

12-11-2009

Reducing Juvenile Recidivism: A Meta-Analysis of Treatment Outcomes

Heidi R. Tennyson
Pacific University

Recommended Citation

Tennyson, Heidi R. (2009). Reducing Juvenile Recidivism: A Meta-Analysis of Treatment Outcomes (Doctoral dissertation, Pacific University). Retrieved from:
<http://commons.pacificu.edu/spp/109>

This Dissertation is brought to you for free and open access by the Theses, Dissertations and Capstone Projects at CommonKnowledge. It has been accepted for inclusion in School of Professional Psychology by an authorized administrator of CommonKnowledge. For more information, please contact CommonKnowledge@pacificu.edu.

Reducing Juvenile Recidivism: A Meta-Analysis of Treatment Outcomes

Abstract

Juvenile criminal activity continues to be a problem in the United States both in terms of its financial burden to society and its impact on quality of life. One adolescent repeat offender may cost tax payers an estimated 1.3 to 1.5 million dollars (Cohen, 1998). Thus, there is an imperative to identify treatments that decrease youthful re-offending. The present meta-analysis analyzed which interventions had the largest effects on decreasing recidivism, and explored in a unique way whether quality of treatment implementation increased treatment efficacy in real-world settings. All programs analyzed were effective in reducing juvenile recidivism except those focused on discipline (i.e., boot camps). Programs offering multiple services were the most effective. In addition, interventions with the highest level of treatment integrity had the strongest outcomes. Finally, researcher-driven studies had larger effects than community-based programs indicating a continued gap between research and practice. The importance of integrity in real-world settings is highlighted in the discussion.

Degree Type

Dissertation

Rights



This work is licensed under a [Creative Commons Attribution 3.0 License](http://creativecommons.org/licenses/by/3.0/).

Copyright and terms of use

If you have downloaded this document directly from the web or from CommonKnowledge, see the “Rights” section on the previous page for the terms of use.

If you have received this document through an interlibrary loan/document delivery service, the following terms of use apply:

Copyright in this work is held by the author(s). You may download or print any portion of this document for personal use only, or for any use that is allowed by fair use (Title 17, §107 U.S.C.). Except for personal or fair use, you or your borrowing library may not reproduce, remix, republish, post, transmit, or distribute this document, or any portion thereof, without the permission of the copyright owner. [Note: If this document is licensed under a Creative Commons license (see “Rights” on the previous page) which allows broader usage rights, your use is governed by the terms of that license.]

Inquiries regarding further use of these materials should be addressed to: CommonKnowledge Rights, Pacific University Library, 2043 College Way, Forest Grove, OR 97116, (503) 352-7209. Email inquiries may be directed to: copyright@pacificu.edu

REDUCING JUVENILE RECIDIVISM:
A META-ANALYSIS OF TREATMENT OUTCOMES

A DISSERTATION
SUBMITTED TO THE FACULTY
OF
SCHOOL OF PROFESSIONAL PSYCHOLOGY
PACIFIC UNIVERSITY
HILLSBORO, OREGON

BY
HEIDI R. TENNYSON
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
OF
DOCTOR OF PSYCHOLOGY

DECEMBER 11, 2009

APPROVED BY THE COMMITTEE:
Alyson Williams, Ph.D.

Paul Michael, Ph.D.

PROFESSOR AND DEAN:
Michel Hersen, Ph.D.

ABSTRACT

Juvenile criminal activity continues to be a problem in the United States both in terms of its financial burden to society and its impact on quality of life. One adolescent repeat offender may cost tax payers an estimated 1.3 to 1.5 million dollars (Cohen, 1998). Thus, there is an imperative to identify treatments that decrease youthful re-offending. The present meta-analysis analyzed which interventions had the largest effects on decreasing recidivism, and explored in a unique way whether quality of treatment implementation increased treatment efficacy in real-world settings. All programs analyzed were effective in reducing juvenile recidivism except those focused on discipline (i.e., boot camps). Programs offering multiple services were the most effective. In addition, interventions with the highest level of treatment integrity had the strongest outcomes. Finally, researcher-driven studies had larger effects than community-based programs indicating a continued gap between research and practice. The importance of integrity in real-world settings is highlighted in the discussion.

Keywords: meta-analysis, juvenile, recidivism, treatment, integrity

ACKNOWLEDGMENTS

The completion of this project represents a benchmark in a long, eventful, and interesting journey that began many years ago as a desire to find effective means of reducing recidivism in criminal populations. Many people have helped, inspired, and supported me along the way and to all of them I am very grateful.

First and foremost I would like to acknowledge my dissertation chair, Dr. Alyson Williams, who helped clarify my ideas and assisted in the conceptualization of this project. Without her support I would not have considered taking on such a formidable task, nor would I have had the courage to attempt a meta-analytic study. Her unwavering confidence in my ability to finish this project helped me see it through to completion. I am very appreciative of her encouragement and gentle nudges to keep moving forward.

I am also very grateful to Dr. Paul Michael whose knowledge of meta-analytic methods kept me on track and without whom this document would not be nearly as strong. I appreciated his warm willingness to answer questions and ability to take abstract and complex topics and explain them in simple and concrete terms. His guidance on this project was invaluable and I am very thankful for his support.

I owe a huge debt of gratitude to my fellow coder on this project, Casey Wixson, who spent countless hours coding articles during both the training process and final reliability study. His many constructive and well-thought-out suggestions also contributed to the integrity of the final version of the coding manual. The strong outcome

of the reliability study is a tribute to his commitment to this project and I was grateful to have his assistance on this challenging task.

I am particularly thankful for the many mentors, friends, and family members who have inspired and encouraged me during this sometimes difficult, but mostly thrilling journey. They have all enriched my experiences and no words adequately capture my appreciation for them. Above all, I am grateful to my parents who have been steadfast sources of support in my life. Their guidance illuminated my way during the many twists and turns of life and allowed me to find sure footing. They always have been and always will be my inspiration.

TABLE OF CONTENTS

	Page
ABSTRACT.....	ii
ACKNOWLEDGMENTS.....	iii
LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
INTRODUCTION.....	1
REVIEW OF LITERATURE ON JUVENILE RECIDIVISM.....	4
METHOD.....	23
RESULTS.....	43
DISCUSSION.....	60
REFERENCES.....	73
APPENDIX	
A. Coding Manual.....	80

LIST OF TABLES

	Page
TABLE 1. Sample Frequencies and Interrater Reliability.....	44
for Treatment and Comparison Groups	
TABLE 2. Treatment Frequencies and Interrater Agreement.....	47
for Treatment and Comparison Groups	
TABLE 3. Integrity Frequencies and Interrater Agreement.....	50
for Treatment and Comparison Groups	
TABLE 4. Research Design Frequencies and Interrater Agreement.....	51
TABLE 5. Homogeneity Analyses for Treatment Groups.....	58
and Treatment Integrity Comparisons	

LIST OF FIGURES

	Page
FIGURE 1. Effect Sizes for Treatment Types as Grouped by Lipsey's (2009) Categories with 95% Confidence Intervals Represented as Lines Around the Effect Size Box.	54
FIGURE 2. Effect Sizes for Treatment Integrity Types with 95% Confidence Intervals Represented as Lines Around the Effect Size Box.	55
FIGURE 3. Effect Sizes for Research-Driven and Real-World Integrity Types with 95% Confidence Intervals Represented as Lines Around the Effect Size Box.	57

REDUCING JUVENILE RECIDIVISM:

A META-ANALYSIS OF TREATMENT OUTCOMES

Crime in the United States takes its toll both in terms of monetary cost and lost quality of life. In fact, according to one estimate the total cost of crime in the U.S. currently exceeds 1 trillion dollars annually (Anderson, 1999). This calculation includes medical expenses, lost earnings, and other services for victims as well as the intangible emotional costs of lost quality of life. The FBI Uniform Crime Reports show that juvenile offenders accounted for only 15% of violent and 30% of property crimes in the U.S in 2002 (FBI, 2004). However, a 1998 study estimated the cost to society for just one juvenile repeat offender at 1.3 to 1.5 million dollars (Cohen, 1998). These figures include the cost to the victim, loss in productivity of the offender, and criminal justice, drug treatment, and medical expenses.

Although adult offenders account for the majority of criminal activity in the United States, about 25 percent of juvenile offenders over the age of 16 will go on to re-offend in their early adult years (Office of Juvenile Justice and Delinquency Prevention, 2006). Thus, if effective interventions can be identified and implemented that target juvenile offenders, a decrease in the number of adolescents that re-offend as adults should follow. Ideally, this would result in a better quality of life for both the young offender and society at large. One method of identifying whether an intervention is effective is to look for a decrease in juvenile recidivism after an intervention has been administered. Recidivism, has been defined as the “repetition of delinquent or criminal behavior,

especially in the case of a habitual criminal, or repeat offender, who has been convicted several times” (VandenBos, 2007, P. 776).

During the 1970s numerous studies were conducted which demonstrated poor outcomes in terms of juvenile rehabilitation. However, in subsequent years studies have shown that a variety of interventions result in varying degrees of improvement with respect to reduced recidivism (Hollin, 1999). Some of these interventions include individual therapy, family therapy, parent training, group treatment, drug treatment, restitution, correctional programs, and multisystemic therapies. The meta-analysis is a means of examining multiple studies and a variety of treatment modalities to arrive at a comprehensive and quantitative review of the literature. Although a large meta-analysis, examining the effect sizes of different treatments in terms of juvenile recidivism, was carried out spanning the years 1958 to 2002 (Lipsey, 2009), no new meta-analysis has been conducted since that time examining the effects of a variety of newer treatments and studies. As up to 30 new studies examining treatments aimed at reducing juvenile reoffending have been carried out since 2002, a new meta-analysis of the literature is indicated.

The present study aims to examine the current characteristics of treatment for juvenile offenders, analyze program effectiveness, and examine which interventions are most successful in decreasing adolescent recidivism. In addition, this quantitative review will examine whether quality of treatment implementation (i.e., treatment integrity) increased treatment efficacy in community settings. The last meta-analytic review of this type indicated that the inability to capture quality of treatment implementation in real-

world venues was a limitation of the study (Lipsey, 2009). The present analysis aims to overcome that weakness by introducing a novel way of examining treatment integrity at community sites.

The next chapter will review the literature pertinent to treatment interventions aimed at reducing youthful reoffending. It will also review meta-analytic techniques and describe their use in the current literature. Finally treatment integrity will be explored as it has been defined in recent studies. The following chapter will lay out methods for the present quantitative review including research questions, procedures, and analytic strategy. Subsequent chapters will describe results of both demographic and descriptive data as well as discuss the findings specific to each research question. Strengths of the project will be analyzed. Specifically, treatment integrity as measured in real-world settings will be highlighted. Limitations of the project will also be examined and future directions for research explored.

LITERATURE REVIEW

To arrive at a current understanding of the literature on recidivism, treatment approaches shown to be ineffective as well as those shown to be effective in reducing reoffending will be explored. Within the context of treatments generally shown to be unsuccessful in reducing recidivism rates, intensive supervision probation, boot camps, and Scared Straight programs will be analyzed. With respect to approaches typically found to be effective in decreasing reoffending, seven types of treatment interventions aimed at reducing recidivism in juvenile offenders will be examined. These treatment interventions include restorative justice, parent training, drug treatment, behavior modification and cognitive behavioral treatment (CBT), as well as family, group, and multisystemic therapies. A combination of literature reviews and meta-analyses will be explored. In addition two meta-analyses from 1992 and 2009 that examined differences in effectiveness for multiple interventions are reviewed here as well. Further, treatment integrity as it relates to meta-analytic literature will be examined. In addition, a debate in the literature with respect to whether treatment conducted in research-driven (i.e., efficacy) studies can be effectively implemented in real-world settings will be explored. Finally, limitations in the current literature will be highlighted. Specifically, a shortcoming in the literature around implementation of quality assurance which restricted its examination to research-driven studies will be assessed and the need for a novel approach to examining treatment integrity in real-world settings will be considered.

Historically Ineffective Interventions

This section provides a review of the literature on programs that have produced negligible or negative results in reducing recidivism. To start, programs based on punishment or coercion have shown little effect in reducing recidivism (McGuire & Priestley, 1995). In fact, these types of programs can actually demonstrate adverse effects (Lipsey, 1992). Lipsey (1992) found that punishment-based programs resulted in up to a 25 percent increase in recidivism. Examples of programs that have typically performed poorly in terms of reduced recidivism rates include intensive parole, military boot camps, and Scared Straight programs. Parole oriented interventions are aimed at increasing monitoring of the offender, while boot camps are based on the idea that harsh and rigorous regimens will deter future criminal behavior. Scared Straight programs involve exposing juveniles to adult prison facilities in an attempt to discourage them from criminal activity.

Intensive Supervision Probation

Intensive supervision probation is designed to increase surveillance of young offenders. It can include house arrest, electronic monitoring, and other types of restrictions. A review of the literature on intensive monitoring of juveniles during probation suggests that this practice does little to decrease recidivism (Lipsey, 2009). Evidence suggests that increased monitoring actually leads to a greater number of technical violations (Giblin, 2002). However, it is surmised that this is in part due to increased surveillance, as probation officers are more likely to become aware of violations with increased contact. Although this type of intervention seems to lead to

more technical violations, there is some evidence that it decreases the number of new offenses (Giblin, 2002).

Boot Camp

Boot camps typically involve confrontations between staff and inmates as the demanding schedule of a military-style basic training is enforced. In a meta-analysis that examined boot camp as one type of intervention, 66 studies were analyzed to examine the effects of five types of incarceration-based drug treatment programs in reducing recidivism (Mitchell, Wilson, & MacKenzie, 2007). The five types of treatment included a narcotics maintenance program, group counseling, residential treatment, a therapeutic community, and boot camp. The researchers found that therapeutic communities, residential treatment, and group counseling showed significant results in reducing recidivism. In contrast, narcotics maintenance programs showed mixed results, while no reduction in recidivism was found for boot camps.

Scared Straight

Scared Straight programs began in the 1970s as an inexpensive means of deterring adolescents from a life of crime. It was based on the premise that young people exposed to the harsh realities of prison life would be dissuaded against law-breaking activities. The first “Juvenile Awareness Program” was developed in New Jersey and following a documentary which proclaimed its success, caught on nationwide (Petrosino, Turpin-Petrosino, & Finckenauer, 2000). However, there was little empirical evidence to support its widespread implementation. In a meta-analysis designed to examine the effects of these types of deterrence program, 150 studies were examined that analyzed

interventions that exposed juveniles to either reformatories or prisons. Results indicated that these types of interventions actually increase the rate of re-offending from 1% to 30% (Petrosino, Turpin-Petrosino, & Finckenauer, 2000)

Historically Effective Interventions

Meta-analyses applied to recidivism research have revealed several types of treatment interventions that are effective at reducing recidivism (Lipsey, 1992, 2009). Specifically, results from Lipsey's (2009) last large quantitative review demonstrated the following: Restorative justice programs showed small effects in reducing reoffending. Behavioral and cognitive-behavioral programs aimed at teaching offenders how to problem solve, cope effectively, and interact successfully demonstrated some success in reducing recidivism. Skill building programs of this type could include parent training and drug treatment programs. Counseling-based treatments in an individual, family, or group format tended to show the strongest results in reducing juvenile reoffending. Finally, multisystemic therapies that were community-based, took place within the offender's home environment, and addressed a variety of factors that influence the likelihood of re-offending also showed success in reducing recidivism. The following sections provide further information about the effectiveness of these types of programs as reported in meta-analytic literature.

Restorative Justice

Restorative justice practices are increasingly being implemented in the criminal justice system (Latimer, Dowden, & Muise, 2005). This practice brings victims and offenders together to collectively determine how an offense can best be resolved. The

aim is for the offender to take responsibility for his or her actions while the victim has the opportunity to state how the offense can most appropriately be repaired (Latimer, Dowden, & Muise, 2005). The following meta-analysis indicates that restorative justice practices are effective in reducing recidivism. A meta-analysis was conducted to examine the effects of restorative justice programs (Latimer, Dowden, & Muise, 2005). The following operational definition was used to determine whether a study met the qualifications of a restorative justice approach: “Restorative justice is a voluntary, community-based response to criminal behavior that attempts to bring together the victim, the offender, and the community, in an effort to address the harm caused by the criminal behavior” (p. 131). Outcome measures included victim and offender satisfaction, restitution compliance, and recidivism. The authors found that compared to traditional criminal justice approaches such as incarceration and probation, restorative justice practices were more effective in terms of all three outcome measures.

Parent Training

Parent training is a therapeutic technique designed to impact parental discipline styles as a means of decreasing adolescent delinquent behavior (Perkins-Dock, 2001). For example, inconsistent parenting practices, harsh discipline, inadequate supervision, and poor boundaries have all been shown to be risk factors for later delinquent behavior in children. Parent-training programs are aimed at assisting parents in developing more effective parenting skills. They often involve teaching parents how to positively reinforce adolescents’ prosocial behavior with increased praise and attention and decrease problem behaviors with less criticism and appropriate disciplinary practices.

