

The Primary Factors that Characterize Effective Interventions with Juvenile Offenders: A Meta-Analytic Overview

Mark W. Lipsey

Peabody Research Institute, Vanderbilt University, Nashville, Tennessee, USA

Abstract: Previous meta-analyses have identified many effective interventions for reducing the recidivism of juvenile offenders and various program factors that are associated with the best outcomes. Most of that work has been focused on only one intervention area and thus has limited scope. Notable exceptions are two relatively comprehensive meta-analyses that have identified a small number of factors or principles that appear to characterize the most effective programs. This paper presents a new analysis of data from one of those meta-analyses designed to test a broader range of intervention factors in a manner that allows identification of both the general principles and the distinct intervention types associated with the greatest reductions in recidivism. Only three factors emerged as major correlates of program effectiveness: a “therapeutic” intervention philosophy, serving high risk offenders, and quality of implementation. With other variables statistically controlled, relatively few differences were found in the effectiveness of different types of therapeutic interventions.

Keywords: juvenile delinquency, rehabilitation, evaluation research, meta-analysis

Meta-analytic reviews of research on the effects of interventions with juvenile offenders have provided ample evidence that a rather broad range of such

The construction of the meta-analysis database on which the analyses in this paper are based was supported in part by grants from the National Institute of Mental Health, the Office of Juvenile Justice and Delinquency Prevention, and the Russell Sage Foundation. Thanks to Paul Gendreau, James Howell, and Darin Carver for useful comments on this paper.

Address correspondence to Mark W. Lipsey, Peabody Research Institute, Vanderbilt University, Box 0181 GPC, 230 Appleton Place, Nashville, TN 37203. E-mail: mark.lipsey@vanderbilt.edu

interventions reduces recidivism (Lipsey & Cullen, 2007). The vast majority of those systematic reviews, however, focus on a particular program or type of program, such as boot camps (MacKenzie, Wilson, & Kider, 2001), cognitive-behavioral therapy (Landenberger & Lipsey, 2005), prison visitation (Petrosino, Turpin-Petrosino, & Buehler, 2003), family therapy (Latimer, 2001), drug court (Wilson, Mitchell, & MacKenzie, 2006), victim-offender mediation (Nugent, Williams, & Umbreit, 2003), multisystemic therapy (Littell, Popa, & Forsythe, 2005), and the like. While those reviews are individually informative about the respective interventions, they provide only a limited angle of vision on the broad patterns that characterize the whole body of research on the effectiveness of programs for juvenile offenders.

Rather than focusing on a predefined type of intervention, an alternate approach is to collect and meta-analyze all the available research on the effects of intervention with juvenile offenders, sorting it according to the types of interventions found, whatever they may be. Though a daunting task, this approach makes it possible to investigate certain important issues that are otherwise difficult to address.

First, a comprehensive meta-analysis of this sort brings to light a number of program types that are unlikely to receive scrutiny in more focused reviews. Much of the delinquency intervention research involves rather generic kinds of programs not likely by themselves to attract the attention of a reviewer. For instance, a large number of studies have been conducted of service-broker-type programs—referral of juveniles to different services based on some assessment of their needs, a kind of program often used with diversion cases. Many of the programs actually used in the juvenile justice system are of this sort and it is informative to consider what is known about their effectiveness along with that of their more crisply defined counterparts.

Another reason to examine the full body of research on delinquency programs in a single meta-analysis is that it allows an integrated analysis of the comparative effectiveness of different program types and approaches. A meta-analysis of, say, cognitive-behavioral programs may demonstrate that they have positive effects on recidivism while another meta-analysis shows that family counseling also has positive effects. But which programs are most effective and for whom and under what circumstances? Answers to those questions are especially critical for practitioners interested in using the most effective programs applicable to their situations. Such comparative assessments are not easy to make across different meta-analyses. The task is not as simple as determining which shows the largest mean effect sizes. Effect sizes are influenced by variation in the subject samples and settings used in the primary studies, by the research methods applied in those studies, and by the procedures employed by the meta-analyst in representing and analyzing the intervention effects. Under these circumstances, simple comparisons of summary effect sizes can be very misleading. Within an integrated meta-analysis, however, common procedures

can be applied and statistical controls used to help level the playing field in a uniform manner so that comparative effectiveness can be better assessed.

