Drug Treatment Courts: A Quantitative Review of Study and Treatment Quality

Leticia Gutierrez
Guy Bourgon
Public Safety Canada

Abstract

The effectiveness of drug courts has been the subject of numerous studies, and three major meta-analyses have examined many of these studies in regard to two main factors: (1) study quality and (2) treatment quality. The current study examines these two factors more closely. Study quality was assessed using the Collaborative Outcome Data Committee Guidelines (CODC); studies were rated as “rejected,” “weak,” “good,” or “strong” based on methodological quality. Drug court treatment quality was assessed by evaluating adherence to the principles of Risk-Need-Responsivity (RNR). The RNR principles have been previously shown to mediate the effectiveness of offender treatment across various offender groups and a variety of criminogenic needs. In total, 96 studies were reviewed and assessed according to study and treatment quality. Results found that the study quality of the literature is poor and that this accounts for much of the variability in findings seen across studies. Furthermore, analyses revealed that although adherence to the RNR principles was poor, increasing adherence to RNR resulted in more effective treatment of offenders and reduced recidivism. Using only methodologically acceptable studies, the least biased estimate of the effectiveness of drug courts in reducing recidivism was found to be approximately 8%.

We would like to thank Karl Hanson for providing assistance with the statistical analyses and methods of the current study and Leslie Helmus for providing training on the Collaborative Outcome Data Committee (CODC) Guidelines and for coding studies for interrater reliability. We would also like to thank Shannon Hodgson, Jan Roehl, and David Wilson, who assisted in providing information (e.g., court evaluations, unpublished reports) that was useful for conducting the present review, as well as Jim Bonta and Tanya Rugge for their feedback and suggestions.

The views expressed are those of the authors and not necessarily those of Public Safety Canada.
The first drug treatment court, which opened in Miami in 1989, came in response to the rising rates of drug-related offenses in the United States. At that time, use of crack cocaine was widespread, and courts were handing down more and longer custodial sentences to substance-abusing offenders. As a result, prison overcrowding became a significant issue, and drug courts (also commonly referred to as drug treatment courts) were created to divert eligible offenders from institutions to judicially supervised treatment in the community. It was believed that these courts, and the associated substance abuse treatment, would assist offenders in overcoming their substance abuse issues and, as a result, reduce recidivism. Since their inception, drug courts have become a popular alternative to incarceration for nonviolent, substance-abusing offenders. Today there are over 3,000 drug treatment courts in the United States, Canada, the United Kingdom, and Australia, with more in the planning stages (National Association of Drug Court Professionals [NADCP], 2011).

Given the popularity of drug courts, a number of researchers have sought to determine whether these courts are effective in reducing recidivism. Three meta-analyses have been conducted to date and all found positive effects. Lowenkamp, Holsinger, and Latessa conducted the first meta-analytic review of drug courts in 2005. Based on weighted effect sizes for 22 studies, they found that drug treatment courts produced an overall reduction in recidivism of 7.5%. The second meta-analysis by Latimer, Morton-Bourgon, andChrétien (2006) reviewed a total of 54 studies and found an overall reduction in recidivism of 12.5%. The third meta-analysis, conducted by Wilson, Mitchell, and Mackenzie (2006), included a total of 50 studies from which they reported an overall reduction in recidivism of 12.3%.

Despite the positive findings of the three meta-analyses regarding the efficacy of drug courts, there is debate in the literature regarding the reliability of these findings due to two potentially moderating factors: study quality and treatment quality.

Study Quality

All three meta-analytic reviews noted the prevalence of problematic study designs among the drug court outcome evaluations. In meta-analytic reviews, as in individual research studies, the quality of the methodology can play a significant role in the interpretation of results (Cook & Campbell, 1979; Farrington, 2002). Therefore, the inclusion of biased studies into a meta-analytic estimation of the effectiveness of programs, such as drug courts, in reducing recidivism can bias the findings.

The introduction of bias to a study can arise from a variety of factors. Random assignment studies are considered to be the “gold standard” of study quality, with other designs considered weaker to varying degrees (Farrington, 1983; Sherman et al., 1997). Although the number of randomized experiments in criminology has increased in recent years, this design is difficult to employ due to a variety of ethical, legal, and practical constraints (Farrington & Welsh, 2005; Sherman et al., 1997).
As a result, weaker quality quasi-experimental designs are more frequently used in criminal justice research. Randomization can control for both known and unknown confounding variables (Farrington, 2002), so that random assignment designs provide stronger evidence that any effects found are more likely attributable to a manipulation (e.g., treatment exposure) rather than preexisting differences between the two groups. In nonrandom designs, researchers must make efforts to reduce and minimize potential sources of bias between the groups that may account for differences in outcomes. This is particularly relevant in quasi-experimental designs, where treatment assignment is often determined by situational (e.g., type of arrest) or personal (e.g., motivation, risk level) characteristics.

In addition to preexisting differences between groups, any and all potential differences in measurement should be considered. For example, bias can be introduced when the reliability and validity of outcome measures are different between the groups (e.g., varying lengths of outcome in a recidivism study). Other factors such as differential attrition rates (i.e., dropouts) and nonblind assignment procedures can also increase bias (Cook & Campbell, 1979; Farrington, 2003).

Confidence in study results can also be influenced by descriptive validity factors (Farrington, 2003). Descriptive validity is the overall amount and quality of description of the various study elements that contribute to an effect size. Sample size, description of intervention, and quality of control variables used in order to determine the effects of an intervention (e.g., risk measures) are examples of descriptive validity indicators. These factors significantly influence the degree of confidence one can place in the results of a study and also contextualize research findings, facilitate replication, and enhance the validity of analytic reviews (Farrington, 2002, 2003).

Due to the structure of drug courts, studies of their effectiveness face some common methodological issues. Inappropriate comparison groups can be a problem given the voluntary nature of drug courts, where clients must “self-select” in order to participate in the program. Secondly, drug courts face high rates of program attrition. Study bias increases as treatment group attrition increases, as there is often little, if any, program attrition in the comparison group. Although program attrition is an issue that affects most studies, it poses a significant threat to evaluations of drug treatment courts due to these programs’ high dropout rates (Cissner & Rempel, 2005; Weekes, Mugford, Bourgon, & Price, 2007). Cissner and Rempel (2005) estimated an average attrition rate of 40% for drug courts. Based on the studies in their meta-analysis, Latimer et al. (2006) reported an average attrition rate of 45.2%, with rates ranging from 9% to 84.4%. Lastly, biased outcome measures are problematic for drug court evaluation studies, as the drug court model is fundamentally different from traditional criminal justice processes and can result in systematic differences between the treatment and comparison groups. For example, for drug court participants, conviction and/or sentencing decisions are delayed until program completion. This can result in some between-group differences on official records where sentencing dates are utilized to identify recidivist events. For a
drug court client, the court internally handles most noncompliant behavior through various sanctions, and noncompliance is often contextualized in terms of treatment behavior and progress rather than criminal outcome. The result can be a systematic bias in outcomes favoring the drug court group over the comparison group.