A meta-analysis of 71 studies was conducted to examine the effects of both behavioral parent-training (BPT) and CBT on outcomes for youth with antisocial behavior problems (McCart, Priester, Cavies, & Azen, 2006). To be included in the meta-analysis studies had to either focus on a behavioral parent-training program or a CBT intervention and target antisocial behavior such as aggression or delinquency. Further, the youth that were the target of treatment had to be 18 years old or younger. The authors found that differences in outcome for the two types of intervention were moderated by age. That is, for children 6 to 12 years of age the BPT interventions had stronger effects, while CBT programs showed stronger effects for older adolescents. The authors suggest that the results make sense in terms of developmental trajectories. For example, younger children tend to look more to their parents for guidance, while adolescents are beginning to seek more independence. Thus, interventions aimed at effecting parenting skill are likely to have the most impact on decreasing recidivism in younger children.

Behavior Modification and CBT

Although implemented within punishment-based settings, results from meta-analyses conducted over the past 20 years demonstrate that treatment programs implemented in correctional facilities can be effective (Pearson, Lipton, Cleland, & Yee, 2002). The results have shown that this type of treatment does reduce recidivism rates. However, it remains unclear exactly which programs are the most effective for reducing recidivism in young offenders.

A meta-analysis of 69 studies was conducted to examine the effects of behavioral and cognitive behavioral treatments on the reduction of recidivism in both adult and

juvenile populations (Pearson, Lipton, Cleland, & Yee, 2002). To be included in the meta-analysis studies had to examine either a behavior modification program or a cognitive-behavioral treatment. Participants were in custody in a correctional facility or were on probation or parole. Outcome measures included drug use and recidivism. The authors found that treated groups showed approximately a 30% reduction in recidivism over untreated groups, but that cognitive behavioral therapy programs were more effective in reducing recidivism than behavioral programs. Although cognitive behavioral programs were shown to be most effective, this study did not address which aspects of cognitive behavioral treatments have the greatest effects on the reduction of recidivism.

In an investigation aimed at overcoming the weaknesses in Pearson et al., (2002) study, Landenberger and Lipsey (2005) conducted a meta-analysis of 58 studies focused exclusively on cognitive behavioral treatments and their effects on the reduction of recidivism in both adult and juvenile populations. To be included in the meta-analysis studies had to be cognitive-behavioral in orientation and similar to recognized programs such as “Aggression Replacement Training” (Goldstein & Glick, 1987). Participants included offenders from general populations who received treatment either while on probation, during incarceration, or through an aftercare program. Recidivism as an outcome measure included re-arrest, reconviction, and incarceration at approximately 12 months post-treatment. The authors found that compared to untreated control groups, there was a 25% decrease in recidivism for offenders who received a CBT treatment. In addition, it did not appear to matter what type of CBT program was used as long as CBT

techniques were well implemented. However, an examination of moderators showed that treatments that included anger control and interpersonal problems solving components were associated with larger effects. The authors also noted that these interventions were just as effective with juveniles as they were with adults.

Drug Treatment

The re-arrest rate for drug offenders in the U.S. in 1994 was 66.7% (Bureau of Justice Statistics, 1994). According to Van Wormer (2003) it is time to implement effective drug and alcohol treatment programs in the prisons because there are far more drug and alcohol abusers in the correctional system than there are currently enrolled in substance abuse treatment programs. In a review of different types of treatment Loxley (2005) examined diversion programs within the criminal justice systems in the United States, Australia, and United Kingdom. In the U.S. participants were court-ordered to treatment, in Australia they were either court-ordered to an education program or treatment, and in the U.K., they were either referred for drug treatment by drug workers or court-ordered to treatment. Loxley (2005) found that the U.S. programs were effective in reducing drug use and criminal behavior. In Australia the programs were found to reduce recidivism, while the programs in the U.K. have not yet shown statistical significance.

Family Interventions

Family therapy interventions are aimed at influencing familial dynamics as one means of addressing criminal behavior in adolescents (Perkins-Dock, 2001). Specific types of family dynamics have been shown to influence delinquent behavior (Perkins-

Dock, 2001). For example, studies show that having a delinquent sibling increases the likelihood of being convicted for a violent offense. Single-parent households also increase the likelihood of violent behavior in adolescents. Experiences of neglect, abuse, and harsh parental discipline also increase the chances for adolescent criminality. Moreover, living below the poverty line particularly influences the likelihood of criminal behavior. One contributing factor is undoubtedly that economic strains mean parents spend more time working outside the home and less time involved with their children thus contributing to delinquency in youth. In a review of the literature, several different types of family therapy were explored. Functional Family Therapy includes concepts based on systems theory and attempts to decrease negative behavior and increase positive interpersonal communication between family members (Sexton & Alexander, 2000). Brief strategic family therapy focuses on improving relationships within the youth's family system as a means of impacting positive behavior change (Perkins-Dock, 2001). One-person family therapy is based on the systemic idea that a change in one family member will lead to corresponding changes in other family members and aims to modify behavior in the adolescent offender as a means of affecting family dynamics in a positive manner (Robbins & Szapocznik, 2000). Meta-analyses conducted in the 1990s indicate that a family systems approach is one of the most effective types of family interventions in decreasing adolescent criminal behavior (Perkins-Dock, 2001).

Group Therapy

According to Wilson, Bouffard, and Mackenzie (2005), because individual therapy is no longer economically feasible in correctional settings, most treatment

interventions for offenders are delivered in groups of approximately 8 to 10 individuals. Studies indicate that cognitive-behavioral-group-oriented programs are effective in reducing recidivism in offender populations (Wilson, Bouffard, and Mackenzie, 2005).

A meta-analysis of 20 studies was conducted to examine the effects of group-oriented cognitive behavioral treatment for juvenile offenders (Wilson, Bouffard, & Mackenzie, 2005). To be included in the meta-analysis studies had to be CBT in orientation and designed to reduce criminal behavior with a focus on cognitive restructuring and the development of life skills and moral reasoning. Participants included youth currently incarcerated, on probation, on parole, or referred by the criminal justice system at the time of treatment. Recidivism as an outcome measure included new arrests and convictions post-treatment. The authors found that compared to untreated control groups, offenders who received a group CBT treatment were less likely to recidivate by 16 percentage points compared to untreated offenders. Specifically, group-based cognitive behavioral therapies that emphasize cognitive restructuring and moral reasoning showed positive reductions in recidivism.

Multisystemic Therapy

Multisystemic therapy (MST) is aimed at decreasing delinquent behavior in youth through both community and home-based interventions (Curtis, Ronan, & Borduin, 2004). MST targets the adolescent as well as his or her family, peer group, school, and community. Meta-analyses indicate that MST is highly effective in treating juvenile delinquent behavior.

A meta-analysis of 10 studies focused on the delivery of multisystemic therapy (MST) to youth who either had antisocial or psychiatric symptoms themselves or had a parent with such symptoms (Curtis, Ronan, & Borduin, 2004). To be included in the meta-analysis studies had to use an approach that adhered to MST principles. Participants were predominantly male youth ages 8 to 17 who were classified as chronic, at-risk, and/or juvenile offenders. Outcome measures included rate of criminal activity, days incarcerated, absence from school, amount of drug use, and out-of-home placement. MST was compared to both parent training and individual therapy. The authors found that adolescents and their families that were treated with MST had better outcomes and that these adolescents were offending less than 70% of the youth treated with alternative treatments. In addition, MST was shown to be effective in reducing youth criminality up to four years post treatment. Another important outcome was a demonstration of strong effects around family relations post MST treatment. The authors suggest that this is consistent with the emphasis placed on family interventions in MST.

The positive and negative results summarized above have come largely from meta-analysis, a technique used to summarize quantitatively a large number of findings. This technique is well-suited for use in updating the current juvenile recidivism literature in increasingly sophisticated ways. Below is a description of meta-analysis as a method and its utility in understanding juvenile recidivism.

Meta-analysis

Meta-analysis is a statistical method that measures the size of a relationship between two or more variables across a collection of studies (Glass, McGraw, & Smith,

1981, cited in Latimer, Dowden & Muise, 2005). The outcome measure in a meta-analysis is the effect size, which describes the strength of a relationship between two variables. Moreover, the meta-analysis allows for a quantitative evaluation considered more rigorous than a narrative literature review. Two large meta-analyses on juvenile recidivism have been carried out over the last fifteen years (Lipsey, 1992; Lipsey, 2009).

Hollin (1999) suggested that the meta-analysis has moved the field of juvenile offender treatment out of the realm of “nothing works” and into domain of “what works.” He commented that the “nothing works” doctrine fit well within the conservative political climate of the 1970s and 1980s, which leaned strongly toward punishment rather than rehabilitation. Although studies were conducted in the 70s and 80s that showed some success in treating offenders, Hollin noted that it was difficult to identify what worked because too many distinct interventions, in varied settings, with different measures of success were examined. However, Hollin stated that the meta-analysis assists in overcoming this problem. Further, he asserted that one of the most important outcomes of applying the meta-analysis to offender treatment studies has been the capacity to compare a large number of treated groups to non-treated groups. The result has been to see a 10% decrease in re-offending when the overall treatment effect is examined. Further, he notes that when different types of interventions are compared through a meta-analysis, some interventions result in up to a 20% decrease in re-offending. According to Hollin, the meta-analysis has shown that treatment interventions do decrease recidivism in offender populations, and he states that the next task is to apply the meta-analysis to determining which treatments are most effective.

Meta-analysis, Juvenile Offenders, and Recidivism

In one of the largest analyses examining the effects of different types of treatment programs on juvenile recidivism, Lipsey (1992) conducted a meta-analysis of 400 studies. Results from this meta-analysis challenged the notion that “nothing works” in juvenile offender treatment. Lipsey found that offender treatment resulted in decreased recidivism, with behaviorally oriented treatments showing the strongest effects. In fact, compared to control groups there was a 20 percent decrease in recidivism for offenders who received behavioral treatments.

In a follow-up study, Lipsey (2009) included his prior research in a meta-analysis that spanned the years 1958 to 2002. In this analysis comprised of 548 studies, Lipsey identified factors associated with treatment efficacy and examined how those interventions that were effective compared to one another. Lipsey also developed a means of categorizing treatment modalities that distinguished between interventions aimed at engaging youth in collaborative processes of change versus approaches focused on external control and coercion. Results showed that the former interventions were more effective than the latter. In addition, he advanced a means of categorizing the quality with which a treatment program was implemented. First, he identified whether there was initial difficulty with implementation (e.g., a large number of dropouts or high staff turnover). Second, he quantified how closely the researcher was involved with treatment implementation on a four-point scale. Results demonstrated that the higher the quality of implementation, the greater the effectiveness of the intervention.

Treatment Integrity

Treatment integrity is currently considered one of the most critical aspects of treatment outcome research (Perepletchikova & Kazdin, 2005). Treatment integrity can be defined as the implementation of an intervention as it was intended to be carried out (Perepletchikova & Kazdin, 2005). High levels of treatment integrity are associated with increased program efficacy. However, in a recent analysis of 202 studies, Perepletchikova, Treat, and Kazdin (2007) found that fewer than 10% of these studies adequately reported treatment integrity. The authors identified three components of treatment integrity including treatment adherence, therapist competence, and treatment differentiation. They defined treatment adherence as the degree to which procedures specified for the intervention are utilized by the therapist. Therapist competence was characterized as both skill level and judgment exemplified by the therapist in delivering services. Finally, treatment differentiation was referred to as the extent to which interventions under analysis differ along important dimensions.

Lipsey (1992) noted that treatment categories in meta-analytic reviews are often approximations due to limited and inconsistent reporting across studies. Thus, finding effective means of grouping treatment types to understand what we know about recidivism becomes particularly salient. In his most recent meta-analysis, Lipsey (2009) characterized his “quality of implementation” variable as a crude composite of two correlated features. Indirectly he was able to roughly examine both treatment adherence and therapist competence. Certainly his variable examining high staff turnover would get at both poor adherence to treatment protocol and possibly low skill level on the part of

new clinicians. In addition, his variable analyzing the point to which the researcher was involved in treatment implementation, would capture high degrees of adherence to treatment protocol with increasing researcher involvement in the intervention. Of significance, Lipsey (2009) developed practical categories that distinguished types of treatment along important dimensions thereby defining a means of examining treatment differentiation between studies. Specifically, Lipsey distinguished between those studies that rely on outside control or coercion and those that are constructive or collaborative in nature. “Discipline” was one category that represented outside control, such as boot camps. Of those studies that Lipsey classified as constructive, he identified four categories: Restorative programs, counseling and its variants, skill building programs, and multiple coordinated services. Restorative programs included interventions that revolved around restitution or mediation. Counseling programs could include individual, family, group or some variant of that type of intervention. Skill building programs focused on instruction aimed at developing specific skills such as social skills training. Finally, multiple coordinated services tended to combine treatment modalities such as might be found in multisystemic treatment. These categories are of particular utility in that they can be employed as a means of assessing current trends in updates of the literature.

Lipsey was able to identify what works in juvenile treatment. Specifically, he demonstrated that high levels of treatment integrity are associated with greater efficacy. However, his category for treatment integrity was based almost exclusively on research-based settings, as his variable for measuring this construct was focused on the level of

involvement of the researcher in treatment implementation. To date, there are no known meta-analyses within the juvenile recidivism literature that examine treatment integrity in real-world settings.

Efficacy and Effectiveness

Within psychological research there continues to be a question as to whether treatments conducted in randomized trials can be implemented in community clinical settings with the same level of success (Hunsley & Lee, 2007). One difficulty in addressing this matter is the lack of current evidence that these treatments work in routine practice. The language often used in the literature to describe the analyses of research-based treatments as opposed to routine clinical practice is “efficacy” versus “effectiveness” studies. Efficacy studies are generally randomized controlled trials conducted by a researcher who has implemented a rigorous research design that likely includes training for all therapy providers and includes a control group (Hunsley & Lee, 2007). Effectiveness studies may have many of these same elements, but are more likely to be conducted in a clinical setting where participants usually receive either “treatment” or “treatment as usual” (Hunsley & Lee, 2007). In addition, effectiveness studies may have less stringent methods in place for assuring quality of treatment implementation. Efficacy and effectiveness studies have been labeled differently throughout the literature and may be described as “research or demonstration studies” versus “routine practice programs” (Lipsey, 2009, p. 145). Descriptors also include “university” versus “community” based studies (Mease, 2004, p. 208). Another common label for effectiveness studies is “real-world practice” (Hunsley & Lee, 2007, p. 21). Hunsley and

Lee (2007) examined 35 effectiveness studies and compared them to benchmarks (i.e., values, such as percent improvement, derived from efficacy trials that serve as the standard for a particular treatment). Their findings suggest cause for optimism that treatments conducted in research-based settings can be effective in real-world practice. In the current study distinctions between efficacy and effectiveness are important, as treatment integrity is being measured. Different labels will be used interchangeably throughout this document relating to efficacy and effectiveness studies. Specifically, efficacy studies will be referred to as either research-driven or university studies, while effectiveness studies will be labeled either real-world or community studies (see Operational Definitions in Method section).

Limitations in the Literature

Throughout the meta-analyses reviewed here, general themes arose as to the limitations in the current literature. Specifically, incomplete reporting on demographic variables was cited, such as a lack of information regarding gender and ethnicity of participants (McCart, Priester, Davies, & Azen, 2006). In addition, inadequate accounting of professional backgrounds of therapy providers was indicated (Latimer, Dowden, & Muise, 2005), which, it was noted, makes it difficult to assess the impact of therapist variables on therapeutic outcomes. More generally, problems with research methodology were revealed. In his quantitative review of the literature, Lipsey (1992) found that over half of the variance in treatment outcomes was attributable to differences in research methods rather than type of intervention employed. Particularly, a dearth of high quality research in real-world settings was observed (Landenberger & Lipsey,

2005). Wilson, Bouffard, and Mackenzie (2005) stated that a question remains as to whether the effectiveness of programs implemented by a researcher who develops a protocol will remain in effect once applied in community setting. Curtis, Ronan, and Borduin (2004) noted a confounding of efficacy and effectiveness study conditions in their meta-analysis. Pearson, Lipton, Cleland, and Yee (2002) called for detailed descriptions of quality assurance measures in future research. As noted earlier, Lipsey (2009) suggested that an examination of treatment integrity in real-world settings within the juvenile recidivism literature is indicated.

