evaluators is to review their prior evaluation reports, especially any evaluations related to DTCs or other problem-solving court programs. Familiarity with how DTCs operate is often essential for knowing what performance indicators, mediator variables, moderator variables and outcomes to examine. It is also important for knowing how to interpret the results and describe the findings in a practical and understandable way to DTC staff and treatment professionals.

When reviewing prior evaluation reports, it is useful to consider the questions listed below. These are not presented in order of importance, and some questions might not be relevant for all evaluations. For example, if the evaluators did not contrast outcomes against those of a comparison sample, then there might be no need to control for baseline differences between the groups. However, under such circumstances the evaluators would be extremely limited in terms of the conclusions they could draw from the study. Lacking a comparison sample, it would not be possible to conclude how well the DTC performed in comparison to other programs. If the evaluators went beyond the data in making such an interpretation, that might bode poorly for their competence as researchers.

- Did the evaluation team analyze important and relevant performance indicators?
- Did they properly account for missing data?

- Did they examine outcomes according to participants' risk factors or moderator variables?
- Did they identify the best target population for the program by examining interaction effects rather than merely examining significant predictors of outcomes?
- Did they rule out potential confounding variables before identifying disparate impacts by gender, race, ethnic or cultural groups?
- Were the analyses conducted on an intent-to-treat basis?
- Did they select an unbiased comparison sample?
- Did they check for baseline differences between the groups at entry and statistically control for them where necessary?
- Did they use a reasonable follow-up window given the outcome variable being measured? For example, did they examine recidivism over a long enough time period for new arrests or convictions to be expected to occur?
- Did they start the clock running at an equivalent time for all participants? For example, did they start it running at the point of entry into either the DTC or an alternative disposition such as probation?
- Did they statistically control, where relevant, for days at risk and days at liberty in the community?
- Did they transform the data or use nonparametric tests for infrequent events?
- Were the interpretations of the findings defensible given the limitations of the research design?

Finally, the most important question of all is: *Did the evaluators' interpretations of the findings make sense, and did they point to concrete actions the DTC could take to improve its performance and enhance outcomes?* The primary goal of a program evaluation is to improve outcomes and continuously move forward. If an evaluation report simply states whether or not the program worked in the past but does suggests how it could work better in the future, it is of very limited utility.

MANAGEMENT INFORMATION SYSTEMS

The reader has undoubtedly found the previous discussions concerning such topics as participant characteristics, performance indicators, attitudinal variables, mediation analyses, and cost analyses extremely daunting, especially when issues of date-stamping and the problem of the missing denominator are considered. The unfortunate reality is that few DTC programs collect the necessary information for a valid evaluation in a reliable or usable manner. When programs do capture the appropriate data elements, they rarely account for missed appointments, connect events to the dates on which they occurred or were supposed to have occurred, or enter the data into an electronic format that permits statistical analyses. In many instances, evaluators are required to extract information from written records or unwieldy statistical spreadsheets with little recourse for reconciling inconsistencies in the data or accounting for missing entries. The result is that many evaluations are completed months or years after the fact—when the results may no longer reflect what is occurring in the program—and there may be so many gaps or caveats in the data that the conclusions which can be drawn are tentative at best.

Date-stamping can be particularly unwieldy if evaluators use traditional forms of statistical spreadsheets, such as Excel. Most spreadsheets are two-dimensional, meaning they are comprised of rows and columns. If, for example, participants' names are listed by rows and the services that are provided are listed by columns, how does one account for the dates on which the services were delivered? It is usually necessary to create separate columns for each service on each date, resulting in hundreds of columns. Fortunately, newer generations of data-entry systems may automatically date-stamp entries. For example, data-entry screens might appear like a professional's appointment calendar. Information can be entered on the appropriate day in the calendar and is automatically date-stamped for analysis. DTC programs are encouraged to use management information systems (MISs)

that automatically date-stamp entries.

It cannot be stressed enough that the secret to valid, timely, and cost-efficient program evaluations depends on the selection of a suitable MIS. The cost of purchasing a useful MIS will be offset many times over by providing greater efficiencies in operations and yielding the type of performance feedback that is necessary to continuously improve and fine-tune one's DTC program.

Some of the older and less sophisticated MISs can be obtained free of charge. For example, the "Buffalo System"—so named because it was developed in a DTC in Buffalo, New York, in the United States—can be obtained for free by contacting the NADCP in the United States. Newer MISs must be purchased, but they are also more likely to be web-based and can, therefore, be accessed simultaneously by multiple programs. Having multiple programs use the same MIS, each with its own secured and encrypted access, can spread the costs of the MIS across multiple users or countries. The Appendix provides information on how to access the Buffalo System and other proprietary MIS products.

Recommended MIS Features

Newer generations of MISs are capable of streamlining program evaluations and reducing the burden on staff members and participants. Where it is feasible, it is recommended that DTC programs select MISs with the following characteristics:

Web-Enabled. Staff members in DTCs are usually employed by different agencies such as the court, probation department or treatment program, and they may have offices in several locations. They may also be required to visit participants in jail, residential treatment facilities or at their homes to conduct assessments or deliver services. This requires staff members to have access to the MIS while they are traveling and from multiple locations. If the MIS is web-enabled, it can be accessed from any location that has internet access, including personal palm devices or laptop computers.

Security Protected. The data should be stored and transferred using indus-

try-standard 128-bit SSL encryption or better. It is also necessary to regulate staff member access to the information based on their job levels and descriptions. Only under carefully specified circumstances should one staff member be able to alter data that was entered by another staff member. For example, the judge should not be able to alter information that was entered by the treatment provider. The judge might, however, have read-only access to certain information entered by the treatment provider, such as a participant's attendance rate at counseling sessions. The authorized level of access for each staff member should be specified by an MIS Administrator and correspond to that staff member's password and username.

Less is More. Staff members should only be required to view data-entry screens that are relevant to their jobs. For example, a treatment provider ordinarily should not be faced with data-entry screens relating to probation contacts or court hearings. The treatment provider might be permitted to view summary reports on probation contacts or court hearings but should not be required to scroll through that material if it is not directly relevant to the treatment provider's duties.

Need to Know. DTC professionals have a right to know why they are being asked to collect information and should avoid duplication of effort. If there is no obvious and empirically defensible reason why particular information is being collected, then perhaps it is unnecessary to collect that information. Redundancies should also be eliminated. For example, once a participant's age is entered in one data-entry screen, it should automatically be entered or "cross-walked" into the respective fields in other screens.

Minimal Burden. It should ordinarily require no more than two to three minutes to enter all data elements that are required for a given participant during a given week. The data-entry screens should also be intuitive and easy to use. Most professionals are accustomed to using the Internet to pay their bills, purchase goods and services, or gather information. A good deal of effort has gone into developing commercial websites that are intuitive and simple to use. The MIS for the DTC should have a similarly intuitive design. The

more complex, confusing, and/or time-consuming the effort required to enter data proves to be, the less likely staff will enter data correctly and appropriately.

Automatic Date Stamping. As was mentioned previously, some data-entry screens appear like a professional's appointment calendar. Information is entered on the appropriate date in the calendar and is automatically date-stamped for analysis.

Reminders/Ticklers for Missing Data. Modern MISs may routinely remind or "tickle" users to enter missing or incomplete data. For example, if a probation officer accesses the MIS to enter attendance information for the week of October 15, but has not entered the attendance information for the previous week of October 8, the system should alert the probation officer about the missing data. Such reminders or ticklers make it less likely that so much time will pass that it becomes difficult to reliably reconstruct the events from memory or handwritten records.

Flexible Input Screens. It is often necessary to add new items, delete or "gray out" unwanted items, or change the wording of items. For example, a DTC might contract with a new treatment program to provide mental health services. This might require a new data-entry screen to be added with items pertaining to participants' attendance at mental health counseling sessions. It should be possible to add new items, delete items, and change the wording of items in no more than an hour.

Hidden Database. DTC professionals should not be required to wade through columns of numeric data. The data should be stored behind the scenes in an analyzable format that permits immediate statistical analysis and queries, and allows easy and quick visualization through an interface that intuitively makes sense to the user. The data should also be stored in specified fields that can be selected readily and entered in statistical models.

Longitudinal Database. Under no circumstance should the system overwrite previous data. For example, if a participant was unemployed when he or she entered the DTC, and then obtained a job a few months later, the participant's employment status should not be overwritten. Doing so would make it difficult or impossible to determine later whether

the participant obtained a job while he or she was in the DTC, or whether the participant entered the program with a job to begin with. Moreover, it would be difficult to determine how long it took the participant to find a job. The database should be arranged longitudinally, meaning that new events are appended alongside older events. If the database is arranged longitudinally and the entries are date-stamped, it will be possible to determine whether participants' employment status changed over time and how long it took for the changes to occur.

Descriptive Reports. There is no limit to the research questions that DTC staff members or stakeholders might ask. However, there are a limited number of basic descriptive analyses that most programs will want to conduct. For example, virtually every DTC will want to know its graduation rate, re-arrest rate, attendance rate in counseling sessions, and proportion of alcohol- and drug-positive tests. In addition, many programs will want to know whether these outcomes differ between males and females or between racial or ethnic groups. It is possible to write statistical syntax so that these common analyses can be performed at the push of a button. For more sophisticated research questions, it may be necessary to send the data to a statistician for analyses; however, for many routine questions it should be possible to generate reports nearly instantly. In line with the adage that "a picture is worth a thousand words," the reports should also include easy-to-interpret graphs, tables and pie charts.