Given that study quality is important when interpreting a study’s findings, especially in meta-analytic reviews, it is not surprising that a number of assessment tools have been developed to assess study quality for outcome evaluations. Although more commonly used in the medical field (e.g., Deeks et al., 2003; Jüni, Witschi, Bloch, & Matthias, 1999), such assessment tools have recently been used in examinations of criminal justice interventions (Sherman et al., 1997; Hanson, Bourgonn, Helmus, & Hodgson, 2009). The present review used the Collaborative Outcome Data Committee Guidelines (CODC, 2007a, 2007b), which were developed for use in examining the effectiveness of sex offender treatment outcome studies. The CODC Guidelines also help researchers differentiate between biased and less-biased evaluations.

The quality of treatment outcome studies has been shown to play a major role when estimating the effectiveness of intervention programs (Cook & Campbell, 1979; Farrington, 2002). The most reliable results come from studies that limit the amount of bias within their design. As noted above, the drug treatment court literature in particular faces many challenges relating to methodological reliability due to the structure of drug court programs. In fact, study quality issues may have affected the reliability of estimations of drug court effectiveness made to date.

**Treatment Quality**

The second factor that may have an influence on the findings of the previous meta-analytic reviews is the variability of treatment quality among drug courts. Views and definitions of what constitutes effective correctional treatment have evolved since the advent of the “nothing works” literature of the 1970s (Martinson, 1974). A movement toward the use of theoretically coherent and evidence-based practice in offender programming has become the focus of the “what works” movement. Treatment programs that adhere to the principles of Risk, Need, and Responsivity (RNR) have been found to be most effective for a variety of offender types (e.g., violent, sexual, substance abusing, etc.; Andrews, Bonta, & Hoge, 1990; Andrews & Bonta, 2006). The drug court system relies on a range of treatment programs offered in the community, which function at arm’s length from the courts. This program structure makes quality assurance and communication between the courts and treatment programs more difficult. Also, the structure of many of these community-based programs is not geared toward evaluation of program effectiveness; rather, the programs are designed to target substance abusers in general. Therefore, the quality of treatment ranges in terms of adherence to the principles of effective correctional programming (i.e., risk, need, and responsivity).
The Risk-Need-Responsivity model of criminal behavior has been paramount in the development of effective correctional programming. It proposes that an offender’s risk level, criminogenic characteristics, and personal characteristics should dictate the level and type of program services. Adherence to the RNR principles has been shown to produce significant reductions in recidivism (Andrews et al., 1990; Andrews & Bonta, 2006).

**Risk Principle**

The first component of the RNR model is the risk principle, which holds that the risk level of an offender can be predicted and must be matched with the frequency and intensity of the correctional intervention. In other words, high-risk offenders should receive a higher frequency and dosage of treatment, as they have a higher probability of negative outcomes compared to low-risk offenders. Low-risk offenders, on the other hand, should receive little to no treatment (Andrews & Bonta, 2006).

**Need Principle**

The second component of the RNR model addresses the importance of identifying and targeting an offender’s criminogenic needs (dynamic risk factors) rather than noncriminogenic needs (factors weakly related to recidivism) in order to reduce recidivism (Andrews et al., 1990; Andrews & Bonta, 2006). Criminogenic needs are factors that when improved or eliminated are likely to result in a reduction of reoffending. There are seven criminogenic need areas (e.g., antisocial attitudes, employment/education, etc.) that have been identified in the literature as being part of the “Central Eight” correlates of criminal behavior (criminal history, a static risk factor, completes the Central Eight).

**Responsivity Principle**

The last principle of the RNR model deals with the issue of general and specific responsivity. This principle can be interpreted as the “what works for whom” principle (Wormith et al., 2007). Responsivity involves the appropriate matching of treatment programs to an offender’s individual learning style and abilities (Andrews et al., 1990; Andrews & Bonta, 2006). General responsivity simply reflects the belief that cognitive-behavioral interventions work best. Specific responsivity is a treatment matching approach that considers an offender’s personality, gender, ethnicity, motivation, age, language, and interpersonal style (Bonta, 1995). Attending to these factors in correctional settings has been shown to result in treatment success and significant reductions in recidivism (Andrews & Bonta, 2006).

Drug courts make use of many different types of treatment programs, often using numerous providers for different types of services (e.g., Alcoholics Anonymous, acupuncture, positive parenting, etc.). This variation introduces the challenge of ensuring that services are being delivered appropriately and that they are being
matched to each offender’s risk level and individual needs. Given the research to date, assessing program adherence to RNR may clarify the meta-analytic estimates of the overall effectiveness of drug treatment courts in reducing recidivism.

Purpose of the Current Study

The meta-analyses that have been conducted to date assessing the effectiveness of drug courts have yielded positive results. However, the influence of the methodological flaws within the individual evaluations of drug courts, as well as the inconsistency of effective programming, have led researchers to question the validity of these meta-analytic findings. The purpose of the present study is to replicate the previous meta-analyses in order to identify the methodological strengths and weaknesses of the evaluations and assess the influence of study quality on the estimation of drug court effectiveness. Also, the role of treatment quality (i.e., adherence to RNR) and its influence on the meta-analytic estimations of drug court efficacy will be examined. Lastly, a least-biased estimate of the effectiveness of drug treatment courts will be provided by examining only those studies deemed minimally biased, as determined by their rating on the CODC Guidelines.

Method

Sample

Since one of the primary purposes of the present investigation was to replicate the previous meta-analytic reviews and examine the effects of study and treatment quality, only those studies that were included in the three previous reviews (Latimer et al., 2006; Lowenkamp et al., 2005; Wilson et al., 2006) were included in the present study. Studies were obtained via the Public Safety Canada library, the Internet (e.g., research institute or evaluation company Web sites), and directly from the authors or drug treatment courts via email, fax, and/or mail. Although efforts were made to obtain all 102 studies used in the original meta-analyses, four studies could not be located. Additionally, two of the studies were collected but excluded from the present investigation, as they did not contain information on a comparison group. Seven studies did not contain sufficient treatment group information; therefore, no effect sizes could be calculated for those studies. Consequently, the present review examined 96 studies/reports, which represent a total of 103 distinct drug treatment courts (some reports included outcomes for more than one court) and a sample of 50,640 offenders.