Purpose of Current Study

Juvenile criminal activity continues to be a costly problem to society. Over the past 60 years a myriad of treatment programs have been studied in an attempt to identify which ones effectively change juvenile antisocial behavior. These programs have shown varying degrees of success, one measure of which has been to examine a reduction in recidivism for juvenile offenders. Examining effect sizes from outcome studies through a meta-analysis has allowed researchers to more accurately compare the benefits of different treatment programs. However, the last meta-analysis to compare a large number of different types of treatment programs for juvenile offenders was carried out six years ago when Lipsey (2009) examined studies spanning the years 1958 to 2002. As up to 30 outcome studies examining treatment programs and their reduction on juvenile recidivism have been conducted since that time, with over 70% carried out in real-world settings, the current study proposes to analyze the most recent research on recidivism through a meta-analysis to examine the effect of current treatment programs on recidivism. In addition,

this study aims to identify which interventions are producing the largest effects in the reduction of juvenile recidivism. More specifically, this study will examine how high levels of treatment integrity in community venues affects treatment efficacy through the introduction of a novel variable able to capture treatment integrity in real-world settings.

METHOD

The primary goal of this study is to update the literature in an increasingly sophisticated manner with respect to therapeutic treatments aimed at reducing juvenile recidivism. Thus, sample, design, and treatment descriptors across studies are characterized. In addition, more specific goals of this study include determining what treatments are currently available, whether these therapies are effective at decreasing youthful reoffending, and if effective, identifying which interventions have the strongest outcomes in reducing recidivism rates. Finally, quality of treatment implementation in real-world settings is also assessed in terms of whether it increases treatment efficacy.

Specific Research Questions

1. What are the current sample, design, and treatment characteristics of the juvenile offender literature?
2. What types of programs are currently used to treat juvenile recidivism?
3. Are current treatment programs effective in reducing youthful offending?
4. Which treatment programs have the largest effect on decreasing juvenile recidivism?
5. Does quality of treatment implementation (i.e., treatment integrity) increase treatment efficacy in real-world settings?

Procedures

Data Collection

A computerized search of specific data bases was carried out to identify all studies between 2003 and 2008 that met criteria for the present meta-analysis. Databases

analyzed include PsycINFO, ProQuest Dissertations and Theses, ERIC, Health and Psychosocial Instruments (HaPI), Social Services Abstracts, Sociological Abstracts, MEDLINE, and Evidence Based Medicine Reviews Multifile. Keywords included descriptions of the population (e.g., juvenile offenders, youthful re-offenders, adolescent recidivists) and treatment (e.g., diversion programs, parent training, multisystemic therapy). Over 1,000 abstracts were acquired from these searches and examined to determine whether each study met criteria for inclusion in the present meta-analysis. From the abstracts inspected, it was established that 33 of the studies met necessary conditions. The full article or dissertation was then obtained and the reference section of each study was reviewed in an effort to find any remaining studies that may have been missed during the original database search. This process yielded 1 last study that met necessary criteria. Of the 34 studies selected for this meta-analysis, three would ultimately be eliminated during the coding process as they did not include adequate data to produce effect sizes for comparison across studies. In addition, during analysis the discovery of outliers would ultimately dictate the removal of one more article (see outliers below), leaving the final number of studies included in the present meta-analysis at thirty.

Sample

Data for this meta-analysis were derived from studies that met specific inclusion criteria. Conditions for inclusion were based upon obtaining data most likely to answer the proposed research questions in this study. Definitions are described below and are

based on those customarily found in the literature. Thus, the range of eligible studies is defined as follows:

Participants. Because the predominant focus of this study is treatment efficacy for juvenile offenders, as in previous meta-analyses analyzing youthful recidivism (Lipsey, 1992, 2009), only those studies that examined offenders between the ages of 12 and 21 years were included. If a study examined both adult and juvenile offenders, it was considered acceptable as long as results for juvenile offenders were presented separately from those of adult offenders. If a study focused exclusively on adult offenders, it was excluded from this meta-analysis. There were no further exclusion criteria with respect to participants.

Treatment. To be included in this meta-analysis studies had to examine at least one intervention aimed primarily at reducing juvenile recidivism. Consistent with definitions in prior studies, in the present meta-analysis, juvenile recidivism was defined as the re-arrest, re-conviction, or re-incarceration of a juvenile offender (VandenBos, 2007). Thus interventions could focus on a variety of treatment modalities including individual, family, group, or multisystemic therapies, as well as correctional programs, parent training, peer influences, or restitution. In addition, treatment could have taken place in either an inpatient or outpatient setting and could have employed a variety of therapeutic orientations including cognitive behavioral, behavioral, or integrative therapies. Finally, consistent with how treatment of recidivism is assessed in the literature, juveniles could have participated in treatment during incarceration, while on probation, or within the context of an aftercare program (Landenberger and Lipsey,

2005). However, in every study an outcome measure must have examined subsequent recidivism rates for the juvenile offenders after treatment.

Time range. The present meta-analysis aims to follow up on the work of (Lipsey, 2009) whose last meta-analytic review of juvenile offender treatments incorporated studies from 1958 to 2002. As the aim of the present meta-analysis is to update the literature subsequent to Lipsey's review, the current meta-analysis includes studies carried out between the years of 2003 and 2008.

Design. Studies of both experimental and quasi-experimental design were included in this meta-analysis. One of the strengths of the present study is that it examines treatments implemented in real-world settings. However, the trade off is that these types of studies do not always include random assignment of participants to treatment. Thus, although random design is considered more methodologically sound, both random and nonrandomized assignment of participants was considered acceptable in the present study in order to capture real-world treatment implementation. In addition, although no exclusion criteria was set on the type of treatment modality measured for this meta-analysis, the treatment had to be compared to either another type of treatment or treatment as usual. If there was no comparison group, the study was deemed unacceptable for current purposes.

Origin. To reduce the possibility of publication bias, both published articles and unpublished dissertations were included in the present meta-analysis. Seventy-three percent of studies that met inclusion criteria were published articles, while twenty-seven percent were unpublished dissertations. In an attempt to eliminate errors due to

translation, only those studies conducted in the English language were included in the present study. In addition, it was decided that only those studies conducted within the United States would be included in the current meta-analysis.

Coding Manual

A coding manual was designed using a model from previous meta-analyses (Lipsey & Wilson, 2001, Mease, 2004). The final version of the coding manual for this study is included in Appendix A. To some degree the literature dictated aspects of the manual, as categories were added or removed based on what was actually reported and therefore available in selected studies. It is considered accepted practice to adjust the manual in this manner (Lipsey & Wilson, 2001). The present coding manual was divided into two distinct sections. The first was the study level, which gave criteria for encoding information about independent variables (i.e., treatment types). The second was the effect size level, which gave criteria for encoding information about dependent variables (i.e., effect size values). An account of the development of the coding manual along with descriptions of the two sections follows.

Development. The first draft of the coding manual was compiled by combining variables drawn from coding manuals used in previous meta-analyses examining similar constructs of interest. This initial draft was then revised after a review of the literature indicated that some variables of interest were simply not reported often enough to justify leaving them in the manual, while variables not previously considered were being reported frequently and therefore were added to the manual. The manual would later go through two more revisions as the coding process (see coder training and testing of

manual below) dictated changes in the manual. It was important to make each category mutually exclusive and to eliminate ambiguity so that coding decisions were as unequivocal as possible. Thus some categories were revised and others eliminated after a pilot test of the manual revealed that certain categories did not meet the above criteria. Finally, an “other” option was added to some of the variables with a note to specify, in order to capture possibilities outside of the scope of the original categories.

Study level. Each study was assigned a specific identification number and a brief citation about the study was noted. Moreover, the name of the coder, the date the study was coded and the source of the study (e.g., journal article, doctoral dissertation) was recorded. In addition, four major categories of variables were encoded including (1) sample descriptors, (2) research design descriptors, (3) treatment descriptors, and (4) therapist descriptors. A description of each of the major categories and its variables follows.

Sample descriptors were aimed at gathering meaningful demographic characteristics of participants in selected studies. It became apparent after coding a few studies that not all authors distinguished between treatment and comparison group characteristics. Thus in the manual coders were instructed to note whether the demographics reported were for the entire sample or specific to treatment and comparison groups. Specific sample descriptors under investigation included (1) mean age; (2) race; (3) gender; (4) mean number of prior offenses; (5) level of crime (i.e., misdemeanor, felony); (6) whether crime was violent or nonviolent; (7) type of crime (i.e., parole violation, arson, assault, etc.); (8) mean age at first arrest; (9) mean number of offenses in

past year; (10) primary caretaker (i.e., biological mother, biological father, relative, etc); (11) household type (i.e., two parent, single parent); (12) mean number of children in household; (13) household income range (i.e., low, medium, high); and (14) income level (i.e., under \$10,000, \$10,001 to \$20,000, etc.).

Research design descriptors were intended to capture both the characteristics and quality of the author's research design. Research design descriptors included (1) total sample size (i.e., at start and end of study); (2) treatment group sample size (i.e., at start and end of study); (3) comparison group sample size (i.e., at start and end of study); (4) number of dropouts (i.e., in treatment group and comparison group); (5) how participants were assigned to treatment (i.e., random, nonrandom); (6) whether the equivalence of the treatment and comparison groups was tested (i.e., yes, no); (7) pretest differences (i.e., no significance, significance); (8) participant referral status (i.e., self referred, solicited, mandated, etc.); and study affiliation (i.e., community, university).

Treatment descriptors were pivotal to the present study as the aim is to identify which types of treatment programs are currently available, determine whether available treatment programs are effective in reducing recidivism, ascertain which treatment programs have the largest effect on reducing recidivism, and examine whether quality of treatment implementation increases efficacy in real-world settings. Early in coding it became clear that it would not be simple to categorize the types of treatments currently in practice, as each study described its identified treatment in unique terms. Therefore, instead of attempting to categorize the types of treatment programs during coding, they

were simply recorded to be later assessed and grouped by meaningful categories based on what was found in selected studies after all studies had been coded.

Upon completion of coding it was decided that treatment groups in the present study would be grouped according to five of Lipsey's (2009) categories, as it was deemed useful to make comparisons between his study and the present meta-analysis. Lipsey used seven categories in his quantitative review, but two of his categories, "surveillance" (2009, p.133) and "deterrence," (2009, p.134) were not employed in the present meta-analysis as none of the 30 included studies fit these categories. However, the following categories were used: "discipline, restorative programs, counseling and its variants, skill building programs, and multiple coordinated services" (Lipsey, 2009, p. 134-135). Discipline programs tended to be regimented in orientation and emphasized obedience to authority. Boot camps analyzed in this study fit into this category. Restorative programs included both victim/offender mediation and restitution programs consistent with those found in the present meta-analysis. A study met criteria for the counseling category if it used one of many types of therapeutic domains, such as individual, family, or group therapy, as its primary intervention, which was consistent with several studies in this quantitative review. Skill building programs tended to be behavioral in approach, such as those in this meta-analysis that used token economies or focused on educational instruction. Finally, the multiple services category included those programs that are designed to provide a number of therapeutic modalities. In the current quantitative review, studies that utilized wraparound services and Multisystemic therapies fit into this category.

There were additional treatment descriptors for both treatment and comparison groups, which included the following: (1) whether participants were treated in an inpatient or outpatient setting; (2) type of treatment setting (i.e., detention center, day treatment, school based, etc.); (3) dominant treatment domain (i.e., individual therapy, group therapy, family therapy, etc.); (4) dominant orientation of program (i.e., cognitive behavioral, behavioral, integrative, etc.); (5) treatment duration in weeks; (6) method of treatment integrity utilized (i.e., manual, training, supervision, etc.); (7) level of treatment integrity indicated (i.e., *low* with one or fewer integrity checks, *medium* with two integrity checks, and *high* with three or more integrity checks); and (8) the nature of the comparison group (i.e., wait list, no treatment, placebo).

Therapist descriptors were aimed at gathering information that might be useful in assessing how therapists' qualities influence treatment outcome. Therapist characteristics included in the coding manual for both treatment and comparison groups are as follows: (1) gender; (2) race; (3) education level (i.e., bachelor's, master's, doctoral); (4) experience (i.e., no experience, less than 1 year, 1 to 5 years, etc.); (5) licensure/certification (i.e., no license/certification, license certification); and treatment adherence by therapist (i.e., measured by self-report, measured by other report, measured by client report).

Effect size level. The dependent variable of interest in this study is recidivism. Recidivism herein was defined as the re-arrest, re-conviction, or re-incarceration of a juvenile offender. Items included in the effect size level coding manual to examine recidivism included: (1) a determination of when outcome data was collected (i.e.,

immediate to two weeks post termination, two weeks plus one day to one month post termination, etc.); (2) Whether the comparison was considered a posttest or a follow-up; (3) type of data the effect size was based on (i.e., means and standard deviations, t-values or F-values, chi-square, etc.); (4) the group raw differences favored (i.e., treatment group, neither, comparison group); (5) treatment group sample size; (6) comparison group sample size; (7) treatment and comparison group means and standard deviations; (8) treatment and comparison group proportions or frequencies; (9) t-values; (10) F-values; (11) chi-square; (12) calculated effect size; and degree of estimation in effect size (i.e., high = chi-square, frequencies, or proportions; medium = t-values or F-values; and low = means and standard deviations).

Coder Training and Testing of Manual

Two coders were employed for this meta-analysis in order to calculate interrater reliability and agreement regarding coding. After this author developed the coding manual and collected studies (see data collection below), the first 15 studies were used for both training purposes and as part of a revision process for the manual. In addition to this author a graduate assistant working on a master's degree in clinical psychology was trained as a second coder. The training procedure entailed coding one to two articles and then reconvening to examine coding discrepancies and to discuss areas of ambiguity in the manual. This process took place over a three-month period and involved about eight such meetings and the coding of 15 articles. In this manner, the coders became more consistent in their coding of studies and the manual was revised to eliminate areas of uncertainty identified in the coding process.

Interrater Reliability and Agreement

The assessment of continuous variables is known as interrater reliability, while evaluation of categorical variables is called interrater agreement. Correlation analyses are often used to calculate interrater reliability, while the kappa statistic is one of the most commonly used to assess reliability for categorical variables (Sim & Wright, 2005) and is considered more robust than simple percent agreement calculations. In order to calculate reliability of coding in the present study, thirty percent of the total studies were selected by random design for the reliability study. A doctoral level graduate student unaffiliated with the current study picked numbers corresponding to the remaining studies to be coded out of a box. The numbers chosen became the 10 studies used for reliability coding. Each coder coded these 10 studies independently. The data were then analyzed via a correlation analysis for continuous level data (e.g., mean age; mean number of prior offenses, etc.) and a Cohen's kappa for categorical variables (e.g., random/nonrandom assignment; inpatient/outpatient treatment, etc.).

Operational Definitions

Specific terms used in this study warrant further explanation. The terms and definitions were sometimes derived explicitly from previous literature and at other times were an amalgamation of concepts gathered from prior studies. A summary of definitions is as follows:

Treatment integrity. As mentioned previously, treatment integrity is defined as the implementation of an intervention as it was intended to be carried out (Perepletchikova & Kazdin, 2005). Important aspects of treatment integrity include

treatment adherence and therapist competence. Using these components of treatment integrity to measure quality assurance in studies included in the present meta-analysis, four treatment integrity measures were employed in the coding manual. These measures included an examination of whether a specific treatment used a manual, provided training to practitioners, maintained supervision of therapists, and/or engaged in adherence checks. All four measures of treatment integrity are associated with quality assurance in the literature and were grouped together here as a novel means of assessing treatment integrity. The last large meta-analysis of juvenile recidivism focused on level of involvement of the researcher in the implementation of treatment, which by its definition only measured treatment integrity in research-driven settings. The unique means of assessing quality of treatment implementation in the present study allowed for assessment of treatment integrity in both community and research-driven studies. Specific definitions of each of the four measures of treatment integrity in this study follow.

Manual. A treatment manual generally outlines both the theoretical and procedural elements underlying the orientation of a specific type of treatment (Nezu & Nezu, 2008). In the present study a manual was defined as any type of treatment protocol specifically referred to as manualized in a particular study.

Training. The training variable in the present study included any type of instruction, guidance, or educational component provided to treatment practitioners. Training could occur at any one point during the study or at multiple times throughout the treatment protocol.

Supervision. An important aspect of treatment integrity includes the level of skill and judgment exercised by the treatment practitioner (Perepletchikova & Kazdin, 2005). Thus, both training and supervision were considered important variables in capturing treatment integrity in the present study. Supervision, as coded in this meta-analysis, included any type of oversight of therapists noted in a study during treatment.

Adherence checks. Adherence refers to the accuracy with which specifics of a treatment protocol are carried out (Perepletchikova & Kazdin, 2005). There are a number of ways that adherence can be measured including self-report, client report, and by supervisor or other report (Mease, 2004). In the present study adherence checks were coded for any of the above types of measures or if a study specifically referred to adherence checks.