The most important advantage of a comprehensive meta-analysis, however, is the opportunity it provides to search for generalizations about the factors associated with effective programs (Cook, 1993). The most useful guidance for practitioners, and the most informative perspective for program developers and researchers, will not come from lists of the names of programs shown by research to have positive effects. Rather, they will come from identification of the factors that characterize the most effective programs and the general principles that characterize “what works” to reduce the recidivism of juvenile offenders.

Various attempts have been made over the years to conduct more or less comprehensive meta-analyses of the research on interventions for juvenile offenders. The two most extensive efforts are those of Don Andrews and his colleagues (e.g., Andrews et al., 1990) and the present author (e.g., Lipsey, 1992; Lipsey & Wilson, 1998). Andrews et al. have focused especially on identifying the principles that characterize effective interventions for offenders (Andrews et al., 1990; Gendreau, 1996). From analysis of the available research guided by their theory of criminal behavior, they have put forward a need principle, a responsivity principle, and a risk principle for explaining the likelihood of positive effects on recidivism. According to the need principle, interventions have larger effects on recidivism if they address criminogenic needs—malleable risk factors predictive of subsequent criminal conduct such as antisocial attitudes and peer associations, self-control and self-management skills, drug dependencies, and the like. The responsivity principle identifies treatment capable of altering those criminogenic needs, especially interventions that use cognitive-behavioral and social learning approaches. The risk principle, in turn, indicates that larger effects are found for higher risk offenders who, thereby, have a greater need for treatment and also more room for improvement.

The meta-analyses conducted by Andrews et al. have demonstrated that studies of interventions they judge as conforming to their need, responsivity, and risk principles show larger effects on recidivism than those that do not (Andrews & Bonta, 2006; Andrews et al., 1990; Dowden & Andrews, 1999, 2000; Gendreau, Smith, & French, 2006). One of those meta-analyses, for instance, reported that programs departing from the need, responsivity, and risk principles had a mean effect size of virtually zero—whereas those that conformed to these principles achieved an effect size of $\phi = .26$, equivalent to a recidivism reduction of around 50% (Andrews & Bonta, 2006, p. 335).

The meta-analysis work of the present author and his colleagues, in contrast, has been largely atheoretical and descriptive. It has involved a large database of studies and has focused on identification of the correlates of recidivism effects—that is, the characteristics of study methods, programs, offenders, and intervention circumstances most strongly associated with the differences between treatment and control recidivism rate (Lipsey, 1992, 1999, 2006;

Lipsey & Wilson, 1998; Wilson, Lipsey, & Soydan, 2003). These analyses have resulted in what, at a global level, is a relatively simple picture of the main factors related to recidivism effects. These fall into four categories. First, a considerable amount of the variability in observed intervention effects is associated with the methods used by the researchers to study those effects rather than substantive characteristics of the intervention. Moreover, methodological and substantive factors are often confounded in ways that make it difficult to disentangle actual program effects from methodological artifacts (Lipsey, 2003).

The three categories of substantive factors most strongly associated with intervention effects are the intervention approach and modality (type of treatment), the quantity and quality of treatment provided, and the characteristics of the juveniles receiving that treatment. Consistent with the Andrews et al. framework, these meta-analytic investigations have found relatively large positive effects associated with cognitive-behavioral and skill building programs but have also sometimes found comparable effects from different approaches (e.g., general counseling). Additionally, in some analyses, the quality with which the intervention is implemented has been as strongly related to recidivism effects as the type of program, so much so that a well-implemented intervention of an inherently less efficacious type can outperform a more efficacious one that is poorly implemented. Quality of implementation is not well documented in the respective research reports, however, and is most evident in the form of a proxy variable—involvement of the researcher/developer in the delivery of the program. That factor allows for other interpretations, but mainly differentiates programs mounted for research and development purposes, presumed to be more carefully implemented and monitored, from those used in routine practice by juvenile justice agencies.

The characteristics of the juvenile offenders treated, by contrast, have proven to be relatively unrelated to the recidivism effects of those interventions. In particular, recidivism effects are not strongly differentiated by age, gender, or ethnicity, at least within the range represented in available evaluation studies. That range is notably restricted for gender, however—the vast majority of studies have used all male or nearly all male samples. The one clear exception to the relatively small differential effects associated with characteristics of the juveniles is risk. As in the Andrews et al. findings, larger recidivism effects have been found for higher risk juveniles (e.g., those with more prior offenses) in the Lipsey et al. analyses.

