

**Drug Treatment Courts:  
A Quantitative Review  
of Study and  
Treatment Quality  
2009-04**

Leticia Gutierrez & Guy Bourgon

Public Safety Canada

## **Abstract**

The effectiveness of drug courts has been debated in regards to two main factors: (1) study quality and (2) treatment quality. The current study examined these two factors. Study quality was examined using the Collaborative Outcome Data Committee Guidelines (CODC), and treatment quality was assessed by evaluating adherence to the principles of Risk-Need-Responsivity (RNR). Using the CODC Guidelines, studies were rated as: “rejected”, “weak”, “good” or “strong” based on methodological quality. These guidelines have been used in meta-analytic reviews of sex offender (Hanson, Bourgon, Helmus & Hodgson, 2009) and community supervision (Simpson, 2008) outcome evaluations. The RNR principles have been previously shown to mediate the effectiveness of offender treatment across various offender groups and a variety of criminogenic needs (e.g., substance abuse, sexual offending). In total, 96 studies were reviewed and assessed according to study and treatment quality. Results of this review found that the study quality of the literature is poor and this accounted for much of the variability found 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 in reducing 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%. Limitations and future research directions are discussed.

## **Author’s Note**

The views expressed are those of the authors and not necessarily those of Public Safety Canada. Correspondence concerning this article should be addressed to Leticia Gutierrez, Corrections Research, Public Safety Canada, 340 Laurier Avenue West, Ottawa, Ontario, Canada, K1A 0P8; email: Leticia.Gutierrez@ps-sp.gc.ca.

## **Acknowledgements**

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 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to thank Shannon Hodgson, Jan Roehl, and David Wilson, who assisted in providing information (e.g., court evaluations, unpublished reports) that was useful for conducting the present review, as well as Tanya Rugge and Jim Bonta for their feedback and suggestions.

## **Product Information**

December 2009

Cat. No.: PS3-1/2009-4E-PDF

ISBN No.: 978-1-100-14490-0

Ottawa

## Introduction

The first drug treatment court, which opened in Miami in 1989, came in response to the rising rates of drug-related offences in the United States. At that time, there was widespread use of crack cocaine, and courts were handing down more and longer custodial sentences to substance-abusing offenders. As a result, prison overcrowding became a significant issue. In an effort to reduce prison overcrowding, 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 the inception of the first drug court, they have become a popular alternative to incarceration for non-violent substance-abusing offenders. Today, there are over 1,700 drug treatment courts in the United States, Canada, the United Kingdom, and Australia, with more in the planning stages (Weekes, Mugford, Bourgon & Price, 2007).

Given the popularity of drug courts, a number of researchers have sought to determine whether these drug courts are effective in reducing recidivism. Three meta-analyses have been conducted to date and all have found positive effects. Lowenkamp, Holsinger and Latessa (2005) conducted the first meta-analytic review of drugs courts. 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 and Chrétien (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, can bias the estimates of their effectiveness in reducing recidivism.

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, Gottfredson, MacKenzie, Eck, Reuter, & Bushway, 1997). Although randomized experiments in criminology have increased in recent years, these designs are difficult to employ due to a variety of ethical, legal and practical constraints (Farrington & Welch, 2005; Sherman et al., 1997). As a result, weaker quality quasi-experimental designs are the most frequently used methodologies in criminal justice research. Randomization can control for both known and unknown confounding variables (Farrington, 2002). Random assignment designs provide stronger evidence that any effects found are more likely attributable to a manipulation (e.g., treatment exposure) rather than pre-existing differences between the two groups. In non-random designs, researchers must make efforts to reduce and minimize any 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 characteristics (e.g., motivation, risk level).

In addition to pre-existing differences between groups, considerations should be made regarding any and all potential differences concerning issues of measurement. 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., drop-outs) and non-blind 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. Some indicators of descriptive validity include the sample size, description of intervention, and quality of control variables used in order to determine the effects of an intervention (e.g., risk measures). These factors significantly influence the degree of confidence one can place in the results of a study. Descriptive validity factors contextualize research findings, facilitate replication, and enhance the validity of analytic reviews (Farrington, 2002; 2003).

Due to the structure of drug courts, there are some common methodological issues. Inappropriate comparison groups are an issue 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 their high dropout rates (Cissner & Rempel, 2005; Weekes et al., 2007). Cissner & Rempel (2005) estimated an average attrition rate of 40% for drug courts. Based on the studies in their meta-analysis, Latimer et al. (2005) reported an average attrition rate of 45.2%, with rates ranging from 9% to 84.4%. Lastly, biased outcome measures are problematic for drug courts as this model is fundamentally different from traditional criminal justice processes and can result in systematic differences between the two groups. For example, the conviction and/or sentencing decisions are delayed until program completion for drug court participants. This can result in some between-group differences on official records where sentencing dates are utilized to identify recidivistic events. For a drug court client, the court internally handles most non-compliant behaviour through various sanctions, and non-compliance is often contextualized in terms of treatment behaviour and progress rather than criminal outcome. The result can lead to a systematic bias in outcomes favouring the drug court group over the comparison group.