Level of treatment integrity. The level of treatment integrity in the present meta-analysis was defined by the number of integrity measures listed above used in a particular study. Due to limited reporting on these variables in the literature, level of treatment integrity was coded in terms of practicality in the present study in order to increase the likelihood of capturing this variable in recent studies. Although, each of the above measures of treatment integrity are unlikely to contribute equally to quality assurance, they were treated as such in the current quantitative literature review in order to most broadly capture treatment integrity in real-world settings in the current literature.

Low integrity. In the present meta-analysis a study was considered to have a low level of treatment integrity if it employed one or fewer integrity measures. However, those studies that made no mention of treatment integrity were coded as “Not reported”

rather than placed in the low integrity category, as the lack of information regarding this highly important variable in the current literature was considered indicative of the probability that no low, medium or high level of treatment integrity had been employed.

Medium integrity. A study was considered to have used a medium level of treatment integrity in the present quantitative review if two measures of treatment integrity were utilized. All studies that mentioned two of the quality assurance variables highlighted above were coded in this category. As reporting even two measures of treatment integrity is rare in the current literature, a study having two measures of treatment integrity was considered to have a medium level of treatment integrity in the present study.

High integrity. High levels of treatment integrity are associated with increased program efficacy (Perepletchikova & Kazdin, 2005). In the present meta-analysis, a study was considered to have a high level of treatment integrity if it employed three or more measures associated with quality assurance and was coded as such. Again, because reporting of quality assurance measures is rare in the current literature (Perepletchikova & Kazdin, 2005), a study reporting three or more measures of quality assurance was considered to have a high level of treatment integrity in the present study.

Treatment as usual. Treatment as usual is a customary way of describing typical services provided to participants in a study, in contrast to being placed on a wait-list or in a control group (Hunsley & Lee, 2007). Treatment as usual was defined in the same manner in the present study and is a common means of examining treatment and comparison groups in the literature.

Efficacy and effectiveness. As noted previously, efficacy refers to research-based treatment, while effectiveness denotes studies carried out in routine clinical practice. In the present study, as in prior literature, a number of terms are employed in discussing efficacy and effectiveness studies. Specifically, in the present study, efficacy studies are referred to as research-driven, or university studies, while effectiveness studies are called real-world or community studies.

Analytic Strategy

Effect Size Considerations

Independence. One complication that can arise in meta-analysis is a violation of statistical independence (Lipsey & Wilson, 2001). This breach can occur when a researcher uses more than one effect size outcome from any one study. Including more than one effect size from a study potentially introduces error by inflating the sample size, as N from a single study gets included for each effect size generated by that study. Lipsey and Wilson, (2001) state that this situation can be managed by one of two common approaches; the meta-analyst can either average the effect sizes or select one of the effect sizes based on specific criteria. Error produced by statistically dependent effect sizes was avoided in the present meta-analysis by using only one effect size from each study. In the current meta-analysis only two of the 30 studies included more than one effect size. The criterion used for selecting an effect size from each of these studies was that it be most consistent with the present analysis.

Weighting. The effect size computation takes into account the sample size from which the effect size is derived (Lipsey & Wilson, 2001). The fact that sample size

influences effect size is potentially complicated by the fact that studies generally vary (sometimes widely) in the size of their samples. Statistically, studies drawn from larger populations have less sampling error and should therefore bear more weight in a statistical computation than studies drawn from smaller sample sizes. In a meta-analysis, data should be weighted accordingly. In the present meta-analysis the inverse of the sampling error variance was employed to weight effect sizes. This is an established method for managing error from differences in sample sizes across studies (Lipsey & Wilson, 2001).

Homogeneity. When there is wide variability among studies in a meta-analysis, the mean effect size does not tend to represent the distribution well (Lipsey & Wilson, 2001). Thus, homogeneity testing is employed to determine whether the variability in effect sizes within the meta-analysis is comparable to the variability that would be found from sampling error alone. In the present study, homogeneity of effect sizes was tested using the Q statistic, which is the customary method for testing homogeneity in meta-analysis (Lipsey & Wilson, 2001). If it is determined that the effect sizes in the meta-analysis are not homogeneous, then the data should be examined to establish whether adjustments can be made.

Outliers. Outliers are those data points that lie at the extreme ends of a distribution (Lipsey & Wilson, 2001). They tend to distort the data and can lead to inaccurate conclusions. Thus, it is prudent to examine a dataset for such data points. If outliers are found, their validity should be assessed and if it is determined that they are spurious, they should be removed or adjusted so that the dataset is not distorted by their

unrepresentative values. If a decision is made to trim a data point, “Windsorizing” is a common approach to adjusting a value (Lipsey & Wilson, 2001, p. 108). In this method (Hoaglin, Iglewicz, & Tukey, 1986) cutoffs for the lower and upper quartiles (or fourth) of a distribution are computed using the following formula: $d_F = \text{upper fourth } (F_U) - \text{lower fourth } (F_L)$. The computation for the lower bound outlier is then as follows: $F_L - (1.5)*d_F$. All effect sizes lower than the obtained value are considered outliers. The computation for the upper bound outlier is: $F_U + (1.5)*d_F$. All effect sizes higher than this value are also considered outliers.

Effects model. In case a heterogeneous Q statistic is found a meta-analyst must choose between three models to help explain what factors beyond sampling error might be influencing effect size values (Lipsey & Wilson, 2001). A random effects model assumes random variability beyond sampling error, a fixed effects model suggests systematic sources of variability, and a mixed model presumes random variability beyond systematic sources. Overton (1998) indicates that fixed effects models are the most commonly used by meta-analysts and are well suited for established research domains, while random effects model should be implemented when a meta-analyst is working within newer research areas. The random effects model is more conservative and is more likely to result in a Type II error (i.e., rejecting results that were actually significant), while the fixed effects model is more likely to result in a Type I error (i.e., finding a significant result when there actually is not one). Thus, Overton (1998) suggests using the random effects model when there is a high degree of uncertainty in a newer research domain and reason for caution in drawing conclusions, while the fixed effects model is

suited to developed areas where there is limited uncertainty and less likelihood for error. The current meta-analysis follows up on a well-developed quantitative literature review where random effects models were employed (Lipsey, 2009). Thus, a fixed effects model was selected, as it is considered well suited for an established research domain such as the current one. In addition, Overton (1998) notes that when a fixed effects model is employed it should be generalized only to those studies included in the current meta-analysis.

Effect size calculations. The effect size is the primary statistic of interest in a meta-analysis. Several forms of research findings may be used to calculate effect size. Means and standard deviations, F-values, chi-squares, and frequencies or proportions were the most commonly reported in the literature and were therefore the types of outcome data employed in this meta-analysis. To meaningfully compare these different statistical measures, effect sizes were calculated for each type of statistic using a computer program that estimates effect size.

In order to most accurately calculate effect size, meta-analysts must also decide on a formula for calculation that fits the data. One method for calculating effect size, the “Standardized Mean Difference,” is indicated for use when making comparisons across treatment and comparison groups (Lipsey & Wilson, 2001, p. 48) and was therefore utilized in this study. For small sample sizes (generally less than 20) there tends to be an upward bias when using this effect size index (Lipsey & Wilson, 2001). The Hedges (1981) formula corrects for this bias and was employed in the present study as a

precautionary measure to avoid over estimation, although the sample size was well above that recommended for this correction.

A positive effect size value indicates that the treatment group is favored, while a negative effect size value suggests a more favorable outcome for the comparison group. The computer program used to calculate effect sizes in this study always generated positive values for both chi-square and F-tests even if the comparison group had a better outcome. Thus, every chi-square and F-value was examined to determine whether an adjustment needed to be made (i.e., the addition of a negative sign) to the effect size statistic produced.

Each computer generated effect size value (d^+) represented a weighted statistic adjusted for sample size. Thus, d^+ values were used in the Q calculation to test for homogeneity. Where significant Q values were obtained, indicating a heterogeneous sample, an analog to the ANOVA was performed. As mentioned previously, a fixed effects model was used in this meta-analysis, suggesting systematic variability in effect sizes. The analog to the ANOVA is a way of explaining this excess variability by partitioning it into within group (Q_W) and between group (Q_B) variability around the means. Significant values for Q_W and Q_B indicated that excess variability was accounted for by both within and between group variability.

Effect sizes. The effect size statistic was employed in the present study to examine differences in treatment outcomes across studies. To understand the meaning of effect size statistics, guidelines were used for interpreting their magnitude. Cohen (1988) offers an approach for interpreting effect size statistics that is commonly cited in the literature.

These guidelines are as follows: small effect sizes equal values of 0-0.20, medium effect sizes equal values of 0.51-0.80, and large effect sizes equal values of 0.51 and higher.

Confidence intervals. Confidence intervals allow for the assessment of precision around the mean effect size. A confidence interval gives the range within which a mean is likely to fall. Confidence intervals are reported in this study, and indicate that there is a 95% chance that the mean is within the two values reported. In addition, the mean effect size is statistically significant when the confidence interval around it does not include zero.

Effect sizes were calculated for this study using the above analytic strategy. Independence, weighting, homogeneity, outliers, and an effect size model were all considered in the computation of effect sizes for the present meta-analysis. Results from these calculations and considerations are presented in the following chapter.

RESULTS

This chapter summarizes descriptive and statistical data from the present meta-analysis. Participant, treatment, design, and therapist characteristics are described. Additionally, outcome effect size data is summarized. Specifically, results with respect to treatment effectiveness and quality of treatment implementation in real-world settings are described.

Demographic and Descriptive Data

Sample characteristics such as sample size, attrition, mean age, gender and race are reported below. Descriptive data with respect to age at first arrest, mean number of prior offenses, and crime and violence level are summarized as well. In addition, sample frequencies and interrater reliability for treatment and comparison groups are outlined in table form, which shows the strong interrater reliability on these variables (see Table 1).

Participants. Over 100,000 youth took part in the 30 research studies examined in this meta-analysis. With respect to attrition, there were close to 300 dropouts reported among studies, however, this variable was not well documented and was likely much higher. The average age of the youth was 16 years with a mean range of 14 to 20 years. Over 50% of the total sample of participants in this meta-analysis was Caucasian, about one-third was African American, and the remaining participants identified as Hispanic, Asian, American Indian, or of other ethnic origin including mixed race. Upwards of 80% of the total sample was male.

Other variables of interest included participants' age at first arrest, number of prior offenses, and type of crime committed, as well as information about their primary

Table 1

 Sample Frequencies and Interrater Reliability for Treatment and Comparison Groups

<u>Treatment</u>	<u>N</u>	<u>%</u>	<u>Reliability</u>	<u>Comparison</u>	<u>N</u>	<u>%</u>	<u>Reliability</u>
<u>Sample Size</u>	20,000		1.00	<u>Sample Size</u>	94,250		1.00
<u>Dropouts</u>	235		**	<u>Dropouts</u>	56		**
<u>Mean Age</u>	16		1.00	<u>Mean Age</u>	16		1.00
<u>Race</u>			1.00	<u>Race</u>			1.00
Caucasian	52			Caucasian	55		
Afr. American	33			Afr. American	28		
Hispanic	12			Hispanic	14		
Asian	1			Asian	1		
Other	2			Other	2		
<u>Gender</u>			1.00	<u>Gender</u>			0.95
Male	82			Male	77		
Female	18			Female	23		
<u>Mean Age at First Arrest</u>	14		1.00	<u>Mean Age at First Arrest</u>	14		1.00
<u>Mean Number of Prior Offenses</u>	5		1.00	<u>Mean Number of Prior Offenses</u>	5		1.00
<u>Crime Level</u>			**	<u>Crime Level</u>			**
Misdemeanor	57			Misdemeanor	57		
Felony	43			Felon	43		
<u>Violence Level</u>			**	<u>Violence Level</u>			**
Nonviolent	62			Nonviolent	**		
Violent	38			Violent	**		

Note. ** indicates that data were undetermined in this category due to limited reporting.

caretakers, number of children in each household, and family socioeconomic status.

Unfortunately, there was inconsistent reporting across studies on these variables. Less than 40% of studies reported the participants' age at first arrest, but of those that did, the mean age at first arrest was 14 years. Closer to 60% of the studies referenced the participants' number of prior offenses, with their mean number of previous offenses

equaling 5. Of the seven (out of thirty) studies that reported crime level, 57% of participants had committed a misdemeanor, while the remaining 43% had committed a felony. Similarly of the eight studies reporting on crime level about 60% of participants' crimes were nonviolent, while close to 40% were considered violent in nature. Although 55% of studies reported the type of crime committed by participants, none of the studies categorized the crimes in precisely the same way. In fact, among studies there were 58 different categories of crimes reported and only eight of those categories were ever repeated in subsequent studies. Too few studies reported statistics on participants' primary caretakers, number of children in each household, or familial socioeconomic status to relate meaningful results.

Treatment. Treatment categories analyzed included the nature of the treatment program, whether participants were treated in an inpatient or outpatient setting, the type of treatment site (i.e., detention center, day treatment, or school based), treatment domain (i.e., individual, group, or family therapy), treatment orientation (i.e., cognitive behavioral, behavioral, or integrative), treatment integrity utilized (i.e., manualized, supervision, adherence checks), level of treatment integrity (i.e., low = 1 or fewer checks, medium = 2 integrity checks, and high = 3 or more integrity checks), treatment duration in weeks, and the nature of the comparison group. In fact, in all 30 studies, the nature of the comparison group was considered "treatment as usual." These treatment characteristics are summarized below. In addition, treatment frequencies and interrater agreement for treatment and comparison groups are outlined in table form, which shows the variability in agreement on these variables (see Table 2).

As mentioned previously, treatment groups in the present study were categorized according to five of Lipsey's (2009) classifications, as it was deemed useful to make comparisons between his study and the present meta-analysis. The percentage of studies in this quantitative review that met criteria for Lipsey's five categories are as follows: discipline (17%), restorative programs (17%), counseling and its variants (17%), skill building programs (26%), and multiple coordinated services (23%).

The majority of treatment groups were seen in an outpatient setting (50%), with the remainder of groups treated in either an inpatient setting (33%) or a mix of inpatient and outpatient (10%). While only 7% of the studies were indistinct with respect to this variable for treatment groups, there was vague reporting in this category for the majority of the comparison groups (40%). Of those studies that did report on this variable for the comparison groups, the majority of the comparison groups were treated in an outpatient setting (37%), with the remainder treated in either an inpatient setting (13%) or in a mix of settings (10%).

Twenty-three percent of the treatment groups received care in a detention center. However, the majority of these groups were seen in other settings (67%), the most common of which was either in the community or at home. Comparison group participants were seen in detention centers (10%), school based settings (3%), and other settings (40%), which also included home- and community-based treatment. The majority of the studies did not specify treatment setting for the comparison groups (47%),

Table 2

Treatment Frequencies and Interrater Agreement for Treatment and Comparison Groups

<u>Treatment</u>	<u>N</u>	<u>%</u>	<u>Agreement</u>	<u>Comparison</u>	<u>N</u>	<u>%</u>	<u>Agreement</u>
<u>Mean Treatment Length in Weeks</u>	23		0.99*	<u>Mean Treatment Length in Weeks</u>	29		1.00*
<u>Nature</u>			**	<u>Nature</u>			**
Discipline	5	17		Discipline	5	17	
Restorative	5	17		Restorative	5	17	
Counseling	5	17		Counseling	5	17	
Skill Building	8	26		Skill Building	8	26	
Multiple	7	23		Multiple	7	23	
<u>In/Out Patient</u>			0.63	<u>In/Out Patient</u>			0.43
Inpatient	10	33		Inpatient	4	13	
Outpatient	15	50		Outpatient	11	37	
Other	3	10		Other	3	10	
Cannot tell	2	7		Cannot tell	12	40	
<u>Setting</u>			0.33	<u>Setting</u>			0.50
Detention	7	23		Detention	3	10	
Day Treatment	0	0		Day Treatment	0	0	
School Based	0	0		School Based	1	3	
Other	20	67		Other	12	40	
Cannot tell	3	10		Cannot tell	14	47	
<u>Domain</u>			0.14	<u>Domain</u>			0.52
Individual	0	0		Individual	0	0	
Group	4	13		Group	1	3	
Family	1	3		Family	0	0	
Parenting	0	0		Parenting	0	0	
Multiple	18	61		Multiple	4	13	
Other	4	13		Other	4	13	
Cannot tell	3	10		Cannot tell	21	71	
<u>Orientation</u>			0.41	<u>Orientation</u>			**
CBT	3	10		CBT	1	3	
Behavioral	2	7		Behavioral	1	3	
Integrative	0	0		Integrative	0	0	
Multiple	6	20		Multiple	0	0	
Other	3	10		Other	1	3	
Cannot tell	16	53		Cannot tell	27	91	

Note. * Indicates that this variable represented interrater reliability. ** indicates that data were undetermined in this category due to limited reporting. For all variables after Mean Treatment Length in Weeks, N = number of studies in the present meta-analysis; % = percent of studies in the present meta-analysis.

while this was true of only a smaller number of studies with respect to the treatment groups (10%).