Against this background, the meta-analysis reported in the present paper has a twofold purpose. The main factors associated with intervention effects identified above have emerged in the Lipsey et al. analyses in a somewhat piecemeal fashion with different analyses and reports examining different aspects, sometimes in different ways with somewhat different results. The analysis presented here aims to investigate those factors in a single integrated

approach that adopts a consistent analytic framework oriented toward two objectives: (1) using that consistent framework to identify the general factors associated with program effects—that is, the descriptive principles that characterize effective programs for juvenile offenders; and (2) providing a balanced, adequately controlled comparative analysis of the differential effectiveness of different intervention modalities.

The second purpose of the analysis is to incorporate and examine the influence of two factors that have not been well represented in previous work. One of these is the extent of juvenile justice control and supervision. Delinquency programs are often identified in terms that conflate the juvenile justice supervisory structure and the treatment provided within that structure. “Programs” such as intensive probation supervision, diversion, boot camps, group homes, and the like may or may not have constructive change-oriented treatments embedded within them. In this analysis we attempt to distinguish variables describing the general level of juvenile justice supervision (e.g., diversion, probation, incarceration) from treatment modalities such as counseling, skill training, and the like. The key question is whether the effectiveness of a given treatment modality is influenced by the associated level of juvenile justice supervision—for example, whether social skills training is equally effective for juveniles in correctional facilities as for those under probation supervision.

In a related vein, this analysis also examines the differential effectiveness of interventions in terms of superordinate categories that contrast broadly different treatment approaches. This categorization brings together treatment modalities that share an intervention philosophy (e.g., deterrence, discipline, skill building, relationships) so that the comparative effectiveness of the different approaches can be assessed along with that of the specific treatment modalities within each approach. Of particular interest is the distinction between what can be called therapeutic program approaches that attempt to engage the youth in a supportive, constructive process of change and approaches that rely more on external control and coercion (e.g., through discipline or surveillance).

In pursuit of these objectives, a new analysis of the Lipsey et al. meta-analytic database of studies of interventions with juvenile offenders was undertaken. In preparation for that analysis, several new variables were defined and coded from the studies represented in that database and others were reconfigured to better support the comparisons of interest.

METHODS

The Meta-Analytic Database

The data used for analysis were based on 548 independent study samples for which information was extracted from 361 primary research reports, many

of which described more than one distinct study or substudy. This meta-analytic database is an updated version of the one described more extensively in previous publications (e.g., Lipsey, 1992; Lipsey & Wilson, 1998). The research reports from which it was derived span the period from 1958 to 2002 and were selected during an extensive search to meet the following criteria:

- Juveniles aged 12–21 received an intervention intended to have positive effects on their subsequent delinquency. Excluded from the present review were studies of variations in juvenile justice processing, such as standard probation versus diversion with no services, and those for which the type of service could not be identified.
- Quantitative results were reported for a comparison between a treatment condition and a control condition for at least one delinquency outcome measure. In addition, the assignment of juveniles to conditions was random or, if not, pretreatment differences were reported or matched. Random assignment was used for 42% of the study samples in the present review and 28% used groups matched on offense histories and/or key demographic characteristics. The remaining 30% did neither but reported pretest differences that were coded and used as control variables.
- The study was conducted in an English-speaking country and reported in English. More than 90% of the studies located were conducted in the United States.

Studies meeting these criteria were retrieved and coded by trained personnel on more than 150 items describing study methods and procedures, subject characteristics, treatment and program characteristics, outcome effect sizes, and other related matters.

Recidivism outcomes in these studies were reported in many different forms, mainly the proportion of juveniles arrested, convicted, or incarcerated or the mean number of such events during various time periods after the conclusion of the intervention. The large number of studies in this database included many with more than one such outcome for the same juvenile sample and intervention. Those cases were used to identify functional relationships among the different forms of recidivism and time periods that were then applied to adjust each outcome to better approximate the most common form of recidivism reported in these studies—the proportion of juveniles rearrested during the 12 months after intervention.

To further standardize the recidivism outcomes so they would be comparable across studies, each was converted to a standardized effect size. The phi coefficient was selected for this purpose; that is, the product-moment correlation between condition (treatment/control) and recidivism (yes/no). To eliminate any sensitivity of the phi coefficient to unequal sample sizes in the treatment

and control groups (McGrath & Meyer, 2006), it was computed from the success versus recidivism proportions for each group rather than the frequencies. The phi coefficient has the advantages of being appropriate for dichotomous outcomes and somewhat less sensitive to differences in the control group recidivism base rate than, say, the odds ratio.