Continuous Performance Feedback. The most important reason for evaluating a DTC is to improve its operations and outcomes. If an intervention is not working, it is important to learn this fact quickly while there is still time to adjust the participant's treatment plan. This requires the MIS to continuously monitor the data and issue automated alerts to staff members whenever a particular action might be called for. For example, the system might automatically alert staff whenever a participant missed two counseling sessions in a row or provided two drug-positive urine specimens. This would prompt the staff to discuss the participant at the next staff meeting and decide on a suitable response. Research indicates that

generating automated feedback in this manner can substantially improve a DTC's operations and participant outcomes (Marlowe et al., 2008b; Rempel, 2007).

User Accountability Reports. One of the biggest threats to a valid program evaluation is the failure of many staff members to enter data in a timely manner. The MIS should generate user accountability reports which indicate how long it took for the data to be entered after the relevant events occurred. For example, supervisors should be made aware that a counselor has been entering attendance information an average of four weeks after the sessions were held, or that drug test results are being listed as "pending" by the lab for an average of more than two weeks. This would enable the supervisor to intervene rapidly to improve the quality of the evaluation and the effectiveness of the DTC.

At the risk of repetition, *if an MIS is easy to use, collects the essential performance indicators, and stores the data in an analyzable format, the likelihood of completing a successful and valid evaluation is high*. And the cost of purchasing such an MIS will be offset many times over by the fact that there is no longer a need for evaluators to spend hundreds of hours attempting (often with limited success) to extract usable information from written charts or spreadsheets. If the necessary information can be handed to researchers in a proper format, useful findings should be obtainable in a short period of time and at a minimal cost.

Of course, if a DTC does not have adequate resources to purchase an MIS with the above characteristics, any effort at evaluation is better than no effort. The DTC could use the Buffalo System described earlier, which is free of charge, or develop written checklists or data spreadsheets to collect information. Even an Excel spreadsheet, if configured and used correctly, can handle a variety of data collection tasks. Getting started evaluating one's DTC is the hardest step. Once evaluation activities are begun, DTC staff members are apt to identify better ways over time to collect useful information and examine the services provided and impacts produced by their program.

ACCESSING ADMINISTRATIVE DATABASES

Much of the information that is required to evaluate a DTC is likely to be in the possession of the program. For example, data on participants' drug tests results, treatment attendance information and graduation rates should be readily available in court records or participants' treatment records. However, to construct a suitable comparison sample, an evaluator will also need information about similar individuals who were not enrolled in the DTC, such as probationers. For example, the evaluator will need information about risk factors of the probationers, including their criminal records or treatment histories. Without this information, it will not be possible to construct a valid matched-comparison sample or to perform a valid propensity-score analysis.

Evaluators will also need data on new arrests, new convictions and incarcerations for both the DTC and comparison samples. This recidivism data will often not be in the possession of the DTC program, but rather may be collected and stored by another government agency, such as the police, department of corrections, or administrative office of the court system. This will require evaluators to obtain information from databases that may not be within their control, which they may not be authorized to access, and which may be unfamiliar to them.

Often there may be technological or attitudinal barriers to accessing other databases. Laws and regulations concerning the sharing of criminal records may be contradictory or difficult to interpret in some countries. This can lead to paralysis on the part of some agencies for fear of committing a breach of privacy or other legal protection. For technological reasons, information often cannot be transferred readily from one database to another due to inconsistent data definitions, incompatible hardware or software, or proprietary technologies that cannot interface with one another. Important terms such as *risk assessment* may also have different meanings for agencies with different

cultures. To a probation officer, a risk assessment may refer to the likelihood of criminal reoffending, whereas to a treatment professional it might refer to the likelihood of leaving treatment prematurely or relapsing to substance use or mental illness. Agencies may be understandably reluctant to share this information out of a concern that it might be misinterpreted.

These barriers, although daunting, are not insurmountable. Evaluators have found effective and ethical ways to access critically important information from government databases without violating legal requirements, ethical mandates or effective clinical practice (Treatment Research Institute, 2012).

Reaching Consensus

The first step to sharing sensitive information is reaching a consensus about what information is available, which data elements may be shared, and under what conditions the data may be shared. Representatives from the interested agencies must come to a mutual understanding about the nature of the data that is collected, the definitions and limitations of the data, and how the data is stored electronically. From there, the agencies must execute data-sharing agreements and other standards that satisfy the legal and ethical concerns of all interested parties.

One resource that may assist agencies to reach this consensus is the National Information Exchange Model (NIEM).[5] The NIEM is considered the standard for information exchange among criminal justice agencies in the United States. It has also been applied internationally and is available to any public, private, or non-profit agency. The NIEM offers basic data-sharing tools, standards and procedures to help agencies share sensitive information. For example, the NIEM can provide samples of data dictionaries containing agreed-upon definitions and recommended formats for criminal justice data. The data dictionaries can help to ensure that information carries a consistent meaning across agencies. The NIEM

5 https://www.niem.gov/aboutniem/Pages/niem.aspx

may also provide sample data-sharing agreements and standards for the transfer of data.

Evaluators and policymakers from OAS member states can use the sample data dictionaries and data-sharing agreements as starting points for reaching agreement about how data might be defined, formatted or shared in their own DTC program evaluations. Legal and ethical mandates will, of course, differ between nations, but the essential principles and practices of effective data-sharing arrangements have common elements that are likely to apply to a wide range of evaluation contexts.

After agencies have reached an agreement about how data should be defined and shared, the NIEM can, if desired, develop computer syntax that allows for the standardization and exchange of the data between MISs. This does not require the agencies to change their own MIS or data-storage methods. Rather, using what is sometimes referred to as *middleware* the NIEM can develop electronic conduits for the transfer of information between databases. This process is likely to be more complicated or costly than is warranted for many DTC pilot programs in OAS member states, but it may become more useful and relevant as dozens of DTCs are developed in multiple nations over time.

Linking Records

It is essential for records to correspond to the same individual across different MISs or databases. Ideally, unique identifying information should be used to link records across agencies or systems. For example, some countries such as Barbados have a national registry number that corresponds to each citizen.

Alternatively, identifying numbers may be assigned to individuals when they contact the criminal justice system. Some of these numbers may correspond to the person, and thus the same number may be assigned to different cases or offenses involving the same individual. Other numbers may correspond to the case, and thus the same number would follow

the person from arrest through sentencing on the same offense but would not be assigned to subsequent offenses. Depending on the country, these identifying numbers might be useful for linking cases across agencies within the criminal justice system, such as linking cases between the court and probation department. Often, however, they cannot be used to link criminal justice records with records from other treatment or social service agencies.

If unique identifying information is not available, there are alternative procedures that evaluators can use to link records with varying degrees of certainty that the records represent the same individual. There are several *probabilistic record linkage* methods which are available freely in the public domain, such as "Link Plus" or "The Link King" (Campbell et al., 2008). These probabilistic methods link cases from different datasets using a combination of available data elements, such as birthdates and first and last names. The links can be accomplished to a specified degree of confidence; for example, with 80 percent confidence that the cases represent the same person. The level of confidence that can be placed in the links is based on such factors as the number of persons in the database and the number of variables that are available to create the matches.

Of course, there will always be some degree of error when using probabilistic record linkages. However, the magnitude of the error should be distributed evenly between the DTC and comparison groups. Therefore, it should not seriously bias the question of which group had lower recidivism rates. However, estimations of the magnitude of the effect of the DTC on recidivism could be off by several percentage points.

If all else fails, bear in mind that even the most restrictive laws and ethical mandates will usually permit de-identified data to be shared or transferred. Many ethics review committees permit evaluators to assign anonymous research identification numbers to cases. After the cases have been linked to records from administrative databases, the evaluator agrees to destroy the lists connecting the subjects' names to their identification numbers. Alternatively, the agency that is storing the administrative data, such as the police or probation department, might be responsible for destroying the links to subjects' names after the data has been transferred. The

data may then be analyzed at the group level in terms of whether a particular case belongs to the DTC or comparison group but cannot be traced back to an individual participant.

Transferring Data

Given the serious resource challenges that confront most countries, data-sharing procedures must start with relatively simple and low-cost methods which are readily available for most programs. For example, *Direct Secure Messaging* (DSM) allows programs to transmit encrypted information to known and trusted recipients over the Internet. Similarly, "cloud" technology may be attractive to small and under-resourced programs because it does not require MIS departments to have sophisticated hardware, networking capabilities, or maintenance staff. Finally, portable external drives can hold substantial amounts of information and can be used to physically transfer information between locations.

Ethics and Legal Training

All professionals who are involved in the evaluation of a DTC or who handle sensitive data should first complete educational courses on ethical and legal protections in research. Because applicable laws may be different in different nations, evaluators might need to identify training programs within their own countries. Web-based courses on research ethics are available in several countries. For example, the U.S. National Institutes of Health provide web-based trainings on legal and ethical protections for research participants.[6] Comparable web-based programs or written educational materials may be available in other countries as well.

6 http://phrp.nihtraining.com/users/login.php

CLOSING COMMENTS

More research has been published globally on the effects of DTCs than on virtually all other criminal justice and substance use treatment programs combined. The results of this intense level of scrutiny have been almost universally positive. In an era of global economic uncertainty and shrinking public dollars, DTCs are being funded and expanded at a steadily increasing rate. There are now more than 3,000 DTCs and other treatment courts in the United States, and new DTC programs are being developed or planned in more than 30 other countries.