The majority of the treatment domain fell into the “multiple” category (61%) for the treatment groups, meaning that treatment involved some combination of therapies such as individual, group, or family therapy. Very few of the studies offered only a single therapy, such as group (13%) or family therapy (3%) for the treatment groups. Similarly, 13% of the comparison groups fell into the “multiple” category for the comparison groups, with only 3% offering only a single therapy domain (i.e., group). For both treatment and comparison groups the “other” category represented 13% of the treatment domain, which involved such domain types as intensive milieu, community service, or probation. In addition, a substantial number of the studies did not specify the treatment domain for either the treatment groups (10%) or the comparison groups (71%).

Treatment orientation (i.e., cognitive behavioral, behavioral, etc.) was specified in few of the studies for either the treatment groups (53%) or the comparison groups (91%). Of the treatment groups that were reported, 10% used a cognitive behavioral orientation, 7% use behavioral, and 20% fell into the “multiple” category meaning that the treatment involved a combination of orientations. In the comparison groups 3% were cognitive behavioral in orientation, while 3% were strictly behavioral. There were also a percentage of studies that fell into the “other” category in terms of orientation for both the treatment (10%) and comparison groups (3%), which included such orientations as systems-based as well as treatments non-theoretical in orientation.

Of studies reporting on treatment length, the average, the duration of treatment for the treatment groups was 23 weeks, with a range of 2 to 52 weeks. However, 29% of studies did not report treatment length for the treatment groups, while 77% of the studies did not report this statistic for the comparison groups. Of those studies that did report treatment duration for the comparison groups, on average, the length of treatment was 29 weeks, with a range of 12 to 44 weeks.

In terms of treatment integrity, in the treatment groups, treatment integrity could not be determined in 57% of the studies, while this was true of 100% of the studies with respect to the comparison groups. Of those studies that did report on treatment integrity, a number of methods were utilized in the treatment groups to maintain integrity including use of a manual (3%), training (7%), adherence checks (3%), and a combination of these that could include supervision (27%). In terms of level of integrity, in the treatment groups 17% of the studies had low integrity, 10% had medium integrity, and 17% had high integrity. Within the treatment groups integrity level could not be determined in 57% of the cases, while this was true of 100% of the studies in the comparison groups. Integrity frequencies and interrater agreement for treatment and comparison groups are outlined in table form below, which shows the excellent agreement on these variables (see Table 3).

Research design. Research design variables were aimed at examining how subjects were assigned to treatment groups, whether the equivalence of groups was tested at pretest, and whether any significant differences were found at pretest. How participants were referred to studies and study affiliation was also examined. Research

Table 3

Integrity Frequencies and Interrater Agreement for Treatment and Comparison Groups

<u>Treatment</u>	<u>N</u>	<u>%</u>	<u>Agreement</u>	<u>Comparison</u>	<u>N</u>	<u>%</u>	<u>Agreement</u>
<u>Treatment Integrity</u>			0.80	<u>Treatment Integrity</u>			**
Manual	1	3		Manual	0	0	
Training	2	7		Training	0	0	
Supervision	0	0		Supervision	0	0	
Adherence	1	3		Adherence	0	0	
Combination	8	27		Combination	0	0	
Other	1	3		Other	0	0	
Cannot tell	17	57		Cannot tell	30	100	
<u>Level of Integrity</u>			0.79	<u>Level of Integrity</u>			**
Low	5	17		Low	0	0	
Medium	3	10		Medium	0	0	
High	5	17		High	0	0	
Cannot tell	17	57		Cannot tell	30	100	

Note. ** indicates that data were undetermined in this category due to limited reporting.

design frequencies and interrater agreement are outlined in table form below, which shows the variability in interrater agreement on these variables (Table 4).

In this meta-analysis 33 % of studies randomly assigned participants to treatment groups, while 63% used a nonrandom design and in 4% of the cases it was unclear what type of research design was employed. In addition, 73% of studies tested the equivalency of the treatment and comparison groups at pretest and of these studies, 40% found no significant differences, 17% found significant differences, and in 43% of the cases it was undetermined. Types of referrals to treatment included self referrals (3%), solicitations by researchers (13%), mandates by courts (27%), a combination of referral types (17%), an “other” category (23%), and indeterminate cases (17%). The majority of the studies in this meta-analysis were community (77%), rather than university based (23%), meaning that they were real-world clinical studies rather than being highly

Table 4

 Research Design Frequencies and Interrater Agreement

Treatment	N	%	Agreement
<u>Assignment</u>			0.53
Random	10	33	
Nonrandom	19	63	
Other	1	4	
<u>Equivalence</u>			0.50
Tested	22	73	
Not Tested	0	0	
Cannot tell	8	27	
<u>Differences</u>			0.56
No	12	40	
Yes	5	17	
Other	3	10	
Cannot tell	10	33	
<u>Referral</u>			0.23
Self	1	3	
Solicited	4	13	
Mandated	8	27	
Combined	5	17	
Other	7	23	
Cannot tell	5	17	
<u>Affiliation</u>			**
Community	23	77	
University	7	23	

Note. ** indicates that data were undetermined in this category due to limited reporting.

research driven, such as having the lead researcher heavily involved with treatment implementation.

Therapists. Therapist descriptors were reported in only 25% of studies in this meta-analysis. Moreover, even when characteristics were reported, they tended to be vague. For example, education level was mentioned in seven studies, but was often reported in indistinct terms such as “graduate level.” Years of clinical experience was cited in four studies, with a range of 1 to 15 years of experience. Two studies indicated

that at least one licensed professional was a part of the treatment team and five studies referenced adherence checks, which predominately took the form of weekly supervision where quality of protocol implementation was assessed. Finally, only one study mentioned the race of the therapists and only two studies referenced whether the clinicians were male or female.

Interrater reliability and agreement. In the present quantitative literature review, interrater reliability averaged across continuous variables was 0.98. In contrast, interrater agreement averaged across categorical variables was 0.52. The discrepancy between interrater reliability and agreement was likely due to difficulties accurately coding categorical variables because of often vague reporting in the literature. For continuous data (e.g., mean age), indistinct reporting frequently resulted in a code of “999” indicating that the data was missing and would likely have been coded the same by both coders (i.e., for continuous data, coders could not code what was not there). However, with categorical variables (e.g., inpatient/outpatient treatment), coders often made their best attempt to accurately categorize data even when descriptions in the literature were rather unclear. For example, a study might discuss an intervention conducted in a residential setting with a day treatment component, but not specify whether participants were in residential care (inpatient), day treatment (outpatient), or both, leading to the potential for more coder discrepancies with categorical variables. However, the interrater agreement averaged across categorical variables was still generally good, as Fleiss (1981) states that when interpreting kappa statistics, values greater than 0.75 represent excellent

agreement beyond chance, values between 0.40 and 0.75 represent fair to good agreement beyond chance, and values below 0.40 represent poor agreement beyond chance.

Treatment Approaches

The first analysis performed examined effects of treatment groups versus comparison groups (i.e., “treatment as usual”). In the present meta-analysis no differences were found between treatment and “treatment as usual” ($d^+ = 0.02$; CI, -0.01-0.04). This was not surprising given that “treatment as usual” represented a wide variety of conditions across the 30 studies in this quantitative review such as services provided by community mental health facilities, residential care, day treatment, court interventions, educational services, and traditional parole services. In other words, “treatment as usual” across studies was not that different than “treatment” across studies. Thus, in the present meta-analysis a more useful comparison is an examination of differences between treatment types.

The present quantitative review analyzed five types of treatment programs based on Lipsey’s (2009) classifications to examine which types of programs had the largest effects in reducing recidivism (see Figure 1). A negative outcome indicated that re-offending increased following participation in a treatment program, while a positive effect was indicative of reduced recidivism rates following treatment. As discussed previously, the discipline category in this meta-analysis was made up of boot camp

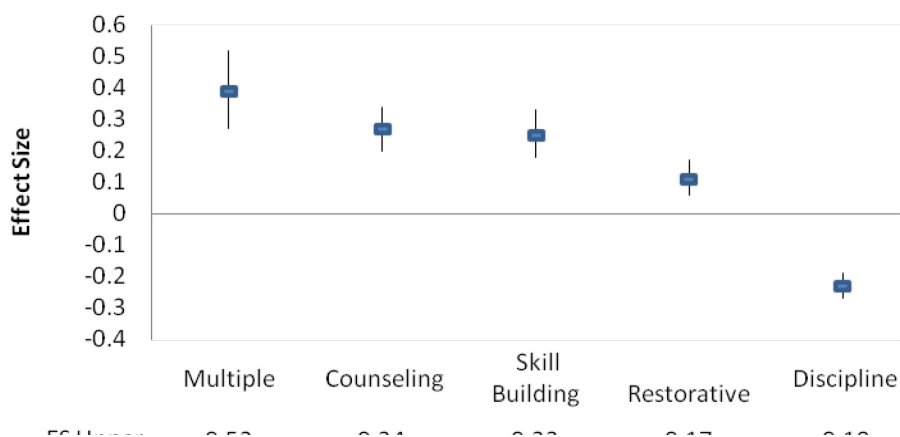


Figure 1. Effect sizes for treatment types as grouped by Lipsey's (2009) categories with 95% confidence intervals represented as lines around the effect size box.

programs ($d^+ = -0.23$; CI, -0.27 to -0.19). The restorative classification predominately contained restitution and restorative justice services ($d^+ = 0.11$; CI, 0.06 to 0.17). Counseling and its variants included interventions primarily employing individual, family, or group therapies ($d^+ = 0.27$; CI, 0.20 to 0.34). Skill building treatments tended to adopt token economies or educational curriculums ($d^+ = 0.25$; CI, 0.18 to 0.33). Finally, the multiple coordinated service classification went to wraparound service programs and multisystemic therapies ($d^+ = 0.39$; CI, 0.27 to 0.52). The results indicate that all treatment categories were effective in reducing recidivism except discipline (boot camps), with multi-coordinated services demonstrating the largest effects.

This study explored whether quality of treatment implementation (i.e., treatment integrity) increased treatment efficacy in real-world settings (see Figure 2). Specifically, a novel variable was employed that allowed for the examination of treatment integrity in either research-driven or community settings. This was a strength of this meta-analysis, as the last quantitative review of this type examined treatment integrity by the level of

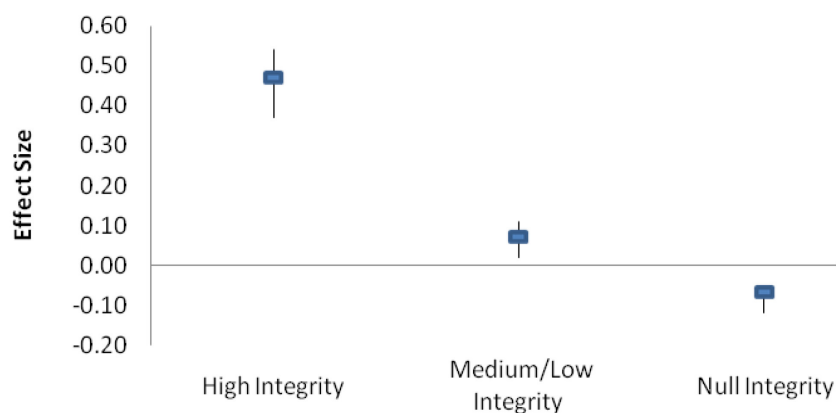


Figure 2. Effect sizes for treatment integrity types with 95% confidence intervals represented as lines around the effect size box. High integrity = 3 or more measures of integrity; medium/low = 1 to 2 measures of integrity; null integrity = no reported measures of integrity.

involvement of the researcher in the study, thus limiting examination of quality of treatment implementation to research-driven studies. In contrast, the present study analyzed treatment integrity by the number of factors associated with quality assurance that the study employed (e.g., a manual, training, supervision, adherence checks), allowing for treatment integrity to be analyzed across a variety of settings. Utilizing the novel variable for treatment integrity developed in this study, three groupings of treatment integrity across studies were explored. Again, negative effects suggested that recidivism increased after participation in treatment, while positive outcomes indicated that recidivism decreased after participation in treatment. Null integrity treatment groups were defined as those with no reported quality assurance measures ($d^+ = -0.08$; CI, -0.12 to -0.05). A medium/low level of integrity indicated that a treatment employed one to two integrity checks ($d^+ = 0.06$; CI, 0.02 to 0.11). Finally, a high level of treatment integrity indicated that three or more quality assurance measures were employed in the

study ($d^+ = 0.46$; CI, 0.37 to 0.54). Results show that those treatment programs with the highest level of treatment integrity had the strongest effects, while those treatment interventions with no reported integrity had negative outcomes.

A follow-up analysis examined treatment integrity when it was partitioned by real-world versus research-driven studies (see Figure 3). The treatment integrity variable employed in the present study assessed for quality assurance measures in both community (real-world) and university (research-driven) studies. The last meta-analysis of this type was limited to exploring treatment integrity in research-driven studies as by definition its quality assurance variable only captured those studies in which the researcher was involved with implementation. Thus, it was of interest to examine how community-based studies compared to university studies when the novel measure of treatment integrity developed in this quantitative review was employed. Specifically, in this analysis real-world studies were considered those carried out in community settings where the researcher who developed the intervention had little or no input in the implementation of the treatment ($d^+ = 0.05$; CI, 0.01 to 0.09). In contrast, research-driven studies were defined as those with which a lead researcher was involved in treatment implementation ($d^+ = 0.90$; CI, 0.78 to 1.01). Results suggest that research-driven studies continue to have the strongest effects. Nevertheless, community settings that employed measures of treatment integrity showed positive effects, albeit much less pronounced than those of university-based research.

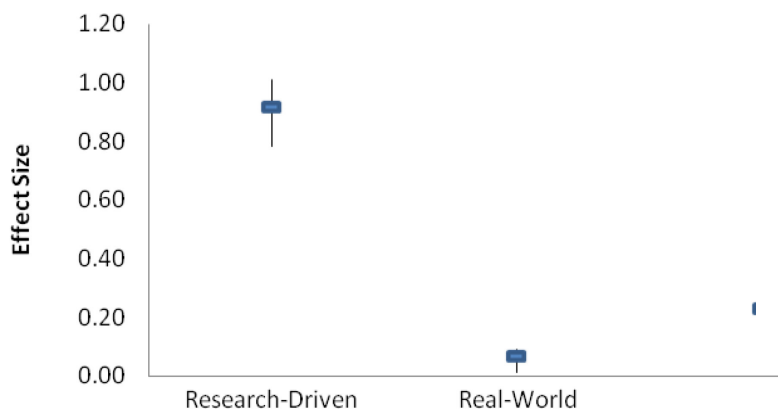


Figure 3. Effect sizes for research-driven and real-world integrity types with 95% confidence intervals represented as lines around the effect size box.

Homogeneity. To examine whether variance across treatment types and treatment integrity was greater than would be expected by chance alone, the Q statistic was analyzed (see Table 5). Specifically, Q_{Within} and Q_{Between} group differences were assessed to determine whether either or both accounted for excess variability. When comparisons were made across the five types of treatments grouped according to Lipsey's (2009) categories, both Q_{Within} and Q_{Between} were significant suggesting unexplained variability in the sample. Q_{Within} and Q_{Between} were also significant when comparison were made across high, medium/low, and null integrity; and research-driven versus real-world treatment integrity, again suggesting unexplained variability in the sample. Unexplained variability in the present meta-analysis is likely due to methodological differences among studies as Lipsey (2009) found in his recent quantitative literature review.

Outliers. As significant Q analyses were obtained in the current study, the data was assessed for outliers to examine whether distortions in the data were responsible for

Table 5

Homogeneity Analyses for Treatment Groups and Treatment Integrity Comparisons

Variable	<i>k</i>	<i>d</i> ⁺	95% CI	<i>Q</i>	<i>Q</i> _W	<i>Q</i> _B
Treatment Type					474.12**	308.24**
Discipline	5	-0.23	-0.27 to -0.19	29.10**		
Restorative	5	0.11	0.06 to 0.17	41.70**		
Skill Building	9	0.25	0.18 to 0.33	68.68**		
Counseling	5	0.27	0.20 to 0.34	200.61**		
Multiple	6	0.39	0.27 to 0.52	44.68**		
Level of Integrity					567.00**	100.75**
Null	19	-0.08	-0.12 to -0.05	346.77**		
Medium/Low	7	0.06	0.02 to 0.11	21.83**		
High	5	0.46	0.37 to 0.54	198.41**		
Partitioned Integrity					516.94**	150.81**
Real-World	6	0.05	0.01 to 0.09	8.63		
Research-Driven	7	0.90	0.78 to 1.01	161.53**		

Note. *k* = number of effect sizes per category; *d*⁺ = weight mean effect size; CI = confidence interval; *Q* = homogeneity test value; *Q*_W = homogeneity within variables; and *Q*_B = homogeneity between variables. ** *p* < 0.01.

the significant outcomes. In the present study the lower quartile effect size cutoff was found to be -0.8, while the upper quartile effect size cutoff was 1.48. There were no outliers found at the lower end of the distribution; however, there were two outliers at the upper end of the distribution. Upon examination, one outlier, a medication study, was determined to be inconsistent with the current meta-analysis and was eliminated. The second outlier was only slightly over the upper bounds of the distribution and was simply Winsorized (trimmed to the upper quartile cutoff of 1.48). Nevertheless, after removal of outliers, both *Q*_{Within} and *Q*_{Between} remained significant for all analyses as shown in Table 5.