Given that study quality is important when interpreting a study’s findings, especially 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 study quality assessment tools have recently been used in examinations of criminal justice interventions (Sherman et al., 1997; Hanson & Bourgon, 2008). 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 outcomes studies. The CODC Guidelines were developed by a group of researchers in order to rate the study quality of treatment outcomes studies and 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 contained within their design. The drug treatment court literature in particular faces many challenges relating to methodological reliability. In fact, the issues relating to study quality may play a role in the estimated effectiveness of drug courts found 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 towards 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 arms length from the courts. This program structure makes quality assurance and communication between the courts and treatment programs more difficult. Also, many of these community-based programs are not specifically geared towards offender samples; rather, they target

substance abusers in general. Therefore, the quality of treatment programs range in terms of their adherence to the principles of effective correctional programming (i.e., risk, need and responsivity).

The Risk-Need-Responsivity theory of criminal behaviour has been paramount in the development of effective correctional programming. This theory 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, Bonta & Hoge, 1990; Andrews & Bonta, 2006).

*Risk Principle.* The first component of the RNR theory is the risk principle, which states 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, a high-risk offender 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 theory addresses the importance of identifying and targeting an offender's criminogenic needs (dynamic risk factors) rather than non-criminogenic 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 re-offending. 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 behaviour (criminal history, a static risk factor, completes the Central Eight).

*Responsivity Principle.* The last principle of the RNR theory deals with the issue of general and specific responsivity. This principle can be interpreted as the "what works for whom" principle (Wormith, Althouse, Simpson, Reitzel, Fagan, & Morgan, 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 states that cognitive-behavioural interventions work best. Specific responsivity is a treatment matching style 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*

The meta-analyses that have been conducted to date assessing the effectiveness of drug courts have yielded positive results. However, debate over the influence of methodological flaws in the evaluations of drug courts and the inconsistency of effective programming have led researchers to question the validity of the results. 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 to be 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., 2006; 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 websites), 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, there were four studies that 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, and there were seven studies which did not contain sufficient treatment group information. As a result, no effect size 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 purposes 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 CODC Guidelines have been used for meta-analytic reviews of sex offender treatment (Hanson, Bourgon, Helmus & Hodgson, 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 21<sup>st</sup> item is specific to cross-institutional designs only) with 9 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 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 produced low confidence and/or contained 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 (i.e., defining treatment, adequacy of search for pre-existing differences and confidence in length of follow-up).

*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 are comprised of two distinct but collaborative components, adherence to the RNR principles was coded separately for the court and for the treatment program.

### *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. Inter-rater 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 where 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 were coded separately and combined later when calculating the effect size.

The treatment quality of each drug treatment court was evaluated based on its adherence to the principles of Risk, Need and Responsivity. In order to adhere to the need principle (given that substance abuse is the main focus of drug courts), we required that the intervention target criminogenic needs beyond substance abuse. 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-behavioural interventions (Andrews et al., 1990). Inter-rater reliability of the treatment quality ratings was assessed in the same fashion as study quality. A third coder rated 10 studies and inter-rater 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 non-recidivism). 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 and Broom (2005), the following 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* ( $H^2$ ) statistics. The *Q* statistic is a commonly used test for homogeneity of variance particularly in meta-analyses (Medina et al, 2006). The distribution of the *Q* statistic is the same as the  $\chi^2$  distribution. The *Birge* statistic ( $H^2$ ) is also reported<sup>1</sup>. Whereas *Q* tests for homogeneity of variance, the  $H^2$  statistic allows the researcher to examine the amount of between-study variability. Smaller  $H^2$  ratios indicate less between study variability and larger  $H^2$  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),  $\chi^2$  analyses were calculated<sup>2</sup>.

## 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.

*Table 1. Frequencies and Percentages of Study Descriptors*

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

### *Study Quality*

In order to evaluate overall study quality, the CODC Guidelines outcome ratings were examined. Of those studies included in previous meta-analyses, 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 were 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

<sup>1</sup> *Birge* statistic, where [ $H^2 = Q / (k-1)$ ].

<sup>2</sup>  $Q_{tot} - \sum Q_i$  where *i* refers to each level of the independent variable and *df* is *i* - 1.

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.

Table 2. Studies corresponding to CODC confidence items and outcomes

CODC Confidence Items	Little Confidence	Some Confidence	High Confidence	Insufficient Information
Global Confidence Rating	56	46	1	--
Defining Treatment	56	46	1	--
Defining Comparison	54	46	3	--
Sample Size of Treatment	7	64	32	--
Sample Size of Comparison	10	63	30	--
Adequacy of Search for Differences	72	27	2	2
Length of Follow-Up	8	72	18	5
Recidivism Validity/Reliability	1	67	14	21
Data Dredging	19	50	34	--
Effectiveness of Statistical Controls	59	39	5	--

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 direction favouring 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 evaluations were deemed “implementation failures” (i.e., attrition rates greater than 49%).