The only way to explain this success is to note that the research has paid off. By proving that DTCs reduce crime, save lives, and save money, evaluators have justified the programs' worth to a skeptical audience of policymakers and the public at large. Political and social beliefs may change with the times and pet projects may fade in and out of favor, but scientific proof is ageless. Convincingly worded arguments and strongly felt emotions may sway public opinion temporarily, but they have a difficult time standing up to scientifically reliable and valid contradictory evidence. In the final analysis, the facts have a way of winning out.

It would be nice to say that DTCs planned these events from the beginning, but they did not. Many of the earliest studies in DTCs violated fundamental principles of good science. For example, they examined outcomes only for DTC graduates or selected unfairly disadvantaged comparison samples. It took nearly two decades for the DTC field to get it right and conduct the proper types of studies that were required to prove that DTCs work.

There is no need for other countries to relive these time-consuming lessons that DTC pioneers had to learn through painstaking effort. Other OAS member states can benefit from more than a quarter century of experience in Australia, Canada, the United States and a few other nations to conduct proper research from the outset.

This is not to suggest that other countries should develop and administer their DTCs in the same manner as Australia, Canada or the United States. Different cultures will find different ways to meet the needs of their own citizens. But the laws of science are universal. There are correct ways and incorrect ways to evaluate a program. Performing flawed evaluations does not merely waste time, resources, and opportunities. It can also result in lost credibility, by making DTCs seem as if they are attempting to cut corners or manipulate the truth. It is not enough to have a worthy concept if one cannot prove its worth in a fair and impartial test.

To professionals working in DTC programs, remember this: It is easy to be a good salesperson if you have a superior product. And DTCs are, indeed, a superior product. Do not hesitate to study them carefully and follow the results wherever they may lead you. The facts are on your side and good scientific methodology will enable you to prove it.

REFERENCES

Andrews, D. A., & Bonta, J. (2010). *The Psychology of Criminal Conduct* (5th ed.) New Providence, NJ: LexisNexis.

Ardito, R. B., & Rabellino, D. (2011). Therapeutic alliance and outcome of psychotherapy: Historical excursus, measurements, and prospects for research. *Frontiers in Psychology, 2*, 11.

Banks, D., & Gottfredson, D. C. (2003). The effects of drug treatment and supervision on time to rearrest among drug treatment court participants. *Journal of Drug Issues,* 33, 385-412.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research – Conceptual, strategic, and statistical considerations. *Journal of Personality & Social Psychology*, *51*, 1173–1182.

Belenko, S., Patapis, N., & French, M. T. (2005). *Economic benefits of drug treatment: A critical review of the evidence for policy makers.* Missouri Foundation for Health, National Rural Alcohol & Drug Abuse Network.

Bhati, A. S., Roman, J. K., & Chalfin, A. (2008). *To treat or not to treat: Evidence on the prospects of expanding treatment to drug-involved offenders.* Washington, DC: The Urban Institute.

Bonomo, E. (2012). *Sanctionable behavior in a felony level drug court: Categorizing noncompliant behavior through a criminal-thinking lens* [Master's thesis]. Atlanta: Georgia State University, Dept. of Criminal Justice.

Bourgon, G., & Armstrong, B. (2005). Transferring the principles of effective treatment into a "real world" prison setting. *Criminal Justice & Behavior, 32*, 3-25.

Breckenridge, J. F., Winfree, L. T., Maupin, J. R., & Clason, D. L. (2000). Drunk drivers, DWI "drug court" treatment, and recidivism: Who fails? *Justice Research & Policy, 2*, 87-105.

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

Burke, K., & Leben, S. (2007). Procedural fairness: A key ingredient in public satisfaction. *Court Review, 44*, 4-25.

Campbell, K., Deck, D., & Krupski, A. (2008). Record linkage software in the public domain: A comparison of Link Plus, The Link King, and a "basic" deterministic algorithm. *Health Informatics Journal, 14*(1), 5-15.

Carey, S. M., & Finigan, M. W. (2004). A detailed cost analysis in a mature drug court setting: a cost-benefit evaluation of the Multnomah County Drug Court. Journal of

Contemporary Criminal Justice, 20(3) 292-338.

Carey, S. M., Finigan, M. W., Crumpton, C. D., & Waller, M. S. (2006). California Drug Courts: Outcomes, Costs and Promising Practices: An Overview of Phase II in a Statewide Study. Journal of Psychoactive Drugs. SARC supplement 3.

Carey, S. M., Mackin, J. R., & Finigan, M. W. (2012). What works? The 10 key components of Drug Court: Research-based best practices. *Drug Court Review*, *8*, 6-42.

Chandler, R. K., Fletcher, B. W., & Volkow, N. D. (2009). Treating drug abuse and addiction in the criminal justice system: Improving public health and safety. *Journal of the American Medical Association, 301,* 183-190.

Cooper, C. S. (2004). *Costs associated with the birth of drug and/ or alcohol addicted/exposed infants*. Washington DC: Bureau of Justice Assistance, Drug Court Clearinghouse & Technical Assistance Center at American University.

Cooper, C. S., Franklin, B., & Mease, T. (2010, April). *Establishing drug treatment courts: Strategies, experiences and preliminary outcomes*. Washington DC: Justice Programs Office, School of Public Affairs, American University. *Available at* http://www.cicad.oas.org/fortalecimiento_institucional/dtca/files/Establishing_DTC_%20Strategies_Experiences_Preliminary_Outcomes_volume%201.pdf.

Cresswell, L. S., & Deschenes, E. P. (2001). Minority and non-minority perceptions of drug court program severity and effectiveness. *Journal of Drug Issues, 31*, 259-291.

Crumpton, D., Carey, S. M., & Finigan, M. W. (2004). Enhancing Cost Analysis of Drug Courts: The Transactional and Institutional Cost Analysis Approach. NPC Research: Portland, OR

Dane, C. B. (2012). *Examining the relationship between procedural justice and recidivism in a jail-based residential substance abuse treatment program* [Master's thesis]. Georgia State University, Andrew Young School of Policy Studies.

Dannerbeck, A., Harris, G., Sundet, P., & Lloyd, K. (2006). Understanding and responding to racial differences in drug court outcomes. *Journal of Ethnicity in Substance Abuse* 5, 1-22.

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

Deschenes. E. P., Ireland, C., & Kleinpeter, C. B. (2009). Enhancing drug court success. *Journal of Offender Rehabilitation, 48*, 19-36.

Edwards, S. J. L., Lilford, R. J., & Hewison, J. (1998). The ethics of randomised controlled trials from the perspectives of patients, the public, and healthcare professionals. *British Medical Journal, 317*, 1209-1212.

Farole, D. J., & Cissner, A. B. (2007). Seeing eye to eye: Participant and staff perspectives on drug courts. In G. Berman, M. Rempel & R. V. Wolf (Eds.), *Documenting Results: Research on Problem-Solving Justice* (pp. 51-73). New York: Center for Court Innovation.

Festinger, D. S., Marlowe, D. B., Lee, P. A., Kirby, K. C., Bovasso, G., & McLellan, A. T. (2002). Status hearings in drug court: When more is less and less is more. *Drug & Alcohol Dependence, 68*, 151-157.

Finigan, M. W. (2009). Understanding racial disparities in Drug Courts. *Drug Court Review, 6*, 135-142.

Flango, V. E., & Cheesman, F. L. (2009). Effectiveness of the SCRAM alcohol monitoring device: A preliminary test. *Drug Court Review, 6*, 109-134.

Frazer, M. S. (2006). *The impact of the community court model on defendant perceptions of fairness: A case study at the Red Hook Community Justice Center*. New York: Center for Court Innovation. Available at http://www.courtinnovation.org/sites/default/files/Procedural_Fairness.pdf.

Fundación Paz Ciudadana, Ministerio de Justicia de Chile and Fiscalía de Chile (2014). *Tribunales de Tratamiento de Drogas Compendio estadístico 2010, 2011 y 2012*. [ebook] Santiago de Chile. Available at: http://www.pazciudadana.cl/wp-content/uploads/2015/03/ttd-analisis-estadistico.pdf [Accessed 16 Apr. 2018].

Fundación Paz Ciudadana/Ulda Omar Figueroa Ossa (2018). *Tribunales de Tratamiento de Drogas: Evaluación de su impacto en la disminución de la reincidencia delictual en Chile.*

Gallagher, J. R. (2012). *Evaluating Drug Court effectiveness and exploring racial disparities in Drug Court outcomes: A mixed methods study* [Doctoral dissertation]. Arlington: Univ. of Texas.

Gendreau, P. (1996). The principles of effective intervention with offenders. In A. Harland (Ed.), *Choosing correctional options that work* (pp. 117-130). Thousand Oaks, CA: Sage.

Gibbs, B. R., & Wakefield, W. (2014). The efficacy of enhanced alcohol-use monitoring: An experiment examining the effects of an increased supervision protocol in an adult drug court. *Drug Court Review, 9*(1), 1-22.

Goldkamp, J. S., White, M. D., & Robinson, J. B. (2002). An honest chance: Perspectives on drug courts. *Federal Sentencing Reporter, 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, 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, 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, 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,* 171-196.

Gutierrez, L., & Bourgon, G. (2012). Drug treatment courts: A quantitative review of study and treatment quality. *Justice Research & Policy, 14*, 47-77.

Harrell, A., Cavanagh, S., & Roman, J. (1999). *Final report: Findings from the Evaluation of the D.C. Superior Court Drug Intervention Program.* Washington, DC: The Urban Institute.