Description of Measures

The Collaborative Outcome Data Committee Guidelines (CODC, 2007a; 2007b). The CODC Guidelines are a comprehensive scale developed for the purpose of rating the study quality of sex offender research. In 1997, leading experts
in this field formed the Collaborative Outcome Data Committee and developed the CODC Guidelines to facilitate the assessment of study quality of sex offender treatment outcome research in order to reduce bias in systematic reviews. The CODC Guidelines postulate that study quality is a combination of the confidence one can place in the results of an evaluation and the amount of bias inherent in the study design. The Guidelines have been used for meta-analytic reviews of sex offender treatment (Hanson et al., 2009; Helmus, 2008) and community supervision (Simpson, 2008) and have been found to reliably assess treatment outcome research (CODC, 2007a, 2007b; Helmus, 2008).

The CODC Guidelines contain 20 items (an additional 21st item is specific to cross-institutional designs only), with nine items assessing confidence and 11 items assessing the amount and direction of bias present in an evaluation. Confidence items are rated on a scale of “little confidence” (0) to “high confidence” (2). Items assessing amount of bias are rated from “considerable bias” (0) to “minimal/negligible bias” (2). The direction of bias is rated as “bias increases magnitude of treatment” (1), “no or minimal bias expected” (0), “bias decreases magnitude of treatment” (-1), or “cannot assess direction of bias” (99). When information is not available to appropriately rate an item, the item is coded as having “insufficient information to evaluate.”

Upon rating of the 20 items, each study is given a global rating for confidence, bias, and direction of bias. Based on the global ratings, studies are divided into overall study quality groups consisting of: rejected, weak, good, or strong. Rejected studies are ones that produce low confidence and/or contain considerable amounts of bias, which are likely to influence the treatment outcome findings. Weak studies are those that produce some confidence and contain little bias. Although these studies may possess significant flaws, they provide useful treatment outcome knowledge that is relatively reliable. Good studies produce high confidence and contain little bias, as they make strong efforts to limit any confounds to study validity. Lastly, strong studies produce high confidence and contain minimal bias, as they possess benign problems that are unlikely to influence study results.

Minor modifications were made to the CODC Guidelines to account for the differences between evaluations of general offender programs rather than sex offender programs. For the purposes of the present study, three items were modified—defining treatment, adequacy of search for preexisting differences, and confidence in length of follow-up. The item “defining treatment” was divided into two items, requiring the coder to assess the quality of both the definition of the treatment as well as the definition of the court. The item assessing the “adequacy of search for preexisting differences” was modified to require a comparison of groups on factors above and beyond the presence of substance abuse. Finally, the item assessing “confidence in length of follow-up” was modified from the original guidelines to allow a shorter minimum required follow-up time for drug court studies (e.g., 12 months rather than 36 months), given that research with sex offenders often requires a longer follow-up time to adequately detect recidivism.
Principles of Risk, Need, and Responsivity (RNR; Andrews, Bonta, & Hoge, 1990). As previously stated, the RNR principles were developed as a model to effectively guide correctional treatment. Each principle is meant to direct treatment providers to create a treatment plan for individual offenders based on their assessed level of risk, areas of criminogenic need, and personal learning style. Adherence to these principles has been shown to reduce recidivism in residential and community correctional treatment settings for a range of offender groups (Andrews & Bonta, 2006). In the current context, treatment quality can be measured by rating a program’s adherence to the RNR principles. For each of the three principles, a rating of (1) for “adherence” or (0) for “non-adherence” is assigned based on the information provided in the evaluations. Programs can adhere to any number of principles, ranging from zero to three. As drug treatment courts comprise two distinct but collaborative components, adherence to the RNR principles was coded separately for the court and for the treatment program.

Coding for adherence to each of the RNR principles required the demonstration of minimal adherence to these factors. At a minimum, in order for a program to be coded as adhering to the risk principle, it needed to demonstrate an assessment of risk for both the treatment and comparison groups by comparing the groups on a minimum of five risk-related factors (above and beyond substance abuse). For a program to meet the minimum requirement for adherence to the need principle, identification and targeting of factors (beyond substance abuse) that are empirically related to offending was required. And finally, to be coded as adhering to the responsivity principle, the program needed to demonstrate the use of cognitive-behavioral interventions and/or the tailoring of treatment to the offender’s unique characteristics (e.g., learning abilities, ethnicity).

Procedure

All 96 studies used in the present evaluation were assigned unique identification numbers. After being trained on the CODC Guidelines, the main author rated study quality for each study using the Guidelines. To ensure the modified CODC Guidelines were coded reliably, a second rater coded 10 of the 96 studies. Interrater agreement was compared for the global ratings of confidence (80%), amount of bias (80%), direction of bias (70%), and global study quality (90%).

Following the assessment of study quality, each study and individual effect size was coded for a variety of study descriptors (e.g., methodological characteristics, offender sample, outcome measures). Information required for calculating effect sizes included: sample sizes, number of recidivists, and statistics used to evaluate treatment effectiveness (e.g., the odds ratio from a logistic regression analysis). In cases in which multiple measures of recidivism were reported, the most inclusive outcome was chosen (e.g., any arrest versus drug convictions). For type of recidivism, the most general outcome was preferred over outcomes such as drug possession, theft, or prostitution. Where possible, recidivism information for program graduates and dropouts was coded separately and later combined to calculate an effect size.
The treatment quality of each drug treatment court was evaluated based on its adherence to the principles of Risk, Need, and Responsivity. To be coded as adhering to the risk principle, a program had to assess risk for both the treatment and comparison groups by comparing the groups on a minimum of five risk-related factors (beyond substance abuse). In order for a program to adhere to the need principle, we required that the intervention target criminogenic needs beyond substance abuse (given that substance abuse is the main focus of drug courts). To adhere to the responsivity principle, the intervention needed to demonstrate its services were tailored to an offender’s individual learning style, including an emphasis on cognitive-behavioral interventions (Andrews et al., 1990). Interrater reliability of the treatment quality ratings was assessed in the same fashion as study quality. A third coder rated 10 studies and interrater agreement for RNR adherence was 100%.

Plan of Analysis

First, descriptive information (i.e., study characteristics and CODC outcome ratings) was gathered. Next, an effect size was calculated for each unique sample. To produce the most reliable and valid estimate of the effects of drug courts, the odds ratio was chosen as the most appropriate effect size as both variables of interest are dichotomous (treatment versus no treatment exposure and recidivism versus nonrecidivism). Also, compared to correlations, odds ratios are more stable estimates of the effect of treatment when study quality is not optimal.

An odds ratio is a comparative measure of risk for a particular outcome. It calculates the likelihood of a specified outcome for someone exposed to a factor of interest (i.e., treatment) as compared to someone who is not exposed (Westergren, Karlsson, Andersson, Ohlsson, & Hallberg, 2001). An odds ratio of 1.0 indicates the ratio of recidivism for the treated group is equal to the ratio of recidivism for the comparison group; thus, treatment has no effect. As an odds ratio approaches 0, it is indicative of small odds of recidivism for the treatment group relative to the odds of recidivism for the comparison group. This translates to more effective treatment.