Table 3. Studies corresponding to CODC quantity of bias and direction of bias items and outcomes

Bias Items	Quantity of Bias				Direction of Bias			
	Negligible	Some	Considerable	Insufficient Information	No Bias	Increases Effect	Decreases Effect	Unknown
Global Bias	1	56	46	--	1	39	3	60
Misc. Factors	--	31	2	70	--	5	3	70
Experimenter Involvement	74	21	2	6	74	15	--	14
Blinding in Data Management	4	12	--	87	4	4	--	95
Subject Selection	8	55	35	5	8	56	--	39
Program Attrition	1	16	66	20	1	1	79	22
Intent-to-Treat	1	16	64	22	1	2	55	45
Attrition in Follow-up	72	9	8	14	72	2	2	27
A Priori Equivalency	10	63	28	2	10	56	3	34
Findings on Equivalency	2	19	18	64	2	9	3	89
Equivalency of Follow-up	54	25	5	19	53	6	1	43
Least Bias Comparison	6	58	39	--	7	49	2	45

### The Effectiveness of Drug Courts

The mean weighted odds ratios estimating the effectiveness of drug courts were calculated from studies included in each of the three previous 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 included ( $k = 96$ ) 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 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 there were only two studies that 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 4. Odds Ratios Overall and by CODC Outcome Ratings

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

\*  $p < .01$ , two-tailed.

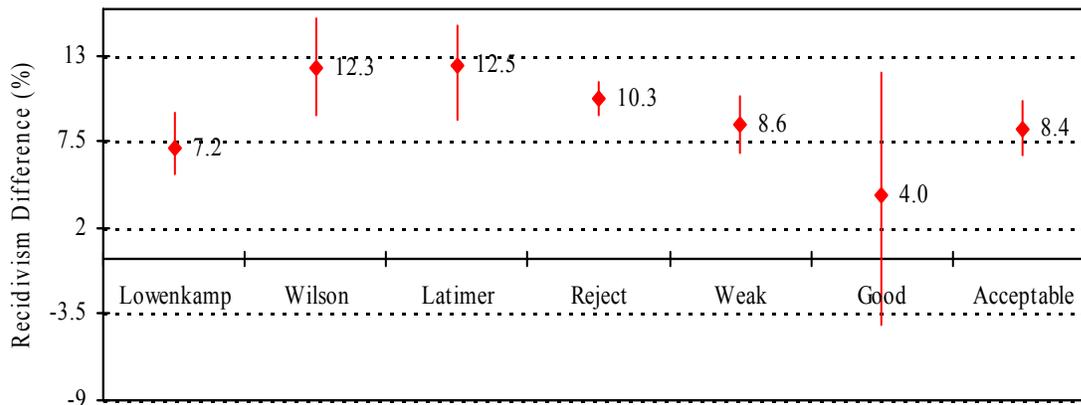
Table 5. Odds Ratios by CODC Global Ratings for Acceptable Studies ( $k = 25$ )

Variable/Factor	Mean OR	95% C. I.		Q	k	N	I <sup>2</sup>
		Low	High				
Acceptable Studies	.711	.660	.766	93.54*	25	14,201	3.90
Confidence:							
Some	.713	.662	.769	92.90*	24	13,966	4.04
High	.531	.259	1.088	—	1	235	—
Quantity of Bias:							
Some	.702	.651	.757	90.84*	24	13,573	3.95
Minimal/Little	.967	.665	1.405	—	1	628	—
Direction of Bias:							
None	.531	.259	1.088	—	1	235	—
Unknown	.705	.650	.764	53.74*	18	11,221	3.16
Increases Treatment	.591	.476	.733	14.73*	5	2,143	3.68
Decreases Treatment	1.836	1.230	2.740	—	1	602	—

\*  $p < .01$ , two-tailed.

After computing the odds ratios, 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%.

Figure 1. Recidivism difference and 95% confidence intervals for three meta-analyses, CODC outcomes and acceptable studies.



### 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 meta-analytic 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 RNR Adherence ( $k = 25$ )

Variable/Factor	Mean OR	95% C. I.		Q	k	N	I <sup>2</sup>
		Low	High				
Acceptable	.711	.660	.766	93.54*	25	14,201	3.90
RNR = 0	.821	.718	.938	38.15*	11	5,511	3.82
RNR = 1	.682	.623	.746	40.57*	13	8,442	3.38
RNR = 2	.306	.179	.523	—	1	248	—
RNR = 3	—	—	—	—	—	—	—
RNR $\geq 1$	.667	.610	.729	48.94*	14	8,690	3.76

\*  $p < .01$ , 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. The present investigation made use of a measure to assess study quality designed for treatment outcome studies with offenders (i.e., CODC Guidelines). 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 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 were 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 pre-existing 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 non-randomized 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, Broom & Stephenson, 2004). 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 pre-existing differences. In addition, such a measure would improve other efforts to control pre-existing differences either through methodological strategies (e.g., subject matching) or post-hoc statistical control procedures (e.g., covariate analysis).

#### *Effects of Drug Treatment Courts: 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 the previous meta-analyses, 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 courts 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.

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

The second component of this study was to assess the relationship between treatment quality and the effectiveness of drug courts. Using only “acceptable” studies, 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 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. 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 favour of treatment. 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 favour 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 increase 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.

#### *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 their adherence to the principles of Risk, Need and Responsivity were obtained solely from the information provided in 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 behaviour.