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

Hawken, A., & Kleiman, M. (2009). Managing drug involved probationers with swift and certain sanctions: Evaluating Hawaii's HOPE [NCJRS no. 229023]. Washington DC: National Institute of Justice. Available at http://www.ncjrs.gov/pdffiles1/nij/grants/229023.pdf.

Heck, C. (2006). *Local drug court research: Navigating performance measures and process evaluations* [Monograph Series No. 6]. Alexandria, VA: National Drug Court Institute.

Henrichson, C., & Galgano, S. (2013). A Guide to Calculating Justice-System Marginal Costs. New York Vera Institute of Justice. https://www.bja.gov/Publications/Vera_Calculating-Justice-System-Marginal-Costs.pdf

Holloway, K. R., Bennett, T. H., & Farrington, D. P. (2006). The effectiveness of drug treatment programs in reducing criminal behavior. *Psicothema*, *18*, 620-629.

Huddleston, W., & Marlowe, D. B. (2011). *Painting the current picture: A national report on drug courts and other problem solving court programs in the United States.* Alexandria, VA: National Drug Court Institute.

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*, 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*, 88-107.

Jones, C. (2011). Intensive judicial supervision and drug court outcomes: Interim findings from a randomized controlled trial. *Crime & Justice Bulletin, 152*, 1-16. New South Wales, Australia: Bureau of Crime Statistics & Research. *Available at* http://www.courtwise.nsw.gov.au/lawlink/bocsar/ll_bocsar.nsf/vwFiles/CJB152.pdf/$file/CJB152.pdf.

King, M. S. (2009). *The solution-focused judging bench book*. Melbourne, Australia: Australasian Institute of Judicial Administration & Legal Services Board of Victoria.

Kirk, J. M. (2012). *Assessing self determining attitudes and behaviors in court mandated treatment clients* [Doctoral dissertation]. Oklahoma State University.

Kleinpeter, C. B., Brocatao, J., & Koob, J. J. (2010). Does drug testing deter drug court participants from using drugs or alcohol? *Journal of Offender Rehabilitation, 49*, 434-444.

Landenberger, N. A., & Lipsey, N. W. (2005). The positive effects of cognitive-behavioral program for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology*, 1(4), 451-476.

Latimer, J., Dowden, C., & Muise, D. (2005). The effectiveness of restorative justice practices: A meta-analysis. *The Prison Journal, 85*, 127-144.

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

Leukefeld, C., Webster, J. M., Staton-Tindall, M., & Duvall, J. (2007). Employment and work among drug court clients: 12-month outcomes. *Substance Use & Misuse, 42*, 1109-1126.

Lindquist, C. H., Krebs, C. P., & Lattimore, P. K. (2006). Sanctions and rewards in drug court programs: Implementation, perceived efficacy, and decision making. *Journal of Drug Issues, 36*, 119-146.

Lowenkamp, C. T., Flores, A. W., Holsinger, A. M., Makarios, M. D., & Latessa, E. J. (2010). Intensive supervision programs: Does program philosophy and the principles of effective intervention matter? *Journal of Criminal Justice, 38*, 368-375.

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.

Lowenkamp, C. T., & Latessa, E. J. (2004). Understanding the risk principle: How and why correctional interventions can harm low-risk offenders. *Topics in Community Corrections-2004,* 3–8.

Lowenkamp, C. T., Latessa, E. J., & Smith, P. (2006). Does correctional program quality really matter? The impact of adhering to the principles of effective intervention. Criminology & Public Policy, 5, 575-594.

MacDonald, J. M., Morral, A. R., Raymond, B., & Eibner, C. (2007). The efficacy of the Rio Hondo DUI Court: A 2-year field experiment. *Evaluation Review, 31,* 4-23.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Erlbaum.

Marlowe, D. B. (2008). Application of sanctions. In C. Hardin & J. N. Kushner (Eds.), *Quality improvement for drug courts: Evidence-based practices* (pp. 107-114) [Monograph Series No. 9]. Alexandria, VA: National Drug Court Institute.

Marlowe, D. B. (2011). Applying incentives and sanctions. In D. B. Marlowe & W. G. Meyer (Eds.), *The drug court judicial benchbook* (pp. 139-157). Alexandria, VA: National Drug Court Institute. Available at http://www.ndci.org/sites/default/files/nadcp/14146_NDCI_Benchbook_v6.pdf.

Marlowe, D. B. (2012a) (Ed.). Special Issue on best practices in Drug Courts. *Drug Court Review, 8 (1)*.

Marlowe, D. B. (2012b). *Targeting the right participants for adult drug courts* [Practitioner Fact Sheet Vol. VII, No. 1]. Alexandria, VA: National Drug Court Institute. At http://www.ndci.org/sites/default/files/nadcp/Targeting_Part_I.pdf.

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

Marlowe, D. B., Festinger, D. S., Arabia, P. L., Dugosh, K. L., Benasutti, K. M., Croft, J. R., & McKay, J. R. (2008b). Adaptive interventions in drug court: A pilot experiment. *Criminal Justice Review, 33*, 343-360.

Marlowe, D. B., Festinger, D. S., Dugosh, K. L., Arabia, P. L., & Kirby, K. C. (2008a). An effectiveness trial of contingency management in a felony pre-adjudication drug court. *Journal of Applied Behavior Analysis, 41*, 565-577.

Marlowe, D. B., Festinger, D. S., Dugosh, K. L., Lee, P. A., & Benasutti, K. M. (2007). Adapting judicial supervision to the risk level of drug offenders: Discharge and six-month outcomes from a prospective matching study. *Drug & Alcohol Dependence, 88S*, 4-13.

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*, 183-198.

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

Marlowe, D. B., Hardin, C. D., & Fox, C. L. (2016). *Painting the current picture: A national report on drug courts and other problem-solving courts in the United States.* Alexandria, VA: National Drug Court Institute. Retrieved from http://www.ndci.org/wp-content/uploads/2016/05/Painting-the-Current-Picture-2016.pdf

Marlowe, D. B., & Kirby, K. C. (1999). Effective use of sanctions in drug courts: Lessons from behavioral research. *National Drug Court Institute Review, 2,* 1-31.

Marlowe, D. B., Patapis, N. S., & DeMatteo, D. S. (2003). Amenability to treatment of drug offenders. *Federal Probation*, *67,* 40-46.

Martin, S. S., Butzin, C. A., Saum, S. A., & Inciardi, J. A. (1999). Three-year outcomes of therapeutic community treatment for drug-involved offenders in Delaware. *Prison Journal, 79,* 294-320.

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

McKee, M. (2010). *San Francisco drug court transitional housing program outcome study*, *at* http://www.sfsuperiorcourt.org/sites/default/files/pdfs/2676%20Outcome%20on%20SF%20Drug%20Court%20Transitional%20Housing%20Program.pdf.

McLellan, A. T., Lewis, D. C., O'Brien, C. P., & Kleber, H. D. (2000). Drug dependence, a chronic medical illness: Implications for treatment, insurance, and outcomes evaluation. *JAMA, 284*, 1689-1695.

Mendoza, N. S., Trinidad, J. R., Nochajski, T. H., & Farrell, M. C. (2013). Symptoms of depression and successful drug court completion. *Community Mental Health Journal*, DOI: 10.1007/s10597-013-9595-5.

Messina, N., Calhoun, S., & Warda, U. (2012). Gender-responsive drug court treatment: A randomized controlled trial. *Criminal Justice & Behavior, 39*, 1539-1558.

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 non-traditional drug courts. *Journal of Criminal Justice, 40*, 60-71.

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

National Institute of Justice. (2018, August 23). Drug Courts. Retrieved December 20, 2018, from https://www.nij.gov/topics/courts/drug-courts/Pages/welcome.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.

Newton-Taylor, B., Patra, J., & Gliksman, L. (2009). Toronto Drug Treatment Court: Participant intake characteristics as predictors of "successful" program completion. *Journal of Drug Issues, Fall*, 965-987.

O'Hear, M. M. (2009). Rethinking drug courts: Restorative justice as a response to racial injustice. *Stanford Law & Policy Review, 20*, 101-137.

Organization of American States, Inter-American Drug Abuse Control Commission. (2010). *Hemispheric drug strategy: Plan of action 2011-2015. Available at* http://www.cicad.oas.org/Main/Template.asp?File=/main/about-cicad/basicdocuments/strategy_2010_eng.asp.

Organization of American States, Inter-American Drug Abuse Control

Commission. (2010b). *Introduction to the New Hemispheric Drug Strategy. Available at* http://www.cicad.oas.org/en/basicdocuments/Hemispheric%20Drug%20Strategy100603.pdf..

Organization of American States, Inter-American Drug Abuse Control Commission. (2012b). *Report on the first intensive technical workshop on the development of methodologies and indicators for DTC evaluation.* Washington DC: Author.

Organization of American States, Observatory on Citizen Security. (2011). *Report on citizen security in the Americas*. *Available at* http://www.oas.org/dsp/alertamerica/Alertamerica_2011.pdf.

Organization of American States, Observatory on Citizen Security. (2012a). *Report on citizen security in the Americas*. *Available at* http://www.oas.org/dsp/alertamerica/Report/Alertamerica2012.pdf.

Peters, R. H. (1996). *Evaluating drug court programs: An overview of issues and alternative strategies.* Washington DC: Justice Programs Office at American University. Available at http://www1.spa.american.edu/justice/documents/220.pdf

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

Peters, R. H., Kremling, J., Bekman, N. M., & Caudy, M. S. (2012). Co-occurring disorders in treatment-based courts: Results of a national survey. *Behavioral Sciences & the Law*, DOI: 10.1002/bsl.2024.