As suggested by Hanson, Broom, and Stephenson (2005), the effect size transformations were conducted prior to calculating mean effect sizes. Odds ratios are not normally distributed (highly skewed from 0 to infinity); therefore, they were converted to log odds ratios to normalize the distribution. Additionally, effect sizes were weighted to allow studies with larger sample sizes to contribute more to the overall effect size than studies with smaller sample sizes. Weighting was accomplished by weighting each effect by the inverse of its variance. Mean effect sizes were ultimately converted back to odds ratios, which are the effect sizes reported.

Effect size tables presented below include mean weighted odds ratios, 95% confidence intervals (CI), k (the number of studies), N (the total number of subjects contributing to the mean odds ratio), and the Q and Birge (H²) statistics. The Q statistic is a commonly used test for homogeneity of variance particularly in meta-
analyses (Medina, Martinez, Meca, & Botella, 2006). The distribution of the Q statistic is the same as the χ² distribution. The Birge statistic (H²) is also reported.1 Whereas Q tests for homogeneity of variance, the H² statistic allows the researcher to examine the amount of between-study variability. Smaller H² ratios indicate less between-study variability and larger H² ratios indicate greater amounts of between-study variability. Finally, in order to compare effect sizes across different levels of a variable (e.g., different levels of study quality), χ² analyses were calculated.2

Results

Study Descriptors

As seen in Table 1, the majority of the studies are unpublished reports (77%) evaluating drug courts in the United States (95%) in the mid to late 1990s. Non-randomized designs were most frequently used (88%), with only 12 evaluations using randomized designs. Most of these studies involved adult offenders (k = 74). For those studies reporting specific retention or graduation rates (k = 74), the average graduation rate was 39.9% (SD = 19.2). Of note, 53 of the 55 studies coded by Latimer et al. (2006), all 23 studies coded by Lowenkamp et al. (2005), and 63 of the 67 studies coded by Wilson et al. (2006) were included in the present investigation (studies were overlapping across the three meta-analyses).

Study Quality

In order to evaluate overall study quality, the CODC Guidelines outcome ratings were examined. Of all the studies included in our sample, over three quarters (k = 78) were rated as “rejected,” 23 studies were rated as “weak,” and only 2 studies rated as “good.” None of the studies was rated “strong.” The studies rated as “weak” or “good” were combined into one group of “acceptable” studies (k = 25). Table 2 provides the CODC confidence items, revealing that over half of the studies (k = 56) received a global confidence rating of “little confidence.” Of particular note, a vast majority of studies (k = 72) were rated as “little confidence” on the item assessing the adequacy of search for differences and over half (k = 59) were rated as producing “little confidence” on the item assessing effectiveness of statistical controls.

Examining the bias ratings of the items on the CODC Guidelines (Table 3) revealed that almost half of the studies (k = 46) received a global bias rating of “considerable bias.” Although in the majority of studies the direction of bias was unclear, when the direction of bias was known (k = 42), it was primarily in the

---

1 Birge statistic, where \[ H^2 = \frac{Q}{(k-1)} \].

2 \[ Q_{tot} - \sum Q_i \] where i refers to each level of the independent variable and df is i – 1.
Table 1
Frequencies and Percentages of Study Descriptors

<table>
<thead>
<tr>
<th>Study Descriptor</th>
<th>k</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Journal report</td>
<td>23</td>
<td>22.3</td>
</tr>
<tr>
<td>Unpublished report</td>
<td>79</td>
<td>76.7</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>1.0</td>
</tr>
<tr>
<td>Design Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized</td>
<td>12</td>
<td>11.7</td>
</tr>
<tr>
<td>Nonrandomized</td>
<td>91</td>
<td>88.3</td>
</tr>
<tr>
<td>Country</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>98</td>
<td>95.1</td>
</tr>
<tr>
<td>Canada</td>
<td>2</td>
<td>1.9</td>
</tr>
<tr>
<td>Australia</td>
<td>3</td>
<td>2.9</td>
</tr>
<tr>
<td>Year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1989-1994</td>
<td>22</td>
<td>21.4</td>
</tr>
<tr>
<td>1995-1999</td>
<td>65</td>
<td>63.1</td>
</tr>
<tr>
<td>2000-2005</td>
<td>16</td>
<td>15.5</td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult</td>
<td>74</td>
<td>71.8</td>
</tr>
<tr>
<td>Juvenile</td>
<td>11</td>
<td>10.7</td>
</tr>
<tr>
<td>Mixed</td>
<td>7</td>
<td>7.0</td>
</tr>
<tr>
<td>Not reported</td>
<td>11</td>
<td>10.7</td>
</tr>
</tbody>
</table>

Table 2
Studies corresponding to Collaborative Outcome Data Committee Guidelines (CODC) Confidence Items and Outcomes

<table>
<thead>
<tr>
<th>CODC Confidence Items</th>
<th>Little Confidence</th>
<th>Some Confidence</th>
<th>High Confidence</th>
<th>Insufficient Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global confidence rating</td>
<td>56</td>
<td>46</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>Defining treatment</td>
<td>56</td>
<td>46</td>
<td>1</td>
<td>--</td>
</tr>
<tr>
<td>Defining comparison</td>
<td>54</td>
<td>46</td>
<td>3</td>
<td>--</td>
</tr>
<tr>
<td>Sample size of treatment</td>
<td>7</td>
<td>64</td>
<td>32</td>
<td>--</td>
</tr>
<tr>
<td>Sample size of comparison</td>
<td>10</td>
<td>63</td>
<td>30</td>
<td>--</td>
</tr>
<tr>
<td>Adequacy of search for differences</td>
<td>72</td>
<td>27</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Length of follow-up</td>
<td>8</td>
<td>72</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>Recidivism validity/reliability</td>
<td>1</td>
<td>67</td>
<td>14</td>
<td>21</td>
</tr>
<tr>
<td>Data dredging</td>
<td>19</td>
<td>50</td>
<td>34</td>
<td>--</td>
</tr>
<tr>
<td>Effectiveness of statistical controls</td>
<td>59</td>
<td>39</td>
<td>5</td>
<td>--</td>
</tr>
</tbody>
</table>
### Table 3

Studies Corresponding to Collaborative Outcome Data Committee Guidelines (CODC) Quantity of Bias and Direction of Bias Items and Outcomes