## References

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

- Andrews, D. A., Bonta, J. & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behaviour*, 17, 19-52.
- Andrews, D. A. & Bonta, J. (2006). *Psychology of Criminal Conduct. 4<sup>th</sup> Edition*. Newark, NJ: LexisNexis.
- \*Anspach, D. F., & Ferguson, A. S. (1999). Cumberland County's drug court program: An evaluation report of Project Exodus. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Drug Courts Program Office.
- \*Anspach, D. F., Ferguson, A. S., & Phillips, L. L. (2003). Evaluation of Maine's Statewide juvenile drug treatment court program: Fourth year outcome evaluation report (Tech. Rep). Portland, ME: University of Southern Maine, Department of Sociology.
- \*Banks, D., & Gottfredson, D. C. (2003). The effects of drug treatment and supervision on time to re-arrest among drug treatment court participants. *Journal of Drug Issues*, 33 (2), 385-414.
- \*Barnoski, R., & Aos, S. (2003). Washington State's drug courts for adult defendants: Outcome evaluation and cost-benefit analysis. (Tech. Rep.). Olympia, WA: Washington State Institute for Public Policy.
- \*Bavon, A. (2001). The effect of the Tarrant County drug court project on recidivism. *Evaluation and program planning*, 24, 13-22.
- Collaborative Outcome Data Committee (2007a). *Sexual Offender Treatment Outcome Research: CODC Guidelines for Evaluation: Part 1: Introduction and Overview*. Ottawa, ON: Public Safety Canada.
- Collaborative Outcome Data Committee (2007b). *The Collaborative Outcome Data Committee's Guidelines for the Evaluation of Sexual Offender Treatment Outcome Research: Part 2: CODC Guidelines*. Ottawa, ON: Public Safety Canada.
- \*Breckenridge, J. F., Winfree, T. L., Maupin, J. R., & Clason, D. L. (2000). Drunk drivers, DWI drug court treatment and recidivism: Who fails? *Justice Research and Policy*, 2 (1), 87-105.
- \*Brewster, M. P. (2001). An evaluation of the Chester County (PA) drug court program. *Journal of drug issues*, 31(1), 177-206.
- Cissner, A. & Rempel, M. (2005). *The State of Drug Court Research: Moving Beyond 'Do They Work?'* New York, NY: Center for Court Innovation.
- Cook, T. & Campbell, D. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.
- \*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.
- \*Craddock, A. (2002). North Carolina drug treatment court evaluation (final report). Washington, DC: U.S. Department of Justice, Office of Justice Programs, Drug Courts Program Office.
- Deeks, J. J., Dinnes, J., D'Amico, R., Sowden, A. J., Sakarovitch, C., Song, F., (2003). Evaluating non-randomized intervention studies. *Health Technology Assessment*, 7, 23-109.
- \*Deschenes, E. P., L. Cresswell, V. Emami, K. Moreno, Z. Klein, and C. Condon. (2001, September). Success of drug courts: Process and outcome evaluations in Orange County, California (Final Report). Submitted to the Superior Court of Orange County, California.
- \*Deschenes, E. P., & Greenwood, P. W. (1994). Maricopa County's Drug Court: An innovative program for first time drug offenders on probation. *Justice System Journal*, 17, 99-115.
- \*Deschenes, E. P., Iman, I., Foster, T., Castellonos, E., Ha, C., Michaels, K., & Ward, D. (2000). Evaluation of Los Angeles county drug courts: 1994-1997 (abstract & executive summary). Richmond, CA: The Center for Applied Local Research.
- \*Deschenes, E. P., I. Imam, T. L., E. Castellonos, T. L. Foster, C. Ha, d. Ward, C., Coley & Michaels, K. (2000). Evaluation of Los Angeles County Drug courts. Richmond, CA: The center for applied local research.
- \*Deschenes, E. P., S. Turner, P. Greenwood, and J. Chiesa. (1996). An experimental evaluation of drug testing and treatment interventions for probationers in Maricopa County, Arizona. Santa Monica, CA: RAND.