President's Commission on Combating Drug Addiction and the Opioid Crisis. (2017). *Final Report to the President* [PDF file]. Retrieved from https://www.whitehouse.gov/sites/whitehouse.gov/files/images/Final_Report_Draft_11-1-2017.pdf.

Reingle, J. M., Striley, C. W., Small, E., Crecelius, R., O'Leary, C. C., & Cottler, L. B. (2012). Can courtroom behavior predict recidivism? An assessment of the Courtroom Behavior Check List for women presenting in drug court. *American Journal of Criminal Justice*, DOI: 10.1007/s12103-012-9190-2.

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, NY: Center for Court Innovation.

Rempel, M., Raine, V., & Lambson, S H. (20142014). *A diagnostic study of the Addiction Treatment Court in Guadalupe, Nuevo León, Mexico: Findings and recommendations*. Washington DC: Inter-American Drug Abuse Control Commission, Organization of American States, and Center for Court Innovation.

Richardson, M., Katsakou, C., Torres-Gonzalez, F., Onchev, G., Kallert, T., & Priebe, S. (2011). Factorial validity and measurement equivalence of the Client Assessment of Treatment

Scale for psychiatric inpatient care – A study in three European countries. *Psychiatry Research, 188*, 156-160.

Roman, J. (2013). Cost-benefit analysis of criminal justice reforms. NIJ Journal, 272. https://ncjrs.gov/pdffiles1/nij/241929.pdf

Rossman, S. B., Rempel, M., Roman, J. K., Zweig, J. M., Lindquist, C. H., Green, M., Downey, P. M., Yahner, J., Bhati, A. S., & Farole, D. J. (2011). *The Multi-Site Adult Drug Court Evaluation: The impact of Drug Courts, volume 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. (2008, July). *Performance measurement of drug courts: The state of the art* [State Technical Assistance Bulletin Vol. 6]. Williamsburg, VA: National Center for State Courts.

Rysavy, P., Cunningham, T., O'Reilly-Martinez, R. (2011). Preliminary analysis of the Northern Territory's illicit drug court diversion program highlights the need to examine lower program completion rates for indigenous clients. *Drug & Alcohol Review, 30*, 671-676. DOI: 10.1111/j.1465-3362.2010.00275.x

Satel, S. (1998). Observational study of courtroom dynamics in selected drug courts. *National Drug Court Institute Review, 1*, 43-72.

Sevigny, E. L., Pollack, H. A., & Reuter, P. (2013). Can drug courts help to reduce prison and jail populations? *Annals of the American Academy of Political & Social Science, 647*, 190-212.

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

Somers, J. M., Currie, L., Moniruzzaman, A., Eiboff, F., & Patterson, M. (2011). Drug Treatment Court of Vancouver (DTCV): An empirical evaluation of recidivism. *Addiction Research & Therapy, 2*, 1-7. Available at http://dx.doi.org/10.4172/2155-6105.1000117.

Somers, J. M., Rezansoff, S. N., & Moniruzzaman, A. (2012). *Mediators of drug treatment court outcomes in Vancouver Canada*. British Columbia, Canada: Simon Fraser Univ.

Sperber, K. M., Latessa, E., & Makarios, M. D. (2013). Examining the interaction between level of risk and dosage of treatment. *Criminal Justice & Behavior, 40*, 338-348.

Sullivan, D., & Tifft, L. (2008). *Handbook of restorative justice*. New York: Routledge.

Summers, R. E., & Barber, J. P. (2003). Therapeutic alliance as a measurable psychotherapy skill. *Academic Psychiatry, 27*, 160-165.

Taxman, F. S., & Bouffard, J. A. (2005). Treatment as a part of drug court: The impact on graduation rates. *Journal of Offender Rehabilitation, 42*, 23-50.

Taxman, F. S., & Marlowe, D. B. (Eds.) (2006). Risk, needs, responsivity: In action or inaction? [Special Issue]. *Crime & Delinquency, 52(1)*.

Treatment Research Institute. (2012). *Increasing effective communication between criminal justice and treatment settings using health information technology*. Philadelphia, PA: Author.

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*, 61-85.

Tyler, T. R. (2007). Procedural justice and the courts. *Court Review, 44*, 26.

U.S. Government Accountability Office. (2005). *Adult drug courts: Evidence indicates recidivism reductions and mixed results for other outcomes* [No. GAO-05-219]. Washington, DC: Author.

Van Wormer, J. (2010). *Understanding operational dynamics of drug courts* [Doctoral dissertation]. Dept. of Criminal Justice, University of Washington.

Wiener, R., Winick, B. J., Georges, L. S., & Castro, A. (2010). A testable theory of problem-solving courts: Avoiding past empirical and legal failures. *International Journal of Law & Psychiatry, 33*, 417-427.

Wodahl, E. J., Garland, B., Culhane, S. E., & McCarty, W. P. (2011). Utilizing behavioral interventions to improve supervision outcomes in community-based corrections. *Criminal Justice & Behavior, 38*, 386-405.

Wolfer, L. (2006). Graduates speak: A qualitative exploration of drug court graduates' views of the strengths and weaknesses of the program. *Contemporary Drug Problems, 33*, 303-320.

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

Zweig, J. M., Lindquist, C., Downey, P. M., Roman, J., & Rossman, S. B. (2012). Drug court policies and practices: How program implementation affects offender substance use and criminal behavior outcomes. *Drug Court Review*, 8, 43-79.

GLOSSARY OF TECHNICAL TERMS

Addiction
A psychiatric diagnosis reflecting compulsive use of alcohol or other drugs. Characteristic symptoms include cravings for the substance, uncontrolled usage, or uncomfortable withdrawal symptoms when levels of the substance decline in the bloodstream or central nervous system. In the most recent version of Diagnostic and Statistical Manual of Mental Disorders (DSM-5 or DSM-V), it is referred to as "substance use disorder." See also: **Substance dependence**.

Adulterated test specimens
Bodily specimens such as urine or saliva that show evidence of having been tampered with or of being fraudulent or unreliable. Most laboratories routinely evaluate drug test specimens for evidence of adulteration by examining such indicators as temperature, pH, creatinine, and specific gravity. Many DTCs assume adulterated specimens to be substance-positive or require that a new specimen be delivered. Evaluators should ordinarily treat adulterated specimens as substance-positive or as not having been delivered as directed (i.e., as an unexcused failure to provide a scheduled sample).

Attendance rate
The ratio of the number of sessions or services that participants received, divided by the number of sessions or services they were scheduled to receive.

Best practices
Specific services or interventions in DTCs that have been demonstrated through empirical research to significantly improve outcomes.

Clinical assessment tools
Standardized and validated questionnaires or interviews that assess the diagnostic criteria for or symptoms of substance dependence and other major psychiatric disorders.

Cohort
A group of participants who entered the DTC during the same specified time period, often defined as an interval of 12 months. For example, all participants who entered a DTC between January 1 and December 31 of a given year might be defined as a cohort.

Community service
A type of restorative justice intervention requiring offenders to work on a volunteer basis in the community. Common examples of community service may include cleaning up litter on the highway or working in a soup kitchen for the homeless.

Comparison group
A sample of individuals who are similar to the DTC participants but did not participate in the DTC. In DTC evaluations, the comparison group is often comprised of drug-involved offenders who were sentenced to probation or who underwent adjudication as usual. The comparison group should be as equivalent as possible to the DTC group with respect to variables that would be expected to affect outcomes. Without a comparison group, it cannot be determined whether the outcomes were affected by the DTC or whether they might have occurred anyway even if the participants had not entered the DTC.

Computer-assisted personal interview (CAPI)
A computerized survey that allows the evaluator to enter items directly into an analyzable database. The computer presents items one at a time; may offer help-menus indicating how to phrase, score, or interpret an item; and skips questions that are logically inapplicable for a participant (e.g., a question concerning pregnancy for a male interviewee).

Computer syntax:
See: **Statistical syntax**.

Correlation
The degree to which two variables are associated with each other. A correlation indicates there is a relationship between the variables, but does not prove causality. For example, more treatment in a DTC might be correlated with better outcomes, but this does not prove that treatment was responsible for the outcomes.

Cost evaluation
A systematic evaluation that attaches monetary values to the results of an impact evaluation to estimate the net financial impacts of a DTC.

Counterfactual hypothesis
An alternative hypothesis to the research hypothesis which predicts that a DTC did not affect outcomes. In most instances, being able to test the counterfactual hypothesis requires the evaluator to use a fair and unbiased comparison sample against which to compare the performance of the DTC participants.

Covariate
A variable that is entered first into a statistical analysis, and whose influence is then factored out to evaluate the effect of the true variable of interest. This procedure helps to rule out confounds or alternative explanations that might have accounted for the effects being observed.

Data & Safety Monitoring Board (DSMB)
A multidisciplinary group of professionals with expertise in research methods, research ethics and statistical analyses who are responsible for overseeing the integrity of data collection and data analyses during randomized studies. Among other duties, the DSMB may perform or oversee interim analyses to determine whether any participants in the study are being disadvantaged or suffering adverse events related to the interventions.

Date-stamping
The practice of connecting services or events to the dates on which they occurred or were supposed to have occurred. For example, indicating that a counseling session was attended on January 1, 2013, is a form of date-stamping. Date-stamping is critically important for measuring many performance indicators in DTC program evaluations. Some computerized data-entry systems automatically date-stamp entries.