Demographic and descriptive data from the present meta-analysis were reported here along with results of analyses comparing five types of interventions grouped according to Lipsey's (2009) categories. Three levels of treatment integrity were also analyzed as well as research-driven versus real-world treatment integrity. In the following chapter these results will be discussed and further elucidated.

DISCUSSION

In this chapter results from the present meta-analysis are discussed and interpreted. Each research question is addressed specifically, with findings reviewed and expanded upon. A summary of the findings is also included along with the strengths and limitations of this study. Finally, future directions for research are considered.

Research Question 1: Description of Recidivism Literature

With respect to the first research question in the present study, information regarding current characteristics of juvenile recidivism literature is presented here. The detailed descriptive analysis in the present study allowed for meaningful review of important variables captured in recent studies examining adolescent re-offending.

Participants

Demographic variables in the present study were consistent with the last large meta-analysis of this type (Lipsey, 2009). The participants most characteristic of studies included in this analysis were Caucasian males of approximately 16 years of age. Recent crime statistics released from the U.S. Department of Justice (Office of Juvenile Justice Delinquency and Prevention, 2007), suggest that this profile is also relatively consistent with the average juvenile offender in the United States at this time, although there is a trend toward more offending by African American youth. In the present study African American offenders made up only 30% of participants in studies reporting this demographic. This smaller proportion of African American adolescents represented in this meta-analysis contrasts with the rising level of crime committed by this population

and suggests that there is a need for more research focused explicitly on treatment protocols effective with this population of offenders.

With regard to severity of crime, the majority of offenses committed by adolescents in this study were nonviolent and tended to be misdemeanors, again consistent with the last study of this type, where only a small proportion of offenders were found to have committed violent or aggressive crimes (Lipsey, 2009). This is also consistent with crime statistic in the U.S., where less than 30% of violent crimes are committed by adolescent offenders (Office of Juvenile Justice Delinquency and Prevention, 2007). One of the most difficult variables to capture in this study was type of crime, as 58 categories of offenses were reported across 30 studies and only 8 of these categories were ever repeated in a subsequent study. Thus, there is a need for the development of meaningful categories of crime type to standardize reporting so that useful comparisons can be made across studies in the future. In addition, in the present meta-analysis, too few studies reported statistics on participants' primary caretakers, number of children in each household, or familial socioeconomic status to relate meaningful results. Thus, a need is indicated for inclusion of family characteristics in studies to understand how family dynamics affect treatment outcomes.

Treatment

The most common type of treatment reported in the present meta-analysis was comprised of multiple services, was conducted in an outpatient setting, and lasted approximately 25 weeks. A large number of participants were treated in their home or community (67%) and received treatment combining multiple domains (61%) such as

individual, group, and family therapy. However, treatment orientation was infrequently reported in the present meta-analysis (i.e., whether the treatment was cognitive behavioral, behavioral, integrative, etc.). It was unclear whether this variable was simply not being reported in studies or whether the majority of treatment protocols analyzed here were simply atheoretical in orientation. If the latter is true, it represents a significant gap between research and practice. Although distinct theories are taught within academic institutions, it may be that in actual practice an indistinct mingling of theories drive emerging treatment protocols. This would suggest a need for further exploration and analysis of the result of such practice on treatment effectiveness.

Treatment Integrity

The present study employed a novel means of examining treatment integrity to capture this variable in real-world settings. This was in contrast to the previous meta-analysis of this type, which focused exclusively on research-driven practice as a means of analyzing treatment integrity (Lipsey, 2009). In this study treatment integrity is viewed as a crucial variable in the analysis of treatment effectiveness, as replication of positive results hinges on clear articulation of treatment programs. When positive results are so important, as is the case with regard to juvenile recidivism, replicating them becomes crucial for financial and safety reasons. In the current meta-analysis, only 43% of studies reported on this variable; however, of those that did so, 27% used a combination of factors associated with treatment integrity such as a manual, training, supervision, or adherence checks. With respect to those studies that reported on treatment integrity, this variable was found to be of either low (17%), medium (10%) or high (17%) integrity.

Treatment integrity and findings regarding it are discussed below in terms of conclusions to draw and research to plan for the future given these findings.

Research Design

The most common research design found in the present study utilized nonrandom assignment of participants to treatment groups, employed testing the equivalence of groups at pretest, found no significant differences between groups, and was conducted in community settings. One of the key findings in the current meta-analysis was that 77% of the studies were real-world investigations aimed at examining how existing treatment programs were fairing, rather than exclusively research-driven analyses carried out, for example, by the researcher who developed the treatment protocol. This represented a shift from the last large meta-analysis of this type, where only 54% of studies were found to have been conducted in real-world settings (Lipsey, 2009). This suggests a move toward examination of treatment as it is being carried out in community settings. This will become particularly important later when an inability to explore treatment integrity in real-world settings is discussed as a limitation of the last meta-analysis of this type.

Therapists

As discussed earlier, therapist descriptors were so infrequently reported upon that no meaningful data were obtained. Each study was examined for therapist characteristics such as gender, race, level of education and experience, licensure, and adherence to treatment protocol. However, even when studies did mention these variables, they tended to be reported in such vague and inconsistent terms to render them meaningless. For example, studies might state that therapists were “educated,” but not specify education

level or give in-house titles for providers (e.g., case manager) that left education level and licensure unclear. This is not the first meta-analysis to note a lack of data regarding characteristics of those individuals implementing treatment (Latimer, Dowden, & Muise, 2005; Mease, 2004). As highlighted in past studies, inclusion of information about therapists would allow researchers to make meaningful statements about those characteristics that contribute to effective treatment implementation and seems long overdue in the literature.

Research Question 2: Types of Treatment for Recidivism

The second research question in the current study addressed the type of treatment programs currently available for adolescent offenders. In the present meta-analysis, types of treatment programs were recorded so that they could later be grouped into meaningful categories. As it was deemed useful to make comparisons to the last large meta-analysis of this type, it was determined that the most meaningful groupings would be consistent with Lipsey's (2009) treatment categories. Thus, within these groups, the types of treatment programs found to be presently available to adolescent offenders included those focused on discipline (e.g., boot camps), restorative justice programs (e.g., restitution), counseling (e.g., individual, group, and family therapies), skill building programs (e.g., token economies and educational programs), and multiple coordinated services (e.g., wraparound services and multisystemic therapy). In contrast, Lipsey had two additional categories in his study including surveillance (e.g., programs with rigorous monitoring such as intensive probation) and deterrence (e.g., programs using fear tactics such as "scared straight"). Lipsey found that surveillance interventions demonstrated small

effects in reducing recidivism while deterrence programs showed negative effects in terms of reducing recidivism rates. Moreover, neither type of program was found in the current literature, suggesting their decreasing usage.

Research Question 3: Overall Effectiveness

The third research question asked whether current treatment programs were effective in reducing juvenile recidivism. In the present quantitative review of the literature, no differences in recidivism were found between participants in treatment groups and those receiving “treatment as usual” in most studies. However, in the current meta-analysis one study’s “treatment” was often another study’s “treatment as usual.” Thus, the outcome, no differences between “treatment” and “treatment as usual,” was as expected. In the present study it was considered more meaningful to group treatments according to those proposed by Lipsey (2009) and then to examine differences among these treatment types. This was considered the best means of answering the question about effectiveness of treatment with respect to juvenile recidivism.

Research Question 4: Comparisons Among Treatments

Addressing which programs have the largest effect on reducing juvenile recidivism was at the core of the fourth research question in this study. The findings in the present study were consistent with those of the last large meta-analysis of this type (Lipsey, 2009). As in the last study, of those interventions found in the current meta-analysis, only discipline-based treatments (i.e., boot camps) had negative effects on reducing recidivism rates. All other categories in this study were effective in reducing juvenile recidivism. In contrast to Lipsey (2009), who found that counseling had the

strongest effects on reducing recidivism, in the present study multiple coordinated services showed the highest effect size differences and were most successful in reducing adolescent re-offending. This is consistent with recent literature, which suggests that programs like multisystemic therapy have some of the highest success rates in reducing youthful reoffending (Curtis, Ronan, & Borduin, 2004). In fact, it may be the strong adherence to treatment integrity in multisystemic therapy (Curtis, Ronan, & Borduin, 2004), that contributes to the higher effect sizes found in this category. Otherwise, the findings in the present meta-analysis were consistent with those found by Lipsey in his study. Specifically, in order of most to least effective, the following types of treatments reduced juvenile re-offending: multiple coordinated services, counseling, skill building, and restorative justice programs.

By Cohen's (1988) guidelines, multiple coordinated services, counseling, and skill building programs all had medium effects in decreasing adolescent reoffending in the present study, while restorative programs produced small effects. The findings in the current meta-analysis were relatively consistent with those of Lipsey's (2009) and prior meta-analyses (Curtis, Ronan, & Borduin, 2004; Latimer, Dowden, & Muise, 2005; Pearson, Lipton, Cleland, & Yee, 2002). Restorative programs showed somewhat weaker effects in the present meta-analysis than in previous quantitative literature reviews. However, in prior analyses restorative programs have shown slightly less effectiveness than other interventions reviewed here (Latimer, Dowden, & Muise, 2005, Lipsey, 2009). Latimer, Dowden, and Muse (2005) suggest that restorative programs may be a complimentary approach best suited to use with other rehabilitative interventions. It

should be noted that although restorative programs demonstrated less effectiveness than other interventions analyzed in this study, they still resulted in reduced juvenile reoffending. More importantly, multiple coordinated services, counseling, and skill building programs all demonstrated moderate effects in decreasing youthful recidivism.

The present meta-analysis sought to examine interventions aimed at decreasing juvenile reoffending, an area of considerable investigation in the current literature. In many ways it replicated, in more recent literature, the results obtained by Lipsey (2009) who examined studies spanning the years 1958 to 2002. In addition to supporting Lipsey's (2009) outcomes, the present study expanded on his findings through the addition of a novel variable for examining treatment integrity in real-world settings.

Research Question 5: Treatment Integrity in Real-World Settings

The fifth and final research question addressed whether the quality of treatment implementation (i.e., treatment integrity) increased treatment effectiveness in real-world treatment settings. A limitation in the prior meta-analysis examining treatment and juvenile recidivism (Lipsey, 2009) was that its measure of treatment integrity did not capture real-world settings, as by definition treatment integrity measured the level of involvement of the researcher in treatment implementation (i.e., a research-driven study). The present meta-analysis overcame this weakness by incorporating a novel way of measuring treatment integrity in either research-driven or real-world settings. It also allowed for examination of null, medium/low, and high levels of treatment integrity, as the more factors associated with treatment integrity that the study implemented (e.g., a manual, training, supervision, or adherence checks), the higher the level of treatment

integrity recorded. However, it should be noted that 57% of studies in the present meta-analysis made no mention of treatment integrity, a surprising find given the importance of this variable in the literature, but consistent with prior results in studies examining this variable (Perepletchikova & Kazdin, 2005).

In the present study, results suggest that the higher the level of treatment integrity the more effective the treatment was in reducing adolescent re-offending. Lipsey (2009) also found that high quality treatment implementation was associated with more effective treatment. Again, the difference is that the present study captured this outcome in real-world settings. In addition, when community (i.e., real-world) versus university (i.e., research-driven) studies were examined separately, community studies that implemented integrity checks continued to demonstrate higher levels of effectiveness than those studies that did not do so. However, research-driven studies with quality assurance measures showed stronger effect size differences than did community studies with integrity checks. Thus, the results suggest that there continues to be a gap between research and real-world treatment implementation. However, community studies that implemented quality assurance measures showed stronger results than those real-world studies that did not do so, indicating that implementing integrity checks in community settings is one way to begin reducing the disparity between research and practice. For example, practitioners can use the same type of checklist employed in this study to assess the strength of their quality assurance measures (e.g., are they using a manual, have they implemented training, do the practitioners receive supervision, and are there adherence checks?). Moreover, when research practices can be reasonably implemented in real-

world settings the obvious benefit is increased treatment effectiveness, which in this study means a much sought after decrease in juvenile recidivism.

Summary of Findings

While programs focused on discipline had negative effects, all other types of treatment demonstrated positive effects in reducing recidivism including from most to least effective: multiple coordinated services, counseling, skill building, and restorative justice programs. Using a measure that captured quality of treatment implementation in real-world settings, results showed that the higher the level of treatment integrity the better the outcomes in both university (research-driven) and community (real-world) settings. When university and community settings were partitioned, research-driven studies demonstrated stronger effects than did real-world practice; however, community settings that implemented integrity checks showed stronger effects in reducing juvenile recidivism than did those real-world settings that did not do so.

Strengths of the Current Study

The implementation of a novel variable assessing treatment integrity in real-world settings was fundamental to the present study and a major contribution and strength of this study. Specifically, the current quantitative literature review articulated a means of assessing treatment integrity in community settings. It followed up on a definitive study in the field (Lipsey, 2009), which indicated that the inability to capture treatment integrity in real-world settings was a limitation of the study. By partitioning treatment integrity into university (i.e., research-driven) and community (i.e., real-world) settings, the present meta-analysis was also able to identify an apparent schism between

research and practice. Nevertheless, results from this study indicated that when integrity checks are used in community settings it increases treatment efficacy. Another strength of this study is that it developed a straightforward list of treatment integrity factors that can easily be adopted by practitioners in community settings as a means of increasing treatment effectiveness by assuring quality of treatment implementation. That is, practitioners can use the same type of checklist employed in this study to assess the strength of their quality assurance measures (i.e., have they used a manual, training, supervision, or adherence checks?). In addition, the present study employed meta-analytic techniques to summarize the most recent juvenile recidivism literature, which is a more sophisticated method than the conventional literature review. Specifically, in the current study, meaningful relationships among variables across studies were able to be examined through an analytically precise method.

Limitations

A common limitation in meta-analysis is difficulty acquiring data on variables of interest across studies due to underreporting. Consistent with prior quantitative literature reviews, inadequate data on important variables was also a limitation of the present study. In addition, a persistent concern with respect to meta-analysis is that easier access to published studies may result in upward bias in the mean effect size (Lipsey & Wilson, 2001). This potential weakness is based on the premise that studies with significant findings are more likely to be published, while equally valid studies without significance remain out of circulation and therefore are less likely to be included in a meta-analysis. Rosenthal (1995) developed the *fail-safe N*, which is a technique that estimates the

number of unidentified studies (with an average effect size of 0) that would be required to change a significant result in meta-analysis (Soeken & Sripusanapan, 2003). After employing Rosenthal's method for calculating *fail-safe N*, no evidence of publication bias was found in the present study. Nevertheless, it is likely that valid unpublished studies, which would likely have contributed to the present results, were not found during the literature search for this study and therefore are not included in the present meta-analysis. Another limitation in the present study was the application of a rudimentary measure for level of treatment integrity. Due to the novel way that treatment integrity was measured in this meta-analysis, a more sophisticated means of labeling level of treatment integrity was not identified in prior literature. Thus, in this study, level of treatment integrity was defined by the number of quality assurance measures found in the current literature, rather than by how each contributed individually to treatment integrity. It is acknowledged that this was a necessary limitation in the present study, as measuring treatment integrity in real-world settings is in its beginning stages of examination in literature on treatments for juvenile recidivism.

Future Directions for Research

The primary goals of this study were to update the literature regarding recidivism and make new statements about effectiveness with particular regard to treatment integrity. In pursuit of these goals, many avenues for future research and improvement of empirical literature in this area were uncovered. Specifically, an area for future study is assessment of the degree to which different quality assurance measures (e.g., a manual, training, supervision, and adherence checks) contribute individually to treatment

integrity. In the present study it was acknowledged that integrity checks are unlikely to contribute equally to treatment integrity; however, due to the novelty of this variable in the literature every quality assurance measure was treated uniformly in the current meta-analysis. In addition, there appears to be a dearth of information in the juvenile recidivism literature regarding family characteristics. It would be helpful to understand how family dynamics affect treatment outcome. Further, identifying a means of categorizing crime types so that there is consistency across the literature would be useful. Finally, examining therapist characteristics that increase therapeutic outcomes (i.e., reduce adolescent re-offending) would benefit the field.