- \*Deschenes, E. P., Iman, I., Foster, T. L., Diaz, L., Moreno, V., Patascil, L., & Ward, D. (1999). Evaluation of Orange County Drug Courts (Tech. Rep.). Richmond, CA: The Center for Applied Local Research.
- \*Deschenes E. P., Turner, S., & Greenwood, P. W. (1995). Drug court or probation: an experimental evaluation of Maricopa County drug court. *Justice System Journal*, 18, 55-73.
- \*Dickie, J. L. (2000). Summit County Juvenile Court Drug Court (evaluation report: July 1, 1999-June 30, 2000). Akron, OH: Institute for Health and Social Policy, University of Akron.
- \*Dickie, J. L. (2001). Summit County Juvenile Court Drug Court (evaluation report: July 1, 2000-June 30, 2001). Akron, OH: Institute for Health and Social Policy, University of Akron.
- \*Dickie, J. L. (2002). Summit County Juvenile Court Drug Court (evaluation report: July 1, 2001-June 30, 2002). Akron, OH: Institute for Health and Social Policy, University of Akron.
- \*Ericson, R., Welter, S., Johnson, T. L. (1999). Evaluation of Hennepin County Drug Court (Tech. Rep.). Minneapolis, MN: Minnesota Citizens Council on Crime & Justice.
- Farrington, D. P. (1983). Randomized experiments on crime and justice. In M. Tony & N. Morris, (Eds.), *Crime and Justice* (pp. 257-308). Chicago: University of Chicago Press.
- Farrington, D. (2002). Methodological quality standards for evaluation research. Symposium conducted at the Third Annual Jerry Lee Crime Prevention symposium, Maryland.
- Farrington, D. (2003). Methodological quality standards for evaluation research. *The Annals of the American Academy*, 587, 49-100.
- Farrington, D. P., & Welsh, B. C. (2005). Randomized experiments in criminology: What have we learned in the last two decades? *Journal of Experimental Criminology*, 1, 9-38.
- \*Fielding, J. E., Tye, G., Ogawa, P. L., Imam, I. J., & Long, A. M. (2002). Los Angeles County drug court programs: initial results. *Journal of Substance Abuse Treatment*, 23, 217-224.
- \*Finigan, M. (1998). An outcome program evaluation of the Multnomah county STOP drug diversion program. Portland, OR: Northwest Professional Consortium.
- \*Gliksman, L., Newton-Taylor, B., Patra, J., & Rehm, J. (2004). Toronto drug treatment court evaluation project final report. London, ON: Centre for Addiction and Mental Health, Social, Prevention and Health Policy Research Department.
- \*Godley, M. D., Dennis, M. L., Funk, R., Siekmann, M., & Weisheit, R. (1998). Madison County alternative treatment and court (final evaluation report). Normal, IL: Lighthouse Institute, Chestnut Health Systems.
- \*Goldkamp, J. S. (1994b). Miami's treatment drug court for felony defendants: Some implications of assessment findings. *The Prison Journal*, 73, 110-166.
- \*Goldkamp, J. S., & Weiland, D. (1993). Assessing the impact of Dade County's felony drug court (executive summary). Philadelphia, PA: Temple University, Crime and Justice Research institute.
- \*Goldkamp, J. S., & Weiland, D. (1993). Assessing the impact of Dade County's felony drug court (Final report). Philadelphia, PA: Temple University, Crime and Justice research institute.
- \*Goldkamp, J. S., & Weiland, D., & Moore, J. (2001). The Philadelphia treatment court, its development and impact: The second phase (1998-2000) (Tech. Rep.). Philadelphia, PA: Temple University, Crime and Justice Research Institute.
- \*Goldkamp, J. S., White, M. D., & Robinson, J. B. (2000). Retrospective evaluation of two pioneering drug courts: Phase I findings from Clark County, Nevada and Multnomah county, Oregon (Tech. Rep.). Philadelphia, PA: Temple University, Crime and Justice Research Institute.
- \*Goldkamp, J. S., White, M. D., & Robinson, J. B. (2001b). Do drug courts work? Getting inside the drug court black box. *Journal of Drug Issues*, 31, 27-72.
- \*Goldkamp, J. S., White, M. D., & Robinson, J. B. (2001c). From whether to how drug courts work: Retrospective evaluation of drug courts in Clark County (Las Vegas) and Multnomah county (Portland) – Phase II report from the National Evaluation of Drug courts (Tech. Rep.). Philadelphia, PA: Temple University, Crime and Justice Research Institute.
- \*Gottfredson, D. C., Coblenz, K., & Harmon, M. A. (1996). A short-term outcome evaluation of the Baltimore city drug treatment court program (Tech. Rep). College Park, MD: University of Maryland, Department of Criminology and Criminal Justice.