Density of services
The amount of treatment or other services that were delivered per unit of time, such as per month or per phase of the program. For example, attendance at five sessions in one month (5 ÷ 1 = 5) reflects a greater density of services than attendance at 15 sessions in six months (15 ÷ 6 = 2.5) even though 15 sessions are more than five sessions.

Descriptive analyses
Statistical information that does not seek to infer cause and effect or attribute the findings to a larger population. For example, calculating the percentage of women who graduated from a DTC or the average num-

ber of treatment sessions that were attended are descriptive analyses.

Direct secure messaging (DSM)
A method for transmitting encrypted information to known and trusted recipients over the Internet.

Distal outcomes
Long-term outcomes often occurring after participants are no longer enrolled in the DTC program.

Dosage or dose
The amount of a service or treatment that participants actually received, as opposed to what they were scheduled to receive. Dosage is typically measured by the number of sessions that were attended or the length of time participants were exposed to the service.

Entry
The specific date on which a participant officially entered a DTC or comparable program and the program obtained legal authority over the individual to order treatment and supervision. Many DTCs and probation programs have a formal entry hearing in court at which the participant pleads or is found guilty (or in some jurisdictions, where the pre-trial/pre-plea criminal proceedings are suspended), may waive certain legal rights, and formally enters the program.

Field visits
Supervision activities conducted by probation officers outside of their offices. Examples may include visits to an offender's home or place of employment.

Follow-up window
The time period during which outcomes are measured. For example, if outcomes are measured for 12 months after each participant's entry into a DTC, then the follow-up window is 12 months in length. In most analyses, follow-up windows should be equivalent or comparable in length for all participants.

Generalizability
The degree to which the same result is found in different programs or among different populations of participants.

Graduation
Successful discharge from a DTC marking completion of the requirements of the program. Outcomes are significantly better for DTCs that have formal criteria for graduation and hold a formal graduation ceremony in court.

High-need offenders
Offenders with relatively severe substance abuse or mental health disorders. High-need offenders have been found to have better outcomes in DTCs as compared to alternative programs such as probation.

High-risk offenders
Offenders with more severe criminal histories or who have a history of failing in treatment or on correctional supervision. These individuals are at a higher risk for re-offending, and tend to have a poor prognosis for success in standard correctional rehabilitation programs and require intensive supervision to succeed in treatment. High-risk addicted offenders have been found to have better outcomes in DTCs as compared to alternative programs such as probation.

Illicit drugs
Drugs that are legally banned and prescription medications that are used for a non-prescribed or non-medically-indicated purpose.

Imputation procedures
Statistical procedures that may be employed to compensate for missing data. Some imputation procedures account for the pattern of results immediately before and after a missing result. Others assume the average or most prevalent result for the participant or the sample to be the likely outcome of a missing score. Expert consultation is usually required to decide whether and how to use these procedures.

Institutional Review Board (IRB)
A multidisciplinary group of professionals and community representatives with knowledge or expertise in research methods, research ethics, and the subject matter of a study. The IRB is responsible for determining whether a study is ethical and safe with regard to such matters as informed consent, confidentiality, and the risk/benefit ratio for participants. The IRB may be empowered to stop a study or require changes to the methods as a condition of pre-approval or annual re-approval.

Intent-to-treat analysis
An analysis that includes data on all individuals who entered the DTC or comparison program regardless of whether they completed, dropped out, or were terminated from the programs. Intent-to-treat analyses should ordinarily be reported as the primary analyses for most DTC program evaluations.

Interaction analyses
Statistical procedures that examine the influence of more than one variable on an outcome and determine how the variables affected the outcome alone and in combination. In DTC evaluations, interaction analyses might be used to determine which types of participants had better outcomes in the DTC as opposed to a comparison condition.

Inter-rater reliability
The degree to which different raters or evaluators assign similar values when measuring a variable.

Investment costs
The additional expenditures required to administer a DTC program, such as the added costs of treatment or frequent court hearings.

Jail
In some jurisdictions, a correctional institution that typically detains individuals on a pre-trial basis or for sentences of less than one year in length.

Log-linear transformation
A mathematical procedure that may compensate for a skewed or non-normal distribution of scores. This procedure has the effect of smoothing out a skewed distribution or making it function as if it were a normal distribution. Expert statistical consultation is often required to employ this advanced statistical procedure appropriately.

Long-term outcome evaluation
A systematic evaluation of participants' performance after they are no longer enrolled in the DTC program. The evaluation typically focuses on post-program outcomes such as new arrest or conviction rates, re-incarceration rates, and employment rates. It may also be referred to as a distal outcome evaluation.

Low-need offenders
Offenders who do not have serious substance abuse or mental health disorders. Low-need offenders do not need to be treated in programs such as DTCs that provide a high level of treatment and rehabilitation.

Low-risk offenders
Offenders with relatively less severe criminal histories who have a good prognosis for success in standard substance abuse treatment or correctional rehabilitation. Such individuals are at a lower risk for re-offending and often do not require intensive programs such as DTCs to have positive outcomes.

Management information system (MIS)
An automated computer system that collects standardized data elements and may run statistical analyses and provide outcome reports. In DTCs, an MIS might collect and analyze data on the services delivered in the DTC and participants' performance.

Matched comparison group
Individuals who are paired with DTC participants based on multiple variables that are known to affect outcomes. For example, an evaluator might match each DTC participant with a probationer who is similar in terms of criminal history, demographic characteristics, and substance abuse problems. The groups should be matched on variables that predict outcomes, and not merely variables that are easy to measure such as basic demographics.

Mediator variable
A variable that affects the relationship between a predictor variable (such as enrollment in a DTC) and an outcome variable (such as recidivism or graduation). The mediator variable can be thought of as an intervening variable that helps to explain the relationship between the predictor variable and outcome variable. Mediator variables are identified through advanced statistical analyses called mediation analyses.

Missing denominator
Refers to a problem commonly encountered in DTC program evaluations in which there is a failure to faithfully record information about events that should have transpired but did not. For example, data might not be recorded on treatment sessions that were scheduled to occur but were not attended. This can complicate the interpretation of findings. It is important for staff members to record information about whether appointments were kept, not kept, excused or rescheduled. See also: **Problem of the missing denominator.**

Moderator variables
Characteristics of participants that differentially predict outcomes in DTCs as compared to alternative programs, such as probation. Moderator variables may be identified through the use of statistical techniques called moderator analyses or interaction analyses. These analyses assist the evaluator to determine which types of participants were helped by a DTC and which participants might not have been helped by the DTC.

Motivation for change
Refers to a body of research finding better outcomes in substance abuse treatment for patients who were intrinsically motivated to improve their lives. Intrinsic motivation for change does not appear to be required when participants first enter a DTC but may become necessary before they are discharged. It may also be referred to as readiness for change.

Neutral discharge
A type of discharge from a DTC that indicates neither failure from the program nor successful completion of the program. It is typically reserved for participants who have serious medical or psychiatric illnesses that the DTC is unable to treat, or who move out of the jurisdiction or enlist in the military with the permission of the court. Partici-

pants who receive a neutral discharge are removed from the equation when calculating most performance indicators for a DTC.

Nonparametric statistical tests

Statistical tests that do not require the scores in the sample to be normally distributed or to have a wide range.

Operationalizing variables

The process of defining variables such as the services provided in a DTC or participants' outcomes in objective, concrete, and measurable terms.

Outcome costs

The expenditures incurred by taxpayers or the government to deal with participants after they completed the DTC program, such as the costs of prosecuting new offenses or incarcerating the participants for new crimes.

Outcome evaluation

A systematic study of how a DTC affected participants' performance on outcomes such as substance use or crime. See also: **Short-term outcome evaluation or Long-term outcome evaluation.**

Outcome savings

The monies repaid to or reclaimed by society as a result of the improved social functioning of DTC participants. For example, DTC participants might pay more taxes, contribute to the financial support of their children, or volunteer to work in charities or social service agencies.

Participant characteristics

Attributes of participants that pre-dated their entry into the DTC or comparison program. These typically include socio-demographic variables such as age, gender, race, and employment status; clinical variables such as participants' primary substances of abuse, psychiatric/substance use diagnoses, and history of substance abuse or mental health treatment; and criminal history variables such as prior arrests, convictions, and incarcerations. **See also: Predictor variables and Risk factors.**

Participant-level performance indicators

Quantifiable measures of how participants performed while they were enrolled in the DTC and after discharge. Examples might include how often participants tested negative for alcohol and other illicit drugs or graduated from the DTC. They may also be referred to as outputs. **See also: Performance indicators.**

Perceived deterrence

Refers to a body of research indicating that offenders have better outcomes when they perceive a direct and rational connection between their own conduct and the imposition of rewards and sanctions by criminal justice authorities. Better perceptions of perceived deterrence have been associated with better outcomes in DTCs.

Performance benchmarks

Specific thresholds or levels of performance that DTCs should strive to achieve. Ideally, performance benchmarks should be based on empirical evidence indicating which practices produced better outcomes in DTCs.