Concluding Comments

The present study was aimed at updating the literature with respect to identifying treatments effective in reducing juvenile recidivism. In addition, treatment integrity was analyzed and a novel variable for assessing quality assurance in real-world settings was introduced. Results suggest that quality implementation of treatment interventions results in increased effectiveness in terms of reduced juvenile reoffending. However, a schism between research and practice was found, with research-driven studies demonstrating stronger treatment integrity effects than those studies conducted in community settings. Nevertheless, this study identified a simple means of bridging this gap by articulating a straightforward set of integrity checks that can be easily implemented in real-world practice. Results of this quantitative literature review indicate that when these integrity checks are in place, it increases treatment effectiveness and ultimately results in decreased juvenile recidivism.

REFERENCES

- *Adams, J. R. K. (2003). The effectiveness of a multifaceted treatment for delinquent and ungovernable adolescents: A long-term follow-up study of recidivism. *Dissertation Abstracts International*, 64(6-B), 2903.
- Anderson, D. A. (1999). The aggregate burden of crime. *Journal of Law and Economics*, 42, 611.
- *Armstrong, T. A. (2003). The effect of Moral Reconciliation Therapy on the recidivism of youthful offenders: A randomized experiment. *Criminal Justice & Behavior* 30(6), 668-687.
- *Aultman-Bettridge, T. (2007). A gender-specific analysis of community-based juvenile justice reform: The effectiveness of family therapy programs for delinquent girls (Washington). *Dissertation Abstracts International*, 68(5-A), pp.2179.
- *Bottcher, J., & Ezell, M. E. (2005). Examining the effectiveness of boot camps: A randomized experiment with a long-term follow up. *Journal of Research in Crime and Delinquency*, 42(3), 309-332.
- *Bouffard, J. A., & Bergseth, K. A. (2008). The impact of reentry services on juvenile offenders' recidivism. *Youth Violence and Juvenile Justice*, 6(3), 295-318.
- Bureau of Justice Statistics. (1994). *Bureau of justice statistics*. Washington, D.C.: The Bureau.
- *Caldwell, M., Skeem, J., Salekin, r., & Van Rybroek, G (2006). Treatment response of adolescent offenders with psychopathy features: A 2-year follow-up. *Criminal Justice and Behavior*, 33(5), 571-596.
- *Carney, M. M., & Buttell, F. (2003). Reducing juvenile recidivism: Evaluating the wraparound services model. *Research on Social Work Practice*, 13(5), 551-568.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, M. A. (1998). The monetary value of saving a high-risk youth. *Journal of Quantitative Criminology*, 14(1), 5-33.
- *Cuellar, A. E., McReynolds, L. S., & Wasserman, G. A. (2006) A cure for crime: Can mental health treatment diversion reduce crime among youth? *Journal of Policy Analysis and Management*, 25 (1), 197-214.

- Curtis, N. M., Ronan, K. R., & Borduin, C. M. (2004). Multisystemic treatment: A meta-analysis of outcome studies. *Journal of Family Psychology, 18*(3), 411-419.
- *Diaz, F. S. L. (2006). Evaluation of the Long Beach Shortstop program: A juvenile crime diversion program. *Dissertation Abstracts International, 66*(11-B), 6256.
- Federal Bureau of Investigation. (2004). *Crime in the United States, 2003: Uniform crime reports*. United States: Federal Bureau of Investigation.
- Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement, 5*(1), 105-112.
- *Forgays, D. K. (2008). Three years of teen court offender outcomes. *Adolescence, 43*(171), 473-484.
- *Eddy, M. M., Whaley, R. B., Chamberlain, P. (2004). The prevention of violent behavior by chronic and serious male juvenile offenders: A 2-year follow-up of a randomized clinical trial. *Journal of Emotional and Behavioral Disorders, 12*(1), 2-8.
- Giblin, M. J. (2002). Using police officers to enhance the supervision of juvenile probationers: An evaluation of the Anchorage CAN Program. *Crime and Delinquency, 48*(1), 116-137.
- Goldstein, A. P., & Glick, B. (1987). *Aggression replacement training: A comprehensive intervention for aggressive youth*. Champaign, IL: Research.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.
- Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association, 81*(396), 991-999.
- Hollin, C. R. (1999). Treatment programs for offenders: Meta-analysis, "What works," and beyond. *International Journal of Law and Psychiatry, 22*, 361-372.
- Hunsley, J., & Lee, C. M. (2007). Research-informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice, 38*(1), 21-33.

- *Jones, C. D., Lowe, L. L., & Risler, E. A. (2004). The effectiveness of wilderness adventure therapy programs for young people involved in the juvenile justice system. *Residential Treatment for Children and Youth*, 22(2), 53-62.
- *Kelley, T. M., Kennedy, D. B., & Homant, R. J. (2003). Evaluation of an individualized treatment program for adolescent shoplifters. *Adolescence*, 38(152), 725-733.
- Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology*, 1, 451-476.
- Latimer, J., Dowden, C., & Muise, D. (2005). The effectiveness of restorative justice practices: A meta-analysis. *The Prison Journal*, 85(2), 127-144.
- Lipsey, M. W. (1992). The effect of treatment on juvenile delinquents: Results from meta-analysis. In F. Losel, D. Bender, & T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 131-143). New York: Walter de Gruyter & Co.
- Lipsey, M. W. (2006) The effects of community-based group treatment for delinquency: A meta-analytic search for cross-study generalizations. In K. A. Dodge & T. J. Dishion (Eds.), *Deviant peer influences in programs for youth: Problems and solutions* (pp. 162-184). New York, NY: Guilford Press.
- Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and Offenders*, 4, 124-147.
- Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE Publications, Inc.
- Loxley, W. (2005). Interventions for illicit drug users within the criminal justice system: A review of some programmes in Australia, the United Kingdom and the United States. In T. Stockwell, P. J. Gruenewald, J. W. Toumbourou, & W. Loxley (Eds.) *Preventing harmful substance Use: The evidence base for policy and practice* (pp. 279-289). New York, NY: John Wiley & Sons Ltd.
- McCart, M. R., Priester, P. E., Davies, W. H., & Azen, R. (2006). Differential effectiveness of behavioral parent-training and cognitive-behavioral therapy for antisocial youth: A meta-analysis. *Journal of Abnormal Psychology*, 34(4), 527-543.
- McGuire, J., & Priestley, P. (1995). Reviewing 'What works': Past, present and future. In J. McGuire (Ed.) *What Works: Reducing Reoffending: Guidelines from research and practice* (pp. 3-34). New York: John Wiley & Sons.

- *McMackin, R. A., Tansi, R., & LaFratta, A. (2004). Special Section: Studies in the rehabilitation of juvenile offenders: Recidivism among juvenile offenders over periods ranging from one to twenty years following residential treatment. *Journal of Offender Rehabilitation*, 38(3), 1-15.
- Mease, A. L. (2004). Meta-analysis of family therapy research. *Dissertation Abstracts International*, 66(1-B), 566.
- *Mills, K.R. (2008). Educational advocacy for court involved youth with disabilities: An empirical analysis of program efficacy towards child find requirements and decreased recidivism. *Dissertation Abstracts International*, 69(2-A), 565.
- Mitchell, O., Wilson, D. B., & MacKenzie, D. L. (2007). Does incarceration-based drug treatment reduce recidivism? A meta-analytic synthesis of the research. *Journal of Experimental Criminology*, 3(4), 353-375.
- *Morrall, A. R., McCaffrey, D. F., & Ridgeway, G. (2004). Effectiveness of community-based treatment for substance-abusing adolescents: 12-month outcomes of youths entering Phoenix Academy or alternative probation dispositions. *Psychology of Addictive Behavior*, 18(3), 257-268.
- *Mulford, C. F. (2004). Restorative justice and the development of empathy, remorse and moral disengagement in adolescent offenders. *Dissertation Abstracts International*, 64(10-B), pp.5256.
- Nezu, A. M., & Nezu, C. M. (2007). Ensuring treatment integrity. In A. M. Nezu & C. M. Nezu (Eds.) *Evidence-based outcome research: A practical guide to conducting randomized control trials for psychosocial interventions* (pp. 263-281). New York, NY: Oxford University Press.
- Office of Juvenile Justice and Delinquency Prevention. (2006). Juvenile Arrests 2004. *Bulletin*, 1-12.
- Office of Juvenile Justice Delinquency and Prevention. (2007). *Statistical Briefing Book*. Retrieved November 14, 2009, from www.ojjdp.gov
- Overton, R. C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3(3), 354-379.
- *Patrick, S., Marsh, R., Bundy, W., Mimura, S., & Perkins, T. (2004). Control group study of juvenile diversion programs: An experiment in juvenile diversion—the comparison of three methods and a control group. *Social Science Journal*, 41(1), 129-135.

- *Pealer, J. A. (2005). A community of Peers: Promoting behavior change. The effectiveness of a therapeutic community for juvenile male offenders in reducing recidivism. *Dissertation Abstracts International*, 65(9-A), 3580.
- Pearson, F. S., Lipton, D. S., Cleland, C. M., & Yee, D. S. (2002). The effects of behavioral/cognitive-behavioral programs on recidivism. *Crime & Delinquency*, 48(3), 476-496.
- Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice*, 12, 365-383.
- Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, 75, 829-841.
- Perkins-Dock, R. E. (2001). Family interventions with incarcerated youth: A review of the literature. *International Journal of Offender Therapy and Comparative Criminology*, 45(5), 606-625.
- Petrosino, A., Turpin-Petrosino, C., & Finckenauer, J. (2000). Well-meaning programs can have harmful effects! Lessons from experiments of programs such as Scared Straight. *Crime and Delinquency*, 46(3), 354-379.
- *Pullman, M. D., Kerbs, J., Koroloff, N., Veach-White, E., Gaylor, R., & Sieler, D. (2006). Juvenile offenders with mental health needs: Reducing recidivism using wraparound. *Crime and Delinquency*, 52(3), 375-397.
- Robbins, M. S., & Szapocznik, J. (2000). *Brief strategic family therapy* (Bulletin). Washington, DC: Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.
- *Rodriguez, N. (2007). Restorative justice at work: Examining the impact of restorative justice resolutions on juvenile recidivism. *Crime & Delinquency*, 53(3), 355-379.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, 118(2), 183-192.
- *Rowland, M. K. (2008). Family-focused reintegration for youth on parole: Evaluation of a state-wide program. *Dissertation Abstracts International*, 68(7-A), 2889.

- *Schaeffer, C. M., & Borduin, C. M. (2005). Long-term follow-up to a randomized clinical trial of multisystemic therapy with serious and violent juvenile offenders. *Journal of counseling and clinical psychology, 73*(3), 445-453.
- Sexton, T. L., & Alexander, J. F. (2000). *Functional Family Therapy (Bulletin)*. Washington, DC: Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.
- *Sloan, J. J., Smykla, J. O., & Rush, J. P. (2004). Do juvenile drug courts reduce recidivism?: Outcomes of drug court and an adolescent substance abuse program. *American Journal of Criminal Justice, 29*(1), 95-115.
- Soeken, K. L., & Sripusanapan, A. (2003). Assessing publication bias in meta-analysis. *Nursing Research, 52*(1), 57-60.
- *Steiner, B., & Giacomazzi, A. L. (2007). Juvenile waiver, boot camp, and recidivism in a northwestern state. *The Prison Journal, 87*(2), 227-240.
- *Timmons-Mitchell, J., Bender, M. B., Kishna, M. A., Mitchell, C. C. (2006). An independent effectiveness trial of multisystemic therapy with juvenile justice youth. *Journal of Clinical Child and Adolescent Psychology, 35*(2), 227-236.
- VandenBos, G. R. (Ed.). (2007). *APA dictionary of psychology*. Washington, D.C.: APA.
- Van Wormer, K. (2003). Review of treatment of drug offenders: Policies and issues. *Journal of Social Work Practice in the Addictions, 3*(4), 101-102.
- *Vappie-Aydin, R. C. (2008). Reducing the rate of recidivism for first-time juvenile offenders with the Parent Monitoring Program, PMP: A family counseling intervention program. *Dissertation Abstracts International, 69*(5-A), 1684.
- *Waite, D., Keller, A., McGravey E. L., Weickowski, E., Pinkerton, R., & Brown, G. L. (2005). Juvenile offender rearrest rates for sexual, violent nonsexual, and property crimes: A 10-year follow-up. *Sexual Abuse: Journal of Research and Treatment, 17*(3), 313-331.
- *Wells, J. B., Minor, K. I., Angel, E., & Stearman, K. D. (2006). A quasi-experimental evaluation of a shock incarceration and aftercare program for juvenile offenders. *Youth Violence and Juvenile Justice, 4*(3), 219-233.
- Wilson, D. B., Bouffard, L. A., & Mackenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders. *Criminal Justice and Behavior, 32* (2), 172-204.

*Winokur, K.P., Smith, A., Bontrager, S. R., & Blankship, J. L. (2008). Juvenile recidivism and length of stay. *Journal of Criminal Justice*, 36(2), 126-137.

APPENDIX A

Study Number* _____

**Coding Manual
For
Reducing Juvenile Recidivism:
A Meta-Analysis of Treatment Outcomes**

For missing data enter 999

StudyID *Assign each study an identification number. If a report presents two independent studies with two independent outcomes, add a decimal to the study ID number to distinguish each study within the report (e.g., 1.1 and 1.2) and code each independent study separately.

Coder Record name of coder _____

DateCode Record date study was coded _____

PubType Select the code that best describes the type of publication
 1 = Journal Article
 2 = Book chapter
 3 = Book
 4 = Doctoral dissertation
 5 = Other (Specify)

PubYear Record publication year _____

CITATION: Write an abbreviated citation in APA format

Sample Descriptors

MeanAge Record the mean age of participants reported in each study

Treatment Group

M = _____

SD = _____

Comparison Group

M = _____

SD = _____

RaceP Record the racial makeup of the sample, providing both the number (N) and percent (%) of participants from each racial background in the study. *If only data from the Total Sample is reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group

1 = Caucasian: N = _____, % = _____

2 = African American: N = _____, % = _____

3 = Hispanic: N = _____, % = _____

4 = Asian: N = _____, % = _____

5 = Other (Specify): N = _____, % = _____

8 = Not reported

9 = Cannot tell

Comparison Group

1 = Caucasian: N = _____, % = _____

2 = African American: N = _____, % = _____

3 = Hispanic: N = _____, % = _____

4 = Asian: N = _____, % = _____

5 = Other (Specify): N = _____, % = _____

8 = Not reported

9 = Cannot tell

GenderP and Record the gender makeup of the sample, providing both the number (N) percent (%) of male and female participants in each study. *If only data from the Total Sample is reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group

1 = Male N = _____, % = _____

2 = Female N = _____, % = _____

8 = Not reported

9 = Cannot tell

Comparison Group

1 = Male

N = _____, % = _____

2 = Female

N = _____, % = _____

8 = Not reported

9 = Cannot tell

PriorOff Record mean number of prior offenses. *If only data from the Total Sample are reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group

M = _____

SD = _____

Comparison Group

M = _____

SD = _____

CrimeLev Record level of crime. *If only data from the Total Sample are reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group

1 = Misdemeanor

N = _____, % = _____

2 = Felony

N = _____, % = _____

3 = Other

N = _____, % = _____

8 = Not reported

9 = Cannot tell

Comparison Group

1 = Misdemeanor

N = _____, % = _____

2 = Felony

N = _____, % = _____

3 = Other

N = _____, % = _____

8 = Not reported

9 = Cannot tell

CrimeVnv Record whether crime was violent or nonviolent. *If only data from the Total Sample are reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group

1 = Violent

N = _____, % = _____

2 = Nonviolent

N = _____, % = _____

8 = Not reported

9 = Cannot tell

Comparison Group

1 = Violent N = _____, % = _____
 2 = Nonviolent N = _____, % = _____
 8 = Not reported
 9 = Cannot tell

CrimeTyp Record type of crime. *If only data from the Total Sample are reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group

1 = Parole violation N = _____, % = _____
 2 = Arson N = _____, % = _____
 3 = Assault N = _____, % = _____
 4 = Robbery N = _____, % = _____
 5 = Rape N = _____, % = _____
 6 = Murder N = _____, % = _____
 7 = Other (Specify) N = _____, % = _____
 8 = Not reported
 9 = Cannot tell

Comparison Group

1 = Parole violation N = _____, % = _____
 2 = Arson N = _____, % = _____
 3 = Assault N = _____, % = _____
 4 = Robbery N = _____, % = _____
 5 = Rape N = _____, % = _____
 6 = Murder N = _____, % = _____
 7 = Other (Specify) N = _____, % = _____
 8 = Not reported
 9 = Cannot tell