- \*Gottfredson, D. C., & Exum, M. L. (2002). The Baltimore City Drug Treatment Court: One-year results from a randomized study. *Journal of Research in Crime and Delinquency*, 39, 337-356.
- \*Gottfredson, D. C., & Najaka, S. S., & Kearley, B. (2003). Effectiveness of drug treatment courts: Evidence from a randomized trial. *Criminology and Public Policy*, 2, 171-196.
- \*Gottfredson, D. C., & Najaka, S. S., & Kearley, B., & Rocha, M. (2005). Long-term effects of participation in the Baltimore City drug treatment court: Results from an experimental study: MD: Department of Criminology and Criminal Justice, University of Maryland.
- \*Gottfredson, D. C., Coblenz, K., & Harmon, M. A. (1997). A short-term outcome evaluation of the Baltimore City drug treatment court program. *American Probation and Parole Association: Perspectives*, Winter, 33-38.
- \*Granfield, R., Eby, C., & Brewster, T. (1998). An examination of the Denver Drug Court: The impact of a treatment-oriented drug offender system. *Law & Policy*, 20, 183-202.
- Hanson, R. K. & Bourgon, G. (2008). A psychologically informed meta-analysis of sex offender treatment outcome studies. *The Proceedings of the 2007 North American Correctional & Criminal Justice Psychology Conference*, Ottawa: Public Safety Canada.
- Hanson, R. K., Broom, I., & Stephenson, M. (2005). Evaluating community sex offender treatment programs: A 12-year follow-up of 724 offenders. *Canadian Journal of Behavioural Science*, 36(2), 87-96.
- Hanson, R. K., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W., Quinsey, V. L., & Seto, M. C. (2002). First report of the collaborative outcome data project on the effectiveness of treatment for sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, 14(2), 169-194.
- Helmus, L. (2008). Structured guidelines for evaluating study quality. In G. Bourgon, R. K. Hanson, J.D. Pozzulo, K. E. Morton-Bourgon, & C. L. Tanasichuk, (Eds.) *The Proceedings of the 2007 North American Correctional & Criminal Justice Psychology Conference*, (pp. 45-49). (Corrections Research User Report). Ottawa, ON: Public Safety Canada.
- \*Harrell, A., Roman, J., & Sack, E. (2001). Evaluation of the Brooklyn treatment court: 1996-1999 (Tech. Rep.). Washington, DC: The Urban Institute.
- \*Harrell, A., Roman, J., & Sack, E. (2001). *Drug court services for female offenders 1996-1999: Evaluation of the Brooklyn Treatment Center*. Washington, DC: Urban Institute. Unpublished manuscript.
- \*Harrell, A., S. Cavanagh, and J. Roman (1998). Final Report: Findings from the evaluation of the D.C. Superior Court Drug Intervention program. Washington, DC: Urban Institute.
- \*Harrell, A., Mitchell, O., Merrill, K., & Marlowe, D. (2003). Evaluation of breaking the cycle. Washington, DC: The Urban Institute.
- \*Harrell, A., Mitchell, O., Hirst, D., Marlowe, D., & Merrill, J. (2002). Breaking the cycle of drugs and crime: findings from Birmingham BTC demonstration. *Criminology and Public Policy*, 1(2), 189-216.
- \*Harrison, R. S., Parsons, B. V., Byrnes, E. I., & Sahami, S. (2001). Salt lake county drug court: evaluation report September 1998- September 2000. Salt Lake City, UT: Intermountain evaluation services.
- \*Harrison, L., Patrick, D., & English, K. (2001). *An evaluation of the Denver drug court: The early years, 1995-1996*. Denver, CO: Colorado Department of public safety, Office of research & statistics, Division of Criminal Justice. Unpublished manuscript.
- \*Hartmann, D. J., Rhineberger, G. M., Gregory, P., Mullins, M., Tollini, C., & Williams, Y. (2003). Evaluation of the Kalamazoo County Juvenile Drug Treatment Court program: October 1, 2001-Sept. 30, 2002, year 5 (Tech. Rep). Kalamazoo, MI: Western Michigan University, Kecher Center for social Research.
- \*Huff, D., Stageberg, P, Wilson, B. S., & Moore, R. G. (2002). *An assessment of the Polk county juvenile drug court*. IA: Iowa Department of Human Rights, Division of Criminal and Juvenile Justice Planning. Unpublished manuscript.
- \*Johnson, G. D., Formichella, C. M., & Bowers, D. J. (1998). Do drug courts work? An outcome evaluation of a promising program. *Journal of Applied Sociology*, 15, 44-62.
- \*Johnson, S., & Latessa, E. J. (2000). The Hamilton County Drug Court: Outcome evaluation findings (Tech. Rep.). Cincinnati, OH: University of Cincinnati, Center for criminal justice research.
- \*Johnson, S., Listwan, S., Koetzle Saffer, D., & Latessa, E. J. (2001). *Akron Municipal drug court: Outcome evaluation findings*. Cincinnati, OH: University of Cincinnati. Unpublished manuscript.