Performance indicators

Performance indicators are quantifiable measures of the services provided in a DTC (called program-level performance indicators or inputs) and how participants performed in the program and after discharge (participant-level performance indicators or outputs). Examples of program-level performance indicators might include how often court hearings were held or how often par-

ticipants attended substance abuse treatment. Examples of participant-level performance indicators might include how often participants tested negative for alcohol and illicit drugs or graduated from the program. See also: **Participant-level performance indicators** and **Program-level performance indicators.**

Policies and procedures manual

A written manual documenting the essential practices, interventions and policies for the DTC. Research indicates DTCs have better outcomes when they (1) have a policies and procedures manual, (2) generally adhere to what is written in the manual, and (3) update the manual on a regular basis.

Post-adjudication DTC

A DTC in which a final conviction has been entered and participants complete the program in lieu of receiving a longer or more severe sentence.

Pre-adjudication DTC

A DTC in which successful graduates can have their guilty plea or conviction withdrawn or vacated and may have the arrest expunged or erased from their criminal record (or in other jurisdictions where a prior guilty plea or conviction is not required for entry, a DTC in which the pre-trial/pre-plea criminal proceedings are suspended, and in which successful graduates have the charges against them dropped).

Predictor variables

Characteristics of participants that pre-dated their entry into the DTC or comparison program and reliably correlate with outcomes.

Prison

In some jurisdictions, a correctional institution that typically incarcerates individuals for sentences of longer than one year.

Probabilistic record linkage

A method for linking cases in different administrative databases when no unique identifying number or variable is available. Cases are linked using a combination of available data elements such birthdates and first and last names. The degree of confidence that can be placed in the links is dependent on such factors as the number of variables that are available to create the linkages, the degree of variability on those variables, and the number of persons that are in the database.

Probation

A criminal sentence requiring an offender to be supervised in the community instead of being detained in a jail or prison.

Probation officer

A criminal justice professional who is primarily responsible for supervising offenders who have been sentenced to probation in the community.

Problem of the missing denominator

See: **Missing denominator**.

Procedural justice or procedural fairness

The phenomenon, demonstrated by a body of research, in which litigants react more favorably to an adverse judgment or punitive sanction if they believe fair procedures were followed in reaching the decision. Greater perceptions of procedural justice have been associated with better outcomes in DTCs.

Process evaluation

A systematic study indicating whether a DTC is functioning as originally planned, treating the intended target population, and delivering the types and dosages of services that are likely to produce favorable outcomes.

Program-level performance indicators
Quantifiable measures of the services that were provided in the DTC. Examples might include how often court hearings were held or how often participants attended substance abuse treatment. These may also be referred to as inputs.

Propensity score analysis
A statistical procedure that may be used to control for differences in participant characteristics between groups. It involves mathematically calculating the probability that an individual would be in the DTC group as opposed to the comparison group, or the relative similarity of the individual to one group as opposed to the other. The analysis statistically accounts for this relative probability when comparing outcomes between the groups. Advanced statistical expertise is often required to implement and interpret the results of this procedure.

Proximal outcomes
Short term outcomes usually occurring while participants are still enrolled in the DTC program. Examples might include counseling attendance or graduation rates.

Quasi-experimental comparison group
A comparison sample of individuals who did not enter the DTC for reasons that are unlikely to have affected their outcomes. A good example might be drug-involved offenders who were eligible for and willing to enter the DTC but could not get in because there were no empty slots available.

Random assignment
A procedure for assigning participants to different groups in an unsystematic and unbiased manner, such as by the flip of the coin. Random assignment provides the greatest assurances that the groups started out with an equal chance of success, and thus that any positive outcomes can be attributed to the effects of the DTC program and not to extraneous factors.

Recidivism
The incidence of new criminal activity occurring after participants entered the DTC or comparison program. This includes criminal activity occurring while participants were enrolled in the program, and after they graduated or were terminated from the program. Recidivism is most commonly measured by the number or percentage of new arrests or new convictions after entry. It may also be referred to as re-offending.

Redundancy or redundant variables
The degree to which predictor variables are correlated with each other and thus do not provide independent prediction of outcomes.

Replication
The process of repeating or reproducing a study to ensure the same results are obtained. Replication increases the likelihood that the findings are valid and reliable.

Restorative justice interventions
Requirements that may be imposed on participants in DTCs to compensate victims or society for their crimes. Examples may include victim restitution, community service, victim impact panels, and payment of fines or fees.

Rewards
Consequences for behavior that are desired by offenders such as verbal praise, applause, small gifts, and reductions in supervision requirements. They may also be referred to as incentives.

Risk factors
Characteristics of participants that pre-dated their entry into the DTC or comparison program and reliably correlate with poorer outcomes. Common examples of risk factors include a younger age, prior failures in treatment or rehabilitation, and a more serious criminal or substance abuse history. Individuals with such risk factors typically require more intensive and structured interventions to succeed in treatment and refrain from criminal activity.

Risk assessment tools
Standardized and validated questionnaires or interviews that measure the likelihood of participants failing in treatment or committing a new offense. Offenders who are identified as being high risk on these tools tend to have better outcomes in DTCs as compared to less intensive dispositions such as probation.

Sanctions
Consequences for behavior that are disliked by offenders such as verbal reprimands, monetary fines, community service, increased supervision requirements, or incarceration. The use of sanctions, and the types of sanctions used, varies from jurisdiction to jurisdiction.

SCRAM®
Secure Continuous Remote Alcohol Monitor. An anklet monitoring device that can detect alcohol in sweat vapors and transmits data wirelessly to a central monitoring station.

Secondary analyses
Data analyses that examine outcomes only for subgroups of participants (e.g., graduates only). Secondary analyses are more likely than intent-to-treat analyses to turn up unreliable or chance findings. Therefore, they should ordinarily be performed only if the intent-to-treat analysis first indicated significant results.

Short-term outcome evaluation
A systematic evaluation of participants' performance while they are enrolled in the DTC program. The evaluation typically focuses on during-treatment outcomes that are likely to predict post-program outcomes such as graduation rates, treatment attendance rates, and rates of drug-negative urine tests. It may also be referred to as a proximal or intermediate outcome evaluation.

Skewed distribution
Refers to scores on a variable that have a restricted range or similar values for a large proportion of participants. Many statistical tests assume that distributions are normal or have a wide range of values, and therefore cannot be used to analyze data with a skewed distribution. The evaluator might need to use a **Nonparametric statistical test** or a **Log-linear transformation** to compensate for data that has a skewed distribution.

Starting clock
The time point from which data collection on performance indicators begins. For DTC program evaluations, the clock should ordinarily be started on the date of entry into the DTC or comparison program.

Statistical imputation procedures
See: **Imputation procedures.**

Statistical significance or statistically significant differences
Differences between groups that have a high mathematical probability (usually 95 percent or higher) of being reproducible in the future. Statistically significant differences permit greater confidence in the reliability of one's findings.

Statistical syntax
Pre-programmed statistical equations instructing how the data should be analyzed. For example, a computer system might have a pre-programmed statistical formula for calculating the average age of participants in a DTC program.

Substance dependence
A psychiatric diagnosis reflecting compulsive use of alcohol or other drugs. Characteristic symptoms include cravings for the substance, uncontrolled usage, or uncomfortable withdrawal symptoms when levels of the substance decline in the bloodstream or central nervous system. See also: **Addiction**.

Survival analysis
A statistical procedure that compares groups of participants on the average length of time until a specified event occurs, such as a new criminal arrest or relapse to drug use.

Target population
The sub-group of drug-involved offenders who perform significantly better in DTCs as compared to alternative programs, and thus should be prioritized for entry into DTCs.

Technical violations
Violations of a court order that do not constitute a crime per se. For example, drinking alcohol is legal for most adults in most countries, but may be a technical violation for a DTC participant and may lead to an arrest or conviction for a probation violation.

Therapeutic alliance
The phenomenon, demonstrated by a body of research, in which patients' perceptions of the quality of the therapist-patient relationship reliably predicts outcomes in psychotherapy. The impact of the therapeutic alliance has not been well studied in DTCs, but has been hypothesized to be an important factor for success in DTC programs. It is also referred to as the working alliance.

Time at liberty
The proportion of time during the follow-up period when participants were relatively free in the community to engage in drug abuse, crime, or other behaviors of interest to evaluators. Restrictions on participants' time at liberty typically include physical barriers such as jail or residential treatment.

Time at risk
The length of time in which participants could have engaged in drug abuse, crime, or other behaviors of interest to the evaluator. Generally speaking, the longer the follow-up period, the longer the time at risk.

Variance
The degree to which participants produced a range of different scores on a measure. For example, if all participants are between the ages of 21 and 23 years, the variance in age is low. For mathematical reasons, it is easier to detect statistically significant differences between groups when the variance on a measure is high.

Victim impact panel
A type of restorative justice intervention that requires offenders to meet with crime victims and learn how they were affected by the crime. The goal is to help offenders develop empathy and gain perspective on how their actions affected other people.

Wait-list comparison group
Drug-involved offenders who were eligible for and willing to enter the DTC but who could not get in because there were no empty slots available. A wait-list comparison sample is often the best alternative after random assignment for a DTC program evaluation.