<table>
<thead>
<tr>
<th>Bias Items</th>
<th>Negligible</th>
<th>Some</th>
<th>Considerable</th>
<th>Insufficient Information</th>
<th>QUANTITY OF BIAS</th>
<th>No Bias</th>
<th>Increases</th>
<th>Decreases</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global bias</td>
<td>1</td>
<td>56</td>
<td>46</td>
<td>--</td>
<td>1</td>
<td>39</td>
<td>3</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Misc. factors</td>
<td>--</td>
<td>31</td>
<td>2</td>
<td>70</td>
<td>--</td>
<td>5</td>
<td>3</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Experimenter involvement</td>
<td>74</td>
<td>21</td>
<td>2</td>
<td>6</td>
<td>74</td>
<td>15</td>
<td>--</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Blinding in data management</td>
<td>4</td>
<td>12</td>
<td>--</td>
<td>87</td>
<td>4</td>
<td>4</td>
<td>--</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Subject selection</td>
<td>8</td>
<td>55</td>
<td>35</td>
<td>5</td>
<td>8</td>
<td>56</td>
<td>--</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Program attrition</td>
<td>1</td>
<td>16</td>
<td>66</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>79</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Intent-to-treat</td>
<td>1</td>
<td>16</td>
<td>64</td>
<td>22</td>
<td>1</td>
<td>2</td>
<td>55</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Attrition in follow-up</td>
<td>72</td>
<td>9</td>
<td>8</td>
<td>14</td>
<td>72</td>
<td>2</td>
<td>2</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>A priori equivalency</td>
<td>10</td>
<td>63</td>
<td>28</td>
<td>2</td>
<td>10</td>
<td>56</td>
<td>3</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Findings on equivalency</td>
<td>2</td>
<td>19</td>
<td>18</td>
<td>64</td>
<td>2</td>
<td>9</td>
<td>3</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>Equivalency of follow-up</td>
<td>54</td>
<td>25</td>
<td>5</td>
<td>19</td>
<td>53</td>
<td>6</td>
<td>1</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Least bias comparison</td>
<td>6</td>
<td>58</td>
<td>39</td>
<td>--</td>
<td>7</td>
<td>49</td>
<td>2</td>
<td>45</td>
<td></td>
</tr>
</tbody>
</table>
direction favoring treatment effectiveness \( (k = 39) \). There were a number of bias items for which “considerable bias” was frequently coded, including: program attrition, intent-to-treat, computation of least bias comparison, and subject selection. Finally, it was found that 53 programs were deemed “implementation failures” (i.e., attrition rates greater than 49%).

**Effectiveness of Drug Courts**

The mean weighted odds ratios estimating the effectiveness of drug courts were calculated for studies included in each of the three meta-analyses. For those studies used by Latimer et al. (2006), the mean weighted odds ratio was .721 (95% CI = .684 to .759). For studies used by Lowenkamp et al. (2005), the mean weighted odds ratio was .671 (95% CI = .623 to .723). Finally, for studies used by Wilson et al. (2006), the mean weighted odds ratio was .669 (95% CI = .638 to .700). The overall mean weighted odds ratio when all studies were grouped together \( (k = 96) \) was calculated to be .671 (95% CI = .646 to .698).

Mean weighted effect sizes were then calculated based on the CODC Guidelines global study quality ratings (i.e., reject, weak, good). Table 4 presents the mean

<table>
<thead>
<tr>
<th>Variable/Factor</th>
<th>Mean OR</th>
<th>95% CI Low</th>
<th>95% CI High</th>
<th>Q</th>
<th>k</th>
<th>N</th>
<th>( H^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>All studies</td>
<td>.671</td>
<td>.646</td>
<td>.698</td>
<td>620.82*</td>
<td>96</td>
<td>50,640</td>
<td>6.53</td>
</tr>
<tr>
<td>Rejected</td>
<td>.657</td>
<td>.627</td>
<td>.688</td>
<td>524.08*</td>
<td>71</td>
<td>36,439</td>
<td>7.49</td>
</tr>
<tr>
<td>Weak</td>
<td>.704</td>
<td>.653</td>
<td>.760</td>
<td>90.25*</td>
<td>23</td>
<td>13,338</td>
<td>4.10</td>
</tr>
<tr>
<td>Good</td>
<td>.850</td>
<td>.611</td>
<td>1.185</td>
<td>2.11</td>
<td>2</td>
<td>863</td>
<td>2.11</td>
</tr>
<tr>
<td>Confidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little</td>
<td>.726</td>
<td>.682</td>
<td>.772</td>
<td>403.82*</td>
<td>50</td>
<td>21,034</td>
<td>8.24</td>
</tr>
<tr>
<td>Some</td>
<td>.636</td>
<td>.606</td>
<td>.671</td>
<td>206.42*</td>
<td>45</td>
<td>29,371</td>
<td>4.69</td>
</tr>
<tr>
<td>High</td>
<td>.531</td>
<td>.259</td>
<td>1.088</td>
<td>--</td>
<td>1</td>
<td>235</td>
<td>--</td>
</tr>
<tr>
<td>Quantity of bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Considerable</td>
<td>.650</td>
<td>.614</td>
<td>.689</td>
<td>336.15*</td>
<td>42</td>
<td>23,055</td>
<td>8.00</td>
</tr>
<tr>
<td>Some</td>
<td>.686</td>
<td>.649</td>
<td>.724</td>
<td>279.23*</td>
<td>53</td>
<td>26,957</td>
<td>5.37</td>
</tr>
<tr>
<td>Minimal/little</td>
<td>.967</td>
<td>.665</td>
<td>1.405</td>
<td>--</td>
<td>1</td>
<td>628</td>
<td>--</td>
</tr>
<tr>
<td>Direction of bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>.531</td>
<td>.259</td>
<td>1.088</td>
<td>--</td>
<td>1</td>
<td>235</td>
<td>--</td>
</tr>
<tr>
<td>Unknown</td>
<td>.721</td>
<td>.684</td>
<td>.760</td>
<td>282.23*</td>
<td>57</td>
<td>28,385</td>
<td>5.04</td>
</tr>
<tr>
<td>Decreases treatment</td>
<td>2.070</td>
<td>1.589</td>
<td>2.697</td>
<td>1.22</td>
<td>3</td>
<td>1,351</td>
<td>0.61</td>
</tr>
<tr>
<td>Increases treatment</td>
<td>.578</td>
<td>.544</td>
<td>.614</td>
<td>236.68*</td>
<td>35</td>
<td>20,669</td>
<td>6.96</td>
</tr>
</tbody>
</table>

* \( p < .01 \), two-tailed.
weighted odds ratios, the 95% confidence intervals, \( k \), \( N \), \( Q \), and \( H^2 \) for the studies grouped by CODC Guidelines global study quality rating, global confidence (i.e., little, some, and high), global rating for quantity of bias (i.e., considerable, some, and minimal) and global rating for direction of bias (i.e., none, unknown, increases treatment, decreases treatment). It is important to note that the \( Q \) statistics are very large and only two studies received a global rating of “good.”