FirstAge Record mean age at first arrest. *If only data from the Total Sample are reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group

M = _____
 SD = _____

Comparison Group

M = _____
 SD = _____

PastOff Record number of offenses in past year. *If only data from the Total Sample are reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group
 M = _____
 SD = _____

Comparison Group
 M = _____
 SD = _____

PrimCare

Select the code that best represents primary caretaker(s). *If only data from the Total Sample are reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group

1 = Biological mother	N = _____, % = _____
2 = Biological father	N = _____, % = _____
3 = Relative (Specify)	N = _____, % = _____
4 = Foster parent	N = _____, % = _____
5 = Adoptive parent	N = _____, % = _____
6 = Other (Specify)	N = _____, % = _____
8 = Not reported	
9 = Cannot tell	

Comparison Group

1 = Biological mother	N = _____, % = _____
2 = Biological father	N = _____, % = _____
3 = Relative (Specify)	N = _____, % = _____
4 = Foster parent	N = _____, % = _____
5 = Adoptive parent	N = _____, % = _____
6 = Other (Specify)	N = _____, % = _____
8 = Not reported	
9 = Cannot tell	

HousType

Select the code that best describes the household type. *If only data from the Total Sample are reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group

1 = Two parent	N = _____, % = _____
2 = Single parent	N = _____, % = _____
3 = Other (Specify)	N = _____, % = _____
8 = Not reported	
9 = Cannot tell	

Comparison Group

1 = Two parent	N = _____, % = _____
2 = Single parent	N = _____, % = _____
3 = Other (Specify)	N = _____, % = _____
8 = Not reported	

9 = Cannot tell

HousNum Record mean number of children in household. *If only data from the Total Sample are reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group

M = _____

SD = _____

Comparison Group

M = _____

SD = _____

IncomRan Record income range. *If only data from the Total Sample are reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group

1 = Low

N = _____, % = _____

2 = Medium

N = _____, % = _____

3 = High

N = _____, % = _____

8 = Not reported

9 = Cannot tell

Comparison Group

1 = Low

N = _____, % = _____

2 = Medium

N = _____, % = _____

3 = High

N = _____, % = _____

8 = Not reported

9 = Cannot tell

IncomLev Record income level. *If only data from the Total Sample are reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

Treatment Group

1 = Under \$10,000

N = _____, % = _____

2 = \$10,001 to 20,000

N = _____, % = _____

3 = \$ 20,001 to 30,000

N = _____, % = _____

4 = \$30,001 to 40,000

N = _____, % = _____

5 = \$40,001 to \$50,000

N = _____, % = _____

6 = Over \$50,001

N = _____, % = _____

8 = Not reported

9 = Cannot tell

Comparison Group

1 = Under \$10,000

N = _____, % = _____

2 = \$10,001 to 20,000

N = _____, % = _____

3 = \$ 20,001 to 30,000	N = _____, % = _____
4 = \$30,001 to 40,000	N = _____, % = _____
5 = \$40,001 to \$50,000	N = _____, % = _____
6 = Over \$50,001	N = _____, % = _____
8 = Not reported	
9 = Cannot tell	

Research Design Descriptors

TotalSizS Record *total* sample size at **start** of study N = _____

TreatSizS Record *treatment group* sample size at **start** of study N = _____

ContrSizS Record *comparison group* sample size at **start** of study N = _____

TotalSizE Record *total* sample size at end of study N = _____

TreatSizE Record *treatment group* sample size at end of study N = _____

ContrSizE Record *comparison group* sample size at end of study N = _____

DropNum Record number of dropouts by end of study. *If only data from the Total Sample are reported record under Treatment Group and label it as such (i.e., cross out Treatment Group and write in "Total Sample").*

1 = Treatment	N = _____, % = _____
2 = Comparison	N = _____, % = _____
8 = Not reported	
9 = Cannot tell	

Assign Record how subjects were assigned to treatment

1 = Random
 2 = Nonrandom
 3 = Other (Specify)
 8 = Not reported
 9 = Cannot tell

Equiv Was the equivalence of the groups tested as pretest?

1 = yes
 2 = No

8 = Not reported

9 = Cannot tell

PreDif Pretest differences between treatment and comparison groups, if tested.
Insert comments regarding how important these differences were from the authors' perspective (e.g., what is the hypothesized impact of the differences—or lack thereof—on the results)

1 = No significant differences. Comment: _____

2 = Significant differences. Comments: _____

3 = Other (Specify)

8 = Not reported

9 = Cannot tell

Referral Participant referral status.

1 = Self referred

2 = Solicited

3 = Mandated

4 = Combination

5 = Other (Specify)

8 = Not reported

9 = Cannot tell

StudyAff Study affiliation. Consider study community or clinically based unless conducted in a university or lab setting.

1 = Community (effectiveness study, realistic setting)

2 = University (efficacy study, high degrees of control/lab based)

3 = Other

8 = Not reported

9 = Cannot tell

Nature of the Treatment Descriptors

InOutPat Record whether participants were treated in an inpatient or outpatient setting.

Treatment Group

1 = Inpatient

2 = Outpatient

3 = Other (Specify)

8 = Not reported

9 = Cannot tell

InOutPatC	<p>Comparison Group</p> <p>1 = Inpatient</p> <p>2 = Outpatient</p> <p>3 = Other (Specify)</p> <p>8 = Not reported</p> <p>9 = Cannot tell</p>
SetType	<p>Record type of treatment setting.</p> <p>Treatment Group</p> <p>1 = Detention Center</p> <p>2 = Day treatment</p> <p>3 = School based</p> <p>4 = Other (Specify)</p> <p>8 = Not reported</p> <p>9 = Cannot tell</p>
SetTypeC	<p>Comparison Group</p> <p>1 = Detention Center</p> <p>2 = Day treatment</p> <p>3 = School based</p> <p>4 = Other (Specify)</p> <p>8 = Not reported</p> <p>9 = Cannot tell</p>
TreatTyp	<p>Record name of treatment program. <i>What do the authors call the type of treatment (e.g., multisystemic, drug treatment, restitution)?</i></p> <hr/>
TreatTypC	<p>Record name of comparison program. <i>What do the authors call the type of comparison(e.g., multisystemic, drug treatment, restitution)?</i></p> <hr/>
TreatDom	<p>Record dominant treatment domain of the program. <i>If there is more than one, specify under "multiple" and specify.</i></p> <p>Treatment Group</p> <p>1 = Individual therapy</p> <p>2 = group therapy</p> <p>3 = family therapy</p>

	<ul style="list-style-type: none"> 4 = Parent training 5 = Multiple (Specify) 6 = Other (Specify) 8 = Not reported 9 = Cannot tell 	
TreatDomC	<p>Comparison Group</p> <ul style="list-style-type: none"> 1 = Individual therapy 2 = group therapy 3 = family therapy 4 = Parent training 5 = Multiple 6 = Other (Specify) 8 = Not reported 9 = Cannot tell 	
Orient	<p>Record dominant orientation of the program. <i>If there is more than one, specify under "Multiple" and specify.</i></p> <p>Treatment Group</p> <ul style="list-style-type: none"> 1 = Cognitive behavioral therapy 2 = Behavioral therapy 3 = Integrative 4 = Multiple (Specify) 5 = Other (Specify) 8 = Not reported 9 = Cannot tell 	
OrientC	<p>Comparison Group</p> <ul style="list-style-type: none"> 1 = Cognitive behavioral therapy 2 = Behavioral therapy 3 = Integrative 4 = Multiple 5 = Other (Specify) 8 = Not reported 9 = Cannot tell 	
TreatDur	<p>Record treatment duration in weeks</p> <ul style="list-style-type: none"> 1 = Treatment 2 = Comparison 8 = Not reported 9 = Cannot tell 	<ul style="list-style-type: none"> N = _____ N = _____

TreatInt	<p>Record method of treatment integrity utilized</p> <p>Treatment Group</p> <p>1 = Manual</p> <p>2 = Training</p> <p>3 = Supervision of therapy</p> <p>4 = Adherence checks</p> <p>5 = Combination (Specify)</p> <p>6 = Other (Specify)</p> <p>8 = Not reported</p> <p>9 = Cannot tell</p> <p>Comparison Group</p> <p>1 = Manual</p> <p>2 = Training</p> <p>3 = Supervision of therapy</p> <p>4 = Adherence checks</p> <p>5 = Combination (Specify)</p> <p>6 = Other (Specify)</p> <p>8 = Not reported</p> <p>9 = Cannot tell</p>
LevelInt	<p>Record level of treatment integrity indicated (<i>use the data above to determine level of integrity below</i>).</p> <p>Treatment Group</p> <p>1 = low (<i>1 or fewer integrity checks</i>)</p> <p>2 = Average (<i>2 integrity checks</i>)</p> <p>3 = High (<i>3+ integrity checks</i>)</p> <p>4 = Other (Specify)</p> <p>8 = Not reported</p> <p>9 = Cannot tell</p>
LevelIntC	<p>Comparison Group</p> <p>1 = low (<i>1 or fewer integrity checks</i>)</p> <p>2 = Average (<i>2 integrity checks</i>)</p> <p>3 = High (<i>3+ integrity checks</i>)</p> <p>4 = Other (Specify)</p> <p>8 = Not reported</p> <p>9 = Cannot tell</p>
NatComp	<p>Nature of Comparison group</p> <p>1 = Wait list</p> <p>2 = No treatment</p>

- 3 = Placebo.
 4 = Other (Specify)
 9 = cannot tell

Therapist Characteristics

GenderT	Record therapist gender	
	Treatment Group	
	1 = Male	N = _____, % = _____
	2 = Female	N = _____, % = _____
	9 = Cannot tell	
GenderTC	Comparison Group	
	1 = Male	N = _____, % = _____
	2 = Female	N = _____, % = _____
	9 = Cannot tell	
RaceT	Record therapist race	
	Treatment Group	
	1 = Caucasian:	N = _____, % = _____
	2 = African American:	N = _____, % = _____
	3 = Hispanic	N = _____, % = _____
	4 = Asian	N = _____, % = _____
	5 = Other (Specify)	N = _____, % = _____
	9 = Cannot tell	
RaceTC	Comparison Group	
	1 = Caucasian:	N = _____, % = _____
	2 = African American:	N = _____, % = _____
	3 = Hispanic	N = _____, % = _____
	4 = Asian	N = _____, % = _____
	5 = Other (Specify)	N = _____, % = _____
	9 = Cannot tell	
EdLevT	Record therapist education level	
	Treatment Group	
	1 = Bachelor's level student	N = _____, % = _____
	2 = Master's level student	N = _____, % = _____
	3 = Doctoral level student	N = _____, % = _____
	4 = Bachelor's level professional	N = _____, % = _____
	5 = Master's level professional	N = _____, % = _____
	6 = Doctoral level professional	N = _____, % = _____

	7 = Multidisciplinary team	N = _____, % = _____
	8 = Other	N = _____, % = _____
	9 = Cannot tell	
EdLevTC	Comparison Group	
	1 = Bachelor's level student	N = _____, % = _____
	2 = Master's level student	N = _____, % = _____
	3 = Doctoral level student	N = _____, % = _____
	4 = Bachelor's level professional	N = _____, % = _____
	5 = Master's level professional	N = _____, % = _____
	6 = Doctoral level professional	N = _____, % = _____
	7 = Multidisciplinary team	N = _____, % = _____
	8 = Other	N = _____, % = _____
	9 = Cannot tell	
ExperT	Therapist experience	
	Treatment Group	
	1 = No experience	N = _____, % = _____
	2 = Less than 1 year	N = _____, % = _____
	3 = 1 to 5 years	N = _____, % = _____
	4 = 5 to 10 years	N = _____, % = _____
	5 = 10 to 15 years	N = _____, % = _____
	6 = 15 to 20 years	N = _____, % = _____
	7 = Over 20 years	N = _____, % = _____
	8 = Other (Specify)	N = _____, % = _____
	9 = Cannot tell	
ExperTC	Comparison Group	
	1 = No experience	N = _____, % = _____
	2 = Less than 1 year	N = _____, % = _____
	3 = 1 to 5 years	N = _____, % = _____
	4 = 5 to 10 years	N = _____, % = _____
	5 = 10 to 15 years	N = _____, % = _____
	6 = 15 to 20 years	N = _____, % = _____
	7 = Over 20 years	N = _____, % = _____
	8 = Other (Specify)	N = _____, % = _____
	9 = Cannot tell	
LicenceT	Therapist Licensure/Certification	
	Treatment Group	
	1 = No license/certification	N = _____, % = _____
	2 = License/certification (Specify)	N = _____, % = _____
	9 = Cannot tell	

ExperTC	<p>Comparison Group</p> <p>1 = No license/certification N = _____, % = _____</p> <p>2 = License/certification (Specify) N = _____, % = _____</p> <p>9 = Cannot tell</p>
AdhereT	<p>Treatment adherence by therapist</p> <p>Treatment Group</p> <p>1 = Measured by self-report (Specify)</p> <p>2 = Measured by other report (Specify)</p> <p>3 = Measured by client (Specify)</p> <p>4 = Other (Specify)</p> <p>9 = Cannot tell</p>
AdhereTC	<p>Comparison Group</p> <p>1 = Measured by self-report (Specify)</p> <p>2 = Measured by other report (Specify)</p> <p>3 = Measured by client (Specify)</p> <p>4 = Other (Specify)</p> <p>9 = Cannot tell</p>

Effect Size Level Coding Manual

StudyID	<p>For each effect size, code all of the following items.</p> <p>Identification number of the study from which the effect size is coded</p> <p>Study ID Number: _____</p>
---------	---

Dependent Measure Descriptor

ESComp	<p>Effect size comparison. Determine when the outcome data was collected.</p> <p>If considered a posttest comparison, place a “P” next to the outcome.</p> <p>If considered a follow-up comparison, place an “F” next to the outcome.</p> <p>(Consider all outcome data a posttest comparison unless explicitly called a follow-up comparison in the study)</p>												
	<table> <tr> <td></td> <td style="text-align: right;">“P”</td> <td style="text-align: right;">“F”</td> </tr> <tr> <td>1 = Immediate to two weeks post termination</td> <td style="text-align: right;">_____</td> <td style="text-align: right;">_____</td> </tr> <tr> <td>2 = Two weeks (+1 day) to one month post termination</td> <td style="text-align: right;">_____</td> <td style="text-align: right;">_____</td> </tr> <tr> <td>3 = One month (+1 day) to three months post termination</td> <td style="text-align: right;">_____</td> <td style="text-align: right;">_____</td> </tr> </table>		“P”	“F”	1 = Immediate to two weeks post termination	_____	_____	2 = Two weeks (+1 day) to one month post termination	_____	_____	3 = One month (+1 day) to three months post termination	_____	_____
	“P”	“F”											
1 = Immediate to two weeks post termination	_____	_____											
2 = Two weeks (+1 day) to one month post termination	_____	_____											
3 = One month (+1 day) to three months post termination	_____	_____											

- 4 = Three months (+1 day) to six months post termination ___ ___
 5 = Six months to one year post termination ___ ___
 6 = One year (+1 day) to 18 months post termination ___ ___
 7 = 18 months (+1 day) to two years post termination ___ ___
 8 = Two years (+1 day) post termination and beyond ___ ___
 9 = Cannot tell

Effect Size Data

ESType Record the type of data effect size is based on

- 1 = Means and standard deviations
 2 = t -value or F -value
 3 = Chi-square ($df=1$)
 4 = Frequencies or proportions, dichotomous
 5 = Frequencies or proportions, polychotomous
 9 = Other (specify)

PageNum Record page number where the data for this effect size was found.

Page number _____

Success Raw difference favors (i.e., shows more success for) which group?

- 1 = Treatment group
 2 = Neither (exactly equal)
 3 = Comparison group
 9 = Cannot tell or statistically insignificant report only

Sample Size

TreatSiz Treatment group sample size _____

ContrSiz Comparison group sample size _____

Means and Standard Deviations

TXMean Treatment group mean _____

CGMean Comparison group mean _____

TXSD Treatment group standard deviation _____

CGSD Comparison group standard deviation _____

Proportions or Frequencies

TXSucces *n* of treatment group with a successful outcome _____

CGSucces *n* of comparison group with successful outcome _____

TXProp Proportion of treatment group with a successful outcome _____

CGProp Proportion of comparison group with a successful outcome _____

Significance Tests

T_Value *t*-value _____

F_Value *F*-value (*df* for the numerator must = 1) _____

ChiSquar Chi-square value (*df* = 1) _____

Calculated Effect Size

ES Effect size _____

CR_ES Degree of estimation in effect size computation

1 = Highly estimated (Chi-Square, Frequencies, Proportions)

2 = Moderate estimation (*t*-value, *F*-value)

3 = Low estimation (Means, Standard Deviations)

4 = Other (Specify)

9 = Cannot tell