- \*Johnson, S., Listwan, S., Koetzle Scaffner, D., & Latessa, E. J. (2001). *Erie county drug court: Outcome evaluation findings*. Cincinnati, OH: University of Cincinnati. Unpublished manuscript.
- Jüni, P., Witschi, A., Bloch, R., & Matthias, E. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *The Journal of the American Medical Association*, *282*, 1054-1060.
- \*Latessa, E. J., Listwan, S. J., Shaffer, D. K., Lowenkamp, C., & Ratansi, S. (2001). Preliminary evaluation of Ohio's drug court efforts. Cincinnati, OH: University of Cincinnati, Center for criminal justice research.
- \*Latessa, E. J., Koetzle Shaffer, D., & Lowenkamp, C. (2002). Outcome evaluation of Ohio's drug court efforts. Unpublished manuscript. Cincinnati, OH: University of Cincinnati.
- Latimer, J., Morton-Bourgon, K., & Chrétien, J. (2006). A Meta-analytic examination of drug treatment courts: Do they reduce recidivism? Ottawa, ON: Department of Justice.
- \*Lind, B., Weatherburn, D., Chen, S., Shanahan, M., Lancsar, E., Haas, M., & De Abreu Lourenco, R. (2002). New South Wales drug court evaluation: Cost-effectiveness. Sydney: NSW Bureau of Crime Statistics and Research.
- \*Listwan, S. J., Shaffer, D. K., & Latessa, E. J. (2003). Effect of drug court programming on recidivism: The Cincinnati experience. *Crime & Delinquency*, *49*(3), 389-441.
- \*Logan, T. K., Hoyt, W., & Leukefeld, C. (2001). Kentucky Drug court outcome evaluation: behaviour costs and avoided costs to society (Tech. Rep.). Lexington, KY: University of Kentucky, Center on Drug and Alcohol Research.
- Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2005). Are drug courts effective: a systematic review. *Journal of Community Corrections*, *3*, 5-10, 28.
- \*Makkai, T., & Verrar, K. (2003). Final report on the South East Queensland drug court. Australia: Australian Institute of Criminology.
- \*Martin, T. (2005). Douglas County drug court targeted capacity expansion grant, year 3 evaluation report. Washington, DC: ISED solutions.
- \*Martin, T. J., Spohn, C.C., Piper, R. K., & Robinson, J. (1999). Phase II Douglas County court evaluation (final report). Omaha, NE: University of Nebraska, Institute for Social and economic development.
- \*Martin, T. J., Spohn, C.C., Piper, R. K., & Frenzel-Davis, E. (2001). Phase III Douglas County Drug court evaluation: Final report. Omaha, NE: Institute for Social and Economic Development.
- Martinson, R. (1979). What works? – Questions and answers about prison reform. *The Public Interest*, *35*, 22-54.
- \*Martinez, A. I., & Eisenberg, M. (2003). Initial process and outcome evaluation of drug courts in Texas (Tech. Rep.). Austin, TX: Criminal Justice Policy Council.
- \*McNeece, C. A., & Byers, J. B. (1995). Hillsborough county drug court two year follow-up study (Tech. Rep.). Tallahassee, FL: Florida State University, Institute for Health and Human Services Research.
- Medina, T., Martinez, F., Meca, J., & Botella, J. (2006). *Assessing heterogeneity in meta-analysis: Q statistic or I2 index?* Unpublished manuscript.
- \*Miethe, T. D., Lu, H., & Reese, E. (2000). Reintegrative shaming and recidivism risks in drug court: Explanations for some unexpected findings. *Crime & Delinquency*, *46*, 522-541.
- \*Miller, M. L., Scocas, E. A., & O'Connell, J. P. (1998). Evaluation of the juvenile drug court diversion program (Tech. Rep.). Washington, DC: Bureau of Justice Assistance.
- \*Millson, W. A., Robinson, D., Stringer, A., & Van Dietan, M. (2005). Drug Treatment court of Vancouver program evaluation: Final evaluation report. Ottawa, ON: Orbis Partners, Inc.
- \*MM Bell Inc. (1998). King county drug court evaluation (final report). Seattle, WA: Author.
- Mohrer, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, P., & Walsh, S. (1995). Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials*, *16*, 62-73.
- \*National Institute of Drug Abuse (2003). Crossing the bridge: an evaluation of the drug treatment alternative-to-prison (DTAP) program. New York, NY: Columbia University, National Center on Addiction and Substance abuse.
- \*O'Connell, J. P., Nestlerode, E., & Miller, M. L. (1999). Evaluation of the Delaware Juvenile drug court diversion program (Tech. Rep.). Dover: DE: Statistical analysis center.

- \*Peters, R. H., & Murrin, M. R. (1998). Evaluation of treatment-based drug courts in Florida's first judicial circuit (Tech. Rep.). Tampa, FL: University of South Florida, Department of Mental health Law & Policy.
- \*Peters, R. H., & Murrin, M. R. (2000). Effectiveness of treatment-based drug courts in reducing criminal recidivism. *Criminal Justice and Behavior*, 27, 72-96.
- \*Peters, R. H., & M. R. Murrin (1998). Evaluation of treatment-based drug courts in Florida's first judicial circuit. Tampa, FL: Department of Mental Health and Policy, Louis de la Parte Florida Mental Health Institute, University of South Florida.
- \*Rodriquez, N., & Webb, V. J. (2004). Multiple measures of juvenile drug court effectiveness: Results of a quasi-experimental design. *Crime & Delinquency*, 50 (2), 292-314.
- \*Roehl, J. (1998). Monterey County drug court (evaluation report No. 1). Pacific Grove, CA: Justice Research Center.
- \*Salt Lake county. (2001). Salt Lake County drug court outcome evaluation. UT: Utah substance abuse & anti-violence coordinating council.
- \*Santa Clara County Courts. (1998). Santa Clara county drug treatment court (3<sup>rd</sup> progress report, March 1, 1996-March 31, 1997). Santa Clara, CA: Author.
- \*Santa Clara County. (1998). Santa Clara county drug treatment court: Two year progress report and outcome comparisons (March 1, 1996-March 31, 1998). Santa Clara, CA: Author.
- Seager, J. A., Jellicoe, D., & Dhaliwal, G. K. (2004). Refusers, Dropouts, and completers: measuring sex offender treatment efficacy. *International Journal of Offender Therapy and Comparative Criminology*, 48(5), 600-612.
- \*Sechrest, D. K., Shichor, D., Artist, K., & Briceno, G. (1998). Riverside county drug court: final research report for the riverside county probation department, riverside county, California (Tech. Rep.). San Bernardino, CA: California State University, Criminal Justice Department.
- \*Shanahan, M., Lancsar, E., Haas, M., Lind, B., Weatherburn, D., & Chen, S. (2004). Cost-effectiveness analysis of the New South Wales adult drug court program. *Evaluation Review*, 28, 3-27.
- Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J. E., Reuter, P., & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising*. Washington, DC: Office of Justice Programs.
- Simpson, K. (2008). What We Know and Where We Need to Go for Effective Community Supervision. Presentation at the 2008 Canadian Psychological Association Convention. Halifax, NS.
- \*Solop, F. I., Wonders, N. A., Hagen, K. K., & McCarrier, K. (2003). Coconino county DUI/Drug court evaluation. AZ: Social Research Laboratory, Northern Arizona University.
- \*Spohn, C., Piper, R. K., Martin, T., & Frenkel, E. D. (2001). Drug courts and recidivism: The results of an evaluation using two comparison groups and multiple indicators of recidivism. *Journal of Drug Issues*, 31, 149-176.
- \*Stageberg, P., Wilson, B., & Moore, R. G. (2001). Final report on the Polk County Adult Drug court (Tech. Rep.). Iowa Department of Human Rights, Division of Criminal and Juvenile Justice Planning.
- \*Tauber, J. S. (1993). The importance of immediate and intensive intervention in a court-ordered drug rehabilitation program (Tech. Rep.). Philadelphia, PA: The president's commission of model state drug laws.
- \*Tauber, J. S. (1995). *An evaluation of the Oakland county drug court after three years: a drug diversion program of the Oakland-piedmont-Emeryville municipal court and the Alameda county probation department*. Paper presented at the National Association of Drug court professionals, national training conference, Las Vegas, NV.
- \*Terry, W. C., III. (1995). Report offences of the first year cohort of Broward County, Florida's drug court (Tech. Rep.). Miami, FL: Florida International University.
- \*Terry, W. C., III. (1999). Broward County's dedicated drug treatment court: From post adjudication to diversion. In Terry, W. C., III (Ed.), *The early drug courts: Case studies in judicial innovation*. Drugs, health, and social policy series (Vol. 7, pp.77-107), Thousand Oaks, CA: Sage Publications, Inc.
- Thomas, H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing*, 2, 91-99.
- \*Thompson, K. M. (2002). A cost-benefit estimate of North Dakota's juvenile drug court: Recidivism cost savings. ND: North Dakota State University.