The odds ratios for each of the global study quality ratings were not significantly different (\( \chi^2 = 4.38; df = 2; p > .05 \)), neither were the odds ratios for global bias ratings (\( \chi^2 = 5.44; df = 2; p > .05 \)). On the other hand, the odds ratios were significantly different based on the ratings of global confidence (\( \chi^2 = 10.58; df = 2; p < .05 \)) and global direction of bias (\( \chi^2 = 100.69; df = 3; p < .05 \)).

In the following analyses, only those studies that were “acceptable” (studies rated “weak” or “good”) were examined (see Table 5). The overall weighted mean odds ratio from “acceptable” studies was found to be .711 (95% CI = .660 to .766). These 25 studies were further broken down by ratings on global confidence, global bias, and global direction of bias. No significant differences were found on global confidence (\( \chi^2 = 0.64; df = 1; p > .05 \)) or global bias (\( \chi^2 = 2.70; df = 1; p > .05 \)). The odds ratios for the global direction of bias, however, were significantly different (\( \chi^2 = 25.07; df = 3; p < .05 \)).

<table>
<thead>
<tr>
<th>Variable/Factor</th>
<th>Mean OR</th>
<th>95% CI</th>
<th>Q</th>
<th>k</th>
<th>N</th>
<th>( H^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptable studies</td>
<td>.711</td>
<td>.660</td>
<td>.766</td>
<td>93.54*</td>
<td>25</td>
<td>14,201</td>
</tr>
<tr>
<td>Confidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some</td>
<td>.713</td>
<td>.662</td>
<td>.769</td>
<td>92.90*</td>
<td>24</td>
<td>13,966</td>
</tr>
<tr>
<td>High</td>
<td>.531</td>
<td>.259</td>
<td>1.088</td>
<td>--</td>
<td>1</td>
<td>235</td>
</tr>
<tr>
<td>Quantity of bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some</td>
<td>.702</td>
<td>.651</td>
<td>.757</td>
<td>90.84*</td>
<td>24</td>
<td>13,573</td>
</tr>
<tr>
<td>Minimal/little</td>
<td>.967</td>
<td>.665</td>
<td>1.405</td>
<td>--</td>
<td>1</td>
<td>628</td>
</tr>
<tr>
<td>Direction of bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>.531</td>
<td>.259</td>
<td>1.088</td>
<td>--</td>
<td>1</td>
<td>235</td>
</tr>
<tr>
<td>Unknown</td>
<td>.705</td>
<td>.650</td>
<td>.764</td>
<td>53.74*</td>
<td>18</td>
<td>11,221</td>
</tr>
<tr>
<td>Increases treatment</td>
<td>.591</td>
<td>.476</td>
<td>.733</td>
<td>14.73*</td>
<td>5</td>
<td>2,143</td>
</tr>
<tr>
<td>Decreases treatment</td>
<td>1.836</td>
<td>1.230</td>
<td>2.740</td>
<td>--</td>
<td>1</td>
<td>602</td>
</tr>
</tbody>
</table>

* \( p < .01 \), two-tailed.
After the odds ratios were computed, they were then converted to percentages representing recidivism differences. Figure 1 shows the recidivism differences that correspond to the findings from each of the three meta-analyses for all of the studies grouped by CODC outcome rating (i.e., reject, weak, good) and “acceptable” studies (i.e., weak or good). Based on only the methodologically acceptable studies ($k = 25$), it was calculated that drug courts produce an 8.4% reduction in recidivism. Excluding studies that were rated weak and only including the best studies (i.e., good studies) showed an overall reduction in recidivism of 4%.

**Effects of Treatment Quality**

The final analysis examined the effectiveness of drug treatment courts based on treatment quality. Only studies deemed methodologically acceptable were included in this analysis. Overall, of the 25 acceptable studies, 11 drug courts demonstrated “no adherence” to any of the three RNR principles, 13 courts showed “adherence to one principle,” and only one showed “adhered to two principles.” None of the drug courts showed “adherence to three principles.” Table 6 presents the metaanalytic summary statistics for the different levels of adherence to the principles of Risk, Need, and Responsivity. The odds ratios were significantly different between the three levels of adherence ($\chi^2 = 14.82; df = 2; p < .05$). Courts that adhered to any of the three principles were compared to those that adhered to none. The odds ratios were significantly different ($\chi^2 = 6.45; df = 1; p < .05$). The linear trend of
increasing adherence to principles of RNR was tested and found to be significant \( (t = 3.05, p < .01) \). In other words, as adherence to RNR increased, the strength of the effectiveness of drug courts respectively increased.

**Table 6**

Acceptable Studies and Risk, Need, and Responsivity (RNR) Adherence \((k = 25)\)

<table>
<thead>
<tr>
<th>Variable/Factor</th>
<th>Mean OR</th>
<th>95% CI</th>
<th>Q</th>
<th>k</th>
<th>N</th>
<th>( H^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptable</td>
<td>.711</td>
<td>.660</td>
<td>.766</td>
<td>93.5*</td>
<td>25</td>
<td>14,201</td>
</tr>
<tr>
<td>RNR = 0</td>
<td>.821</td>
<td>.718</td>
<td>.938</td>
<td>38.15*</td>
<td>11</td>
<td>5,511</td>
</tr>
<tr>
<td>RNR = 1</td>
<td>.682</td>
<td>.623</td>
<td>.746</td>
<td>40.57*</td>
<td>13</td>
<td>8,442</td>
</tr>
<tr>
<td>RNR = 2</td>
<td>.306</td>
<td>.179</td>
<td>.523</td>
<td>--</td>
<td>1</td>
<td>248</td>
</tr>
<tr>
<td>RNR = 3</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>RNR ≥ 1</td>
<td>.667</td>
<td>.610</td>
<td>.729</td>
<td>48.94*</td>
<td>14</td>
<td>8,690</td>
</tr>
</tbody>
</table>

* \( p < .01 \), two-tailed.

**Discussion**

There has been some skepticism regarding the accuracy of the three recent meta-analyses assessing the effectiveness of drug courts. Major methodological issues and variability in treatment quality have been cited as two potential sources of bias in the evaluations of drug courts to date (Belenko, 2001). The present investigation made use of a measure (i.e., CODC Guidelines) designed for treatment outcome studies with offenders to assess the quality of drug court studies. Treatment quality was also evaluated by rating program adherence to the principles of effective correctional programming (i.e., risk, need, and responsivity). Utilizing this method allowed for an empirical investigation of the influence of these two factors on meta-analytic estimates of the effectiveness of these specialty courts.

**Findings on Study Quality**

Using the CODC Guidelines, almost three quarters of the studies included in the sample were rejected on the basis of major methodological problems, and only 25 studies were deemed methodologically “acceptable” (i.e., “weak” or “good”). Only two of the acceptable studies were rated as “good” and none of the studies was rated “strong.” In other words, a vast majority of the studies used in the previous meta-analyses contained major methodological problems that were likely to have influenced interpretations of the meta-analytic findings. The results of the current
study demonstrate that the drug court literature suffers from serious methodological weaknesses that limit the confidence researchers can place in the findings.