APPENDIX: RESOURCES FOR DRUG TREATMENT COURTS

Performance Indicators and Performance Measurement

National Center for State Courts (NCSC) & Justice Programs Office at American University Translating Drug Court Research into Practice (R2P)
https://nicic.gov/series/r2p-translating-drug-court-research-practice

Performance Measurement of Drug Courts: The State of the Art
National Center for State Courts (2008)
http://cdm16501.contentdm.oclc.org/cdm/ref/collection/spcts/id/171

Local Drug Court Research: Navigating Performance Measures and Process Evaluations
National Drug Court Institute (2006)
http://www.ndci.org/publications/monograph-series/navigating-performance-measures-and-process-evaluations

Evaluating Drug Court Programs: An Overview of Issues and Alternative Strategies
Justice Programs Office at American University
https://jpo.wrlc.org/handle/11204/3306

Introductory Handbook for DWI Court Program Evaluations.
National Center for DWI Courts (2010)
http://www.dwicourts.org/sites/default/files/nadcp/DWI%20Ct%20Eval%20Manual%20REVISED-8-10.pdf

Recidivism 101: Evaluating the Impact of Your Drug Court
Center for Court Innovation (2005)
http://www.courtinnovation.org/sites/default/files/Recidivism_101%5B2%5D.pdf

Management Information Systems (MIS)

Treatment Research Institute Court Evaluation Program (TRI-CEP)
http://triweb.tresearch.org//?s=tricep

Drug Court Case Management System (DCCM)
Advanced Computer Technologies
http://www.actinnovations.com/Products/DCCM

eCourt System
U.S. National Institute of Drug Abuse

Criminal Justice-Drug Abuse Treatment Studies (CJ-DATS)
http://www.gmuace.org/documents/prod-pub/cjdats/cjdats-summary-ecourt.pdf

Buffalo, NY, Drug Court Case Management System
http://www.ndci.org/contact
lfleming@nadcp.org

Accessing Administrative Databases

National Information Exchange Model (NIEM)
https://www.niem.gov/aboutniem/Pages/niem.aspx.

Staff Surveys of DTC Policies and Practices

U.S. National Institute of Drug Abuse
Criminal Justice-Drug Abuse Treatment Studies (CJ-DATS)
National Drug Court Survey
http://www.gmuace.org/documents/prod-pub/cjdats/cjdats-summary-drug-court.pdf

U.S. National Institute of Justice
Multisite Adult Drug Court Evaluation (MADCE)
https://www.ncjrs.gov/pdffiles1/nij/grants/237110.pdf (Appendix A)

NPC Research, Inc.
Adult Drug Court Typology Interview
http://npcresearch.com/wp-content/uploads/Drug-Court-typology-guide-NPC-Research-01-26-04-copyrighted.pdf

Rating Scales of Correctional Program Quality

Correctional Program Assessment Inventory (CPAI)
http://www.state.in.us/idoc/files/CPAI_overview.ppt

Correctional Program Checklist (CPC)
https://www.uc.edu/content/dam/uc/gencounsel/docs/CPC%20Training%20MOU%2011.12.14.pdf

Risk-Assessment Instruments

Level of Service Inventory-Revised (LSI-R) * † ‡
https://ecom.mhs.com/(S(zhkd5d55qlwc3lr2gzqq5w55))/product.aspx-?gr=saf&prod=lsi-r&id=overview

Risk and Needs Triage (RANT)
http://www.trirant.org/

Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)
https://www.cdcr.ca.gov/rehabilitation/docs/FS_COMPAS_Final_4-15-09.pdf

Ohio Risk Assessment System (ORAS)
https://www.uscourts.gov/sites/default/files/74_1_2_0.pdf

Federal Post Conviction Risk Assessment (PCRA)
http://www.uscourts.gov/FederalCourts/ProbationPretrialServices/Supervision/PCRA.aspx

Risk Prediction Index (RPI)
https://www.fjc.gov/content/328578/fjc-50th-improving-criminal-justice-administration-and-sentencing-federal-courts

Wisconsin Risk and Need Assessment Scale (WRN)
https://csgjusticecenter.org/corrections/projects/wisconsin-state-risk-assessment-validation/

Clinical-Assessment Instruments

Addiction Severity Index (ASI) * † ‡
Treatment Research Institute
http://triweb.tresearch.org/index.php/tools/download-asi-instruments-manuals/

Global Appraisal of Individual Needs (GAIN) *
GAIN Coordinating Center
http://gaincc.org/instruments/

Alcohol Use and Associated Disabilities Interview Schedule (AUDADIS) * †
U.S. National Institute of Alcoholism and Alcohol Abuse
http://www.ncbi.nlm.nih.gov/pubmed/10606491

Drug Abuse Screening Test (DAST) * †
http://www.camh.ca/en/education/about/camh_publications/Pages/drug_abuse_screening_test.aspx

Multisite Adult Drug Court Evaluation (MADCE)
Participant Survey
https://www.ncjrs.gov/pdffiles1/nij/grants/237109.pdf (Appendix A)

Structured Clinical Interview for the DSM-IV (SCID) *
http://www.scid4.org/

Texas Christian University (TCU) Drug Dependence Screen (available in English, Spanish, Russian, Chinese (simplified and traditional), and Vietnamese) *
https://ibr.tcu.edu/forms/tcu-drug-screen/

Procedural Justice / Procedural Fairness Scales

Multisite Adult Drug Court Evaluation (MADCE)
Participant Survey
https://www.ncjrs.gov/pdffiles1/nij/grants/237109.pdf (Appendix A)

Procedural Justice Questionnaire (PJQ)
Treatment Research Institute
http://triweb.tresearch.org/index.php/about-us/contact-us/

Perceived Deterrence Scales

Multisite Adult Drug Court Evaluation (MADCE)
Participant Survey
https://www.ncjrs.gov/pdffiles1/nij/grants/237109.pdf (Appendix A)

Perceived Deterrence Questionnaire (PDQ)
Treatment Research Institute
http://triweb.tresearch.org/index.php/about-us/contact-us/

Motivation / Readiness for Change Scales

Inventario de Procesos de Cambio * †
European Monitoring Centre for Drugs and Drug Addiction
http://www.emcdda.europa.eu/html.cfm/index3662EN.html

Personal Concerns Inventory (PCI)
http://pubs.niaaa.nih.gov/publications/AssessingAlcohol/InstrumentPDFs/47_PCI.pdf

University of Rhode Island Change Assessment (URICA)
https://elcentro.sonhs.miami.edu/research/measures-library/urica/index.html
https://www.ncbi.nlm.nih.gov/books/NBK64976/table/A62309/

Multisite Adult Drug Court Evaluation (MADCE)
Participant Survey
https://www.ncjrs.gov/pdffiles1/nij/grants/237109.pdf (Appendix A)

Circumstances, Motivation, Readiness, and Suitability Scales (CMRS)
http://www.ncbi.nlm.nih.gov/pubmed/7832182

Satisfaction with the Program / Treatment Scales

Treatment Perception Questionnaire (TPQ)
European Monitoring Centre for Drugs and Drug Addiction
http://www.emcdda.europa.eu/html.cfm/index4322EN.html

Client Satisfaction Questionnaire (CSQ)
http://www.csqscales.com

Client Assessment of Treatment Scale (CATS) *
http://www.ncbi.nlm.nih.gov/pubmed/21342706

Therapeutic Alliance Scales

Helping Alliance Questionnaire-II (HAQ-II)
http://www.med.upenn.edu/cpr/instruments.html

Working Alliance Inventory
http://wai.profhorvath.com/

Observer Rating Scales for Court Sessions

Multisite Adult Drug Court Evaluation (MADCE)
https://www.ncjrs.gov/pdffiles1/nij/grants/237109.pdf

On-Site Training and Technical Assistance

Institutional Strengthening and Integral Programs Section
Inter-American Drug Abuse Control Commission (CICAD)
Organization of American States
http://www.cicad.oas.org/Main/Template.asp?File=/fortalecimiento_institucional/default_eng.asp

U.S. National Drug Court Institute (NDCI)
http://www.ndci.org/training/advanced-training

Drug Court Clearinghouse and Technical Assistance Project at American University
https://www.american.edu/spa/jpo/initiatives/drug-court/

On-Line Webinars and Distance Learning Programs

U.S. National Drug Court Institute (NDCI)
http://www.ndci.org/training/online-trainings-webinars

National Drug Court Resource Center (NDCRC)
https://www.ndci.org/resources/training/

Center for Court Innovation
http://drugcourtonline.org/

National Institute of Justice, National Center for State Courts (NCSC) & Justice Programs Office at American University
Translating Drug Court Research into Practice (R2P)
https://www.nij.gov/topics/courts/drug-courts/pages/research2practice.aspx

Evidence-Based Treatments for Substance Abuse

National Registry of Evidence-Based Programs and Practices (NREPP)
U.S. Substance Abuse and Mental Health Services Administration
https://www.samhsa.gov/nrepp

Training on Research Ethics

U.S. National Institutes of Health
Office of Research Integrity
http://phrp.nihtraining.com/users/login.php

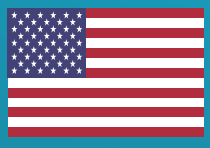
*Spanish translation available.

†Validated among individuals of Hispanic or Latino ethnicity.

‡Validated in at least one South American or Caribbean Nation.

Organization of American States
Secretariat for Multidimensional Security
Inter-American Drug Abuse Control Commission

OAS | CICAD



This publication was prepared by the Executive Secretariat of the Inter-American Drug Abuse Control Commission (ES-CICAD), Secretariat for Multidimensional Security of the Organization of American States (SMS/OAS). It was developed under the aegis of the Drug Treatment Court Program for the Americas, an initiative coordinated by ES-CICAD/SMS/OAS. This publication was made possible through the financial contribution of the Government of the United States of America, through the Bureau of International Narcotics and Law Enforcement Affairs (INL) of the U.S. Department of State, and of the Government of Canada, through the Anti-Crime Capacity Building Program (ACCBP) of Global Affairs Canada.

ISBN 978-0-8270-6830-8