- \*Truitt, L., Rhodes, W. M., Hoffman, N. G., Seeherman, A. M., Jalbert, S. K., Kane, M., Bacani, C. P., Carrigan, K. M., & Finn, P. (2002). Evaluating treatment drug courts in Kansas City, Missouri and Pensacola, Florida: final report for Phase I and Phase II (Tech. Rep.). Cambridge, MA: ABT Associates Inc.
- \*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*. Washington, D.C.: U.S. Department of Justice: National Drug Court Institute.
- \*Utah Substance Abuse & Anti-Violence Coordinating council. (2001). Salt Lake County Drug Court: Outcome evaluation. Salt Lake City, UT: Author
- \*Vito, G. F., & Tewksbury, R. A. (1998). Jefferson County Drug Court Program: Impact evaluation, 1997 (Tech. Rep.). Louisville, KY: University of Louisville.
- \*Vito, G. F., & Tewksbury, R. A. (1998b). The impact of treatment: The Jefferson County (Kentucky) drug court program. *Federal Probation*, 62, 46-51.
- Weekes, J., Mugford, R., Bourgon, G., & Price, S. (2007). *Drug Treatment Courts FAQs*. Canadian Centre for Substance Abuse. Retrieved May 1, 2007 from <http://www.ccsa.ca/NR/rdonlyres/FFBA90ED-2E2F-408D-A6C9-4F9E9F9B9155/0/ccsa00113482007.pdf>
- Westergren, A., Karlsson, S., Andersson, P., Ohlsson, O. & Hallberg, I. (2001). Odds Ratio. *Journal of Clinical Nursing*, 10, 257-269.
- Wexler, D. & Winick (1991). Therapeutic jurisprudence as a new approach to mental health law and policy analysis and research. *University of Miami Law Review*, 975.
- Wilson, D. B., Mitchell, O., & Mackenzie, D. L. (2006). A systematic review of drug court effects on recidivism. *Journal of Experimental Criminology*, 4, 459-487.
- \*Wolfe, E., Guydish, J., & Termond, J. (2002). A drug court outcome evaluation comparing arrests in a two year follow-up period. *Journal of Drug Issues*, 32, 1155-1172.
- Wormith, J.S., Althouse, R., Simpson, M., Reitzel, L.R., Fagan, T.J. & Morgan, R.D. (2007). The rehabilitation and reintegration of offenders: The current landscape and some future directions for correctional psychology. *Criminal Justice and Behaviour*, 34, 879-892.
- \*Wright, D., & Clymer, B. (2001). Evaluation of Oklahoma drug courts, 1997-2000. Oklahoma city, OK: Oklahoma Criminal Justice Resource Centre.
- \*Wright, D., & Clymer, B. (2001). *Beckham county juvenile drug court: Phase II analysis and evaluation*. Oklahoma City, OK: Oklahoma Statistical Analysis Center. Unpublished manuscript.
- \*Wright, D., Clymer, B., & Simpson, D. (2001). *Evaluation of Oklahoma drug courts*. Oklahoma City, OK: Oklahoma Statistical Analysis Center. Unpublished manuscript.