Regarding bias specifically, it was found that 44% of the studies were rated as having considerable bias. When the direction of bias was known, it was almost exclusively artificially increasing the effect of the drug court program. Only three studies were rated as containing bias that decreased the effect of treatment, and only a single study was rated without bias. The assessment of study quality suggests that there can be little confidence placed in the results of most of these studies as they possess considerable amounts of bias, which inflate the positive findings of the individual evaluations.

Further examination of the CODC Guidelines items highlighted specific problem areas. Consistent with what has been reported in the drug court literature, program attrition emerged as a major methodological issue. Attrition rates in drug treatment court programs have been well-documented to be around 50% (Cissner & Rempel, 2005). This is similar to the findings of the present investigation, with over half of the studies reporting attrition rates greater than 49% and average graduation rates of approximately 40%. Such high attrition rates have a major impact on study quality, as it becomes increasingly difficult to measure the magnitude of intervention effectiveness as dropout rates increase.

The recommended strategy to handle program attrition is to use intent-to-treat analyses (CODC, 2007a; 2007b; Thomas, Ciliska, Dobbins, & Micucci, 2004). This involves using outcome information for all subjects, regardless of whether they dropped out of treatment, in the estimation of treatment effectiveness. It is important to note that intent-to-treat analyses produce more conservative estimates of the effects of treatment given that fewer individuals actually receive the intervention as attrition rates increase. However, intent-to-treat analyses are a better estimate of treatment effectiveness compared to analyses that only use treatment completers. Such strategies overestimate the effectiveness of a program by ignoring program failures. Also, the proportion of potential dropouts in a comparison group is larger than in a treatment completer group. It has also been shown that dropouts tend to have higher recidivism rates than even untreated participants, further biasing the completer-untreated group comparison (Hanson et al., 2002; Seager, Jellicoe & Dhaliwal, 2004).

In the current study, all effect sizes were calculated by utilizing an intent-to-treat analysis, requiring recidivism information for both completers and dropouts. As for the individual studies, less than 30% of studies reported their findings based on intent-to-treat analyses. For future drug treatment court evaluations, researchers should collect outcome information on all admissions in order to calculate a less-biased estimate of program effectiveness. Furthermore, researchers should consider dosage factors (i.e., intensity and duration) when deciding what to include in their estimations of treatment effectiveness (e.g., covariate analyses).

The second major methodological issue that emerged from the assessment of study quality was preexisting group differences (i.e., treatment versus comparison).
Based on the CODC Guidelines ratings, it was found that 70% of studies conducted inadequate searches for differences between groups. Even though it was common to compare the treatment and comparison groups on demographic variables (e.g., age, race, gender), the studies rarely compared groups on risk-related factors or validated risk scores. Particularly in nonrandomized designs, it is critical that researchers demonstrate group equivalency, as risk-relevant differences could account for differences found in outcomes.

In the future, researchers can improve evaluations by making use of validated risk assessments to establish equivalency. In some cases, however, this may not be possible (e.g., in retrospective designs). A reasonable alternative to risk assessments is to construct a scale of risk-relevant indicators of recidivism (i.e., criminogenic needs), validate the scale on a comparison group, and use the scale to demonstrate group equivalency (Hanson et al., 2005). By using a validated risk measure or constructing and validating a scale within a study, an evaluation can produce more confidence that differences in outcomes between groups are due to treatment effects rather than preexisting differences. In addition, such a measure would improve other efforts to control preexisting differences either through methodological strategies (e.g., subject matching) or post-hoc statistical control procedures (e.g., covariate analysis).

**Evaluations of Drug Treatment Courts: Effects of Study Quality**

One of the main goals of the present investigation was to estimate the effectiveness of drug courts while controlling for the influence of study quality. Using all of the studies in our sample, the overall odds ratio was found to be .671 (95% CI = .646 to .698; k = 96). This translates to an 11% difference in recidivism rates between drug treatment court and comparison groups. This result most closely resembles the effect found in the Lowenkamp et al. (2005) study. In fact, the effect size estimates found in the current study based on Latimer’s and Wilson’s studies both resulted in a smaller treatment effect than those reported in the original meta-analyses. Two factors may account for this. Firstly, there are differences in the manner in which the researchers calculated the overall effect sizes. Contact with the authors revealed that of the three meta-analyses, only Lowenkamp et al. (2005) utilized intent-to-treat analyses when calculating the overall effect sizes (personal communication, 2006). Secondly, it is likely that study quality played a role as well. The overall magnitude of the Q and $H^2$ clearly indicates a substantial proportion of between-study variability.

The comparison of effect sizes based on global study quality ratings did not yield significant differences ($\chi^2 = 4.38; df = 2; p > .05$), suggesting effect sizes were not significantly related to overall study quality. Some caution is warranted when interpreting these findings, as there were relatively few studies rated as “weak” (k = 23) and only two rated “good.” Given that the majority of studies were rejected based on the CODC Guidelines, it is important to note that there was major between-study variability (see Table 4). In fact, the only studies that revealed
homogeneity of between-study variance were the two studies rated as “good” on global study quality and the three studies whose bias was rated as decreasing the effect of treatment.

Nonetheless, it was found that as study quality improved, the effect of drug treatment courts decreased (i.e., the value of the odds ratio approached 1). In fact, when only acceptable studies were included, the overall effect of drug treatment courts was smaller, translating to an 8% difference in recidivism. Consequently, this estimate is likely a more reliable estimate of the effects of drug treatment courts in reducing recidivism than previous estimates, as only those studies deemed methodologically acceptable were included.

**Evaluations of Drug Treatment Courts: Effects of Treatment Quality**

The second component of this study was to assess the relationship between treatment quality and the effectiveness of drug courts. Using only the “acceptable” studies in our sample, the overall odds ratio was calculated to be .711 (95% CI = .660 to .766; $k = 25$). Consistent with criminal justice research assessing treatment effectiveness and adherence to the principles of Risk, Need and Responsivity (Andrews & Bonta, 2006; Simpson, 2008), it was found that the effects of drug courts significantly increased as adherence to RNR increased ($t = 3.05, p < .01$). In terms of reductions in recidivism, adherence to none, one, or two of the principles corresponded to a 5%, 11%, and 31% reduction in recidivism, respectively. None of the studies included in the current meta-analysis adhered to all three RNR principles; however, other meta-analytic reviews have shown that adherence to all three principles can produce up to 35% reductions in recidivism (Andrews & Bonta, 2006). It was also found that RNR played a role in the homogeneity of acceptable studies. This trend is best illustrated by comparing the $H^2$ statistics for rejected studies ($k = 71; H^2 = 7.49$), acceptable studies ($k = 25; H^2 = 3.90$), and acceptable studies that adhere to one or more of the RNR principles ($k = 14; H^2 = 3.76$). This suggests that study quality and treatment quality can account for much of the variance that is seen among drug treatment court evaluations.

An exploration of the need areas targeted by drug courts revealed that in addition to targeting substance abuse, treatment was often geared towards the enhancement of general well-being (e.g., mental health, relaxation). Drug treatment courts could gain from integrating knowledge from the “what works” literature regarding effective correctional programming. For example, making use of validated risk assessments to identify individual criminogenic need areas could improve the effectiveness of these correctional programs for criminal justice outcomes in addition to improving an individual’s well-being. It is important for community-based treatment programs to note that there are differences between substance-abusing offenders and substance abusers outside of the legal system. This distinction makes it necessary for drug court programs to integrate evidence-based strategies that have proven effective with offender populations (i.e., RNR).
Summary

In summary, the present study found that study quality and treatment quality greatly influenced the results of the drug court evaluations included in three previous meta-analyses. Issues surrounding quasi-experimental study designs, comparison groups, management of high attrition rates, as well as inadequate searches and controls for group differences are methodological problems that often biased evaluations in favor of treatment effectiveness. The assessment of study quality with the CODC Guidelines showed estimates of drug court effectiveness in reducing recidivism are mostly based on studies with highly biased methodologies. These findings suggest that study quality influences study results. As methodology gets poorer, the variance among studies increases. And, since bias in the drug court literature tends to favor treatment outcomes, as methodology gets worse and variance increases, reported treatment effect sizes increase respectively.

The role of treatment quality was also explored and the results suggest that treatment quality is related to drug court effectiveness. For methodologically acceptable studies, as adherence to RNR increased, the effectiveness of treatment increased respectively. The homogeneity of studies also decreased as adherence to the principles of RNR increased.

Overall, the findings from this study suggest that the drug treatment court literature is littered with methodological problems, study quality greatly influences study outcomes, and attention must be paid to the direction of bias contained within a study. This study also found that treatment efficacy was dependent upon adherence to the RNR principles of effective correctional programming. An assessment of adherence to the RNR principles showed that very few programs adhered to at least one of the principles, and none of the programs adhered to all three of the principles.

Although drug treatment courts present an alternative to incarceration, appropriate implementation of the drug treatment court model and adherence to the principles of effective correctional practices are required to produce the desired results (i.e., reduce recidivism). Accurately translating what takes place behind the closed doors of drug treatment courts depends on good quality evaluations. Currently, it is difficult to draw conclusions from the few acceptable studies. The least biased estimate of overall reductions in recidivism was approximately 8%. More methodologically sound research is needed in order to estimate the effectiveness of drug courts.

Current Research

Although the majority of evaluations reviewed in this meta-analysis were conducted five to 10 years ago, issues concerning the methodological quality of empirical studies will continue to play a substantive role in how we evaluate and build knowledge about what works and what is considered evidence-based practice. A review of the Multi-site Adult Drug Court Evaluation conducted by the Urban Institute, a large scale and ambitious drug court evaluation (Rossman, Roman,
Zweig, Rempel, & Lindquist, 2011) serves to illustrate how current research and evaluation has, on the one hand, improved on addressing some of the methodological issues raised in the meta-analysis. On the other hand, it also serves to illustrate that there is still work to be done.

In this Multi-site Adult Drug Court Evaluation, 23 drug courts in the United States were comprehensively evaluated to improve our knowledge about effectiveness. In this evaluation, the drug courts appeared to be placing a greater emphasis on treatment quality through the use of cognitive-behavioral programs in the community that teach the link between thoughts and behaviors. The use of cognitive-behavioral interventions is integral to the Responsivity principle of effective offender treatment and is an essential component in the reduction of recidivism with offenders in the community.

In terms of issues that impact methodological quality, greater efforts are being made to apply post-hoc statistical controls (e.g., propensity score modeling) for group differences on demographic variables (e.g., race and gender). These propensity scores, however, based on simple and easily accessible demographic variables, are limited and provide a less accurate (less predictive) measure of risk than the use of validated third-generation risk/need instruments (particularly those that include assessments of dynamic risk factors empirically related to recidivism, such as antisocial associates and procriminal attitudes). The use of validated risk/need assessments in drug court evaluations would enhance the researcher’s ability to demonstrate group equality (treatment versus comparison group) as well as provide a better statistical control in any post-hoc analysis. In addition, researchers still need to develop a better strategy to handle the high attrition rates found in drug courts. In the Multi-site Adult Drug Court Evaluation, an average 41% drop-out rate (of those participants who exited the program) was reported for the 23 courts. Using intent-to-treat analysis (for both the drug court group as well as the comparison groups) and ensuring survival times commence equally for both groups would provide more confidence in (and decrease potential bias of) empirical evidence regarding the effectiveness of these courts.

Limitations and Future Research

This study has limitations that can guide future research. We acknowledge that information used to rate studies on the CODC guidelines as well as to judge the drug courts’ adherence to the principles of Risk, Need, and Responsivity was obtained solely from the reports. A continuation of this research will further explore the quality of treatment by contacting the evaluation teams as well as the drug treatment courts for more detailed information.

It is hoped that future research in this area will consider the influence of methodology on study results and use this knowledge to guide decisions in designing and conducting future evaluations. It is also hoped that drug treatment courts will make greater use of what is known regarding effective correctional treatment practices in order to improve treatment quality and reduce criminal behavior.
References

References marked with an asterisk indicate studies included in the meta-analysis.


*Cosden, M., Crothers, L., & Peerson, S. (1999). *Superior Court of California County of Ventura: Drug court (Summary findings).* Santa Barbara, CA: University of California, Graduate School of Education.


Leticia Gutierrez, M.A., (Carleton University) is a doctoral student in the Forensic Psychology program at Carleton University and a Research Analyst in the Corrections Research Unit of the Department of Public Safety. Her areas of research include drug treatment courts, community supervision, aboriginal offenders, and restorative justice. Currently, Ms. Gutierrez is assessing the role of therapeutic relationships in community supervision settings in Canada.

Guy Bourgon, Ph.D., (University of Ottawa) is a clinical psychologist with a specialization in correctional and criminal justice psychology. He has over 25 years of clinical experience in the assessment and treatment of adults and youths involved in the criminal justice system. Dr. Bourgon has been involved in the development of various effective correctional programs and has extensive experience in the training and supervision of frontline professionals. He is presently a Senior Researcher for Public Safety Canada, conducting research on effective correctional services and practices, and has published articles on effective correctional treatment across a range of offender populations and programs, risk assessment, research methodology, and effective knowledge transfer to everyday practice. Dr. Bourgon maintains a private practice, is an adjunct professor at Carleton University, and is a member of the Editorial Board of Criminal Justice and Behavior.