The phi coefficient has the disadvantage that it presents relatively small numerical values for effects that may have practical significance. For example, consider a 12-month recidivism rate of .50 for a control group; that is, 50% of the juveniles were rearrested within 12 months. A 10% reduction in recidivism from that base rate, which might be judged a worthwhile improvement in many practical contexts, is represented by a phi of only .05 and a 20% reduction corresponds to a phi of .10. Care must therefore be taken not to minimize the importance of modest phi values simply because they are numerically small relative to what we usually expect from meaningful correlation coefficients.

For all analyses, the phi coefficients were Fisher-z transformed and outliers more than 1.5 times the interquartile range beyond the 25th and 75th percentiles (Tukey, 1977) were recoded to the respective inner fence values. Inverse variance weights were applied in all analyses following standard meta-analytic procedures (Hedges & Olkin, 1985; Lipsey & Wilson, 2001). Outliers among the sample sizes were recoded to the 95th percentile value to prevent a few extraordinarily large samples from having an excessive influence on the weighting function. Random effects models were used throughout, implemented with SPSS macros developed for that purpose by David Wilson (Lipsey & Wilson, 2001).

Overall Recidivism Effects

The mean random effects phi coefficient across all 548 study samples was .062. This overall mean is statistically significant (95% confidence interval, .048 to .075) but it tells us very little about the effects of the interventions implemented in these studies. The Q-test for the heterogeneity of the effect sizes in this distribution showed that there was an enormous amount of between-study variability ($Q = 2134.9$, $df = 547$, $p < .0005$). Some effect sizes were thus very large—the 75th percentile was .12 and the 90th percentile was .23—and others were very small, even negative—the 25th percentile was $-.02$ and the 10th percentile was $-.09$. In these circumstances, the grand mean does not characterize any of the effect sizes very well.

Moderator Variables

The descriptive variables coded for each study were organized into thematic groups representing the main factors expected to influence the magnitude of study effects. Within each group, some variables were selected for inclusion in the analysis because they provided the primary summary information for a particular characteristic (e.g., the proportion of males in the

study sample). In other cases, more than one variable presented relevant information and they were combined into a composite variable—for example, initial differences between treatment and control groups on gender, age, ethnicity, and prior offense history. The categories of moderators and the variables selected to represent them are described below.

Characteristics of the study methods. Variables in this category were intended as control variables to adjust for methodological differences among studies that might influence effect sizes. One set of these related to the way recidivism was measured. As noted earlier, all the recidivism rates were converted to the approximate equivalent of rearrest recidivism during the 12 months subsequent to intervention if they were not already in that form. To check on the adequacy of those conversions and control for any remaining variability associated with the way recidivism was measured, the following variables described the form of recidivism used in the original study:

- Court contact/conviction (yes/no dummy code); 21% of the studies
- Incarceration/institutionalization (yes/no dummy code); 12% of the studies
- Mean frequency of arrests (yes/no dummy code); 12% of the studies
- Self-report delinquency (yes/no dummy code); 3% of the studies
- Probation violations (yes/no dummy code); 6% of the studies
- Weeks of exposure in the delinquency period (ranging from 4 to over 300); 59% of the studies

A second set of methodological control variables dealt with variations in the study design.

- Random assignment (yes/no dummy code); 42% of the studies
- Matched control group (yes/no dummy code); 28% of the studies
- Unmatched control group (yes/no dummy code); 30% of the studies
- Outcomes were covariate adjusted for pretest differences (yes/no dummy code); 4% of the studies
- Direction and magnitude of initial differences between intervention and control group. This composite variable was created as the principal components factor score from a set of codes for whether any initial differences in age, gender, ethnicity, prior offense history, or severity of offenses favored the intervention or control group and the magnitude of those differences. Not all studies reported initial differences on all these variables. Missing values were imputed using maximum likelihood estimates based on those differences that were reported plus other variables relating to the overall nature of the design.

A third set of methodological variables dealt with attrition from the intervention and control samples on the recidivism outcome. These variables were represented as the proportion of the sample initially assigned to each condition that was not included in the recidivism data collected after the intervention period. For attrition from the intervention sample, the mean was 7.4%; from the control sample it was 7.1%.

A fourth set of variables was included to address possible publication bias, the tendency for studies with small samples and nonsignificant effects to be censored out of the published research literature (Rothstein, Sutton, & Borenstein, 2005). These included sample size and a code for whether the primary study report was a journal article or book chapter versus a technical report or thesis (yes/no for journal/book); 35% of the studies were reported in journal articles or book chapters. It should be noted that relationships with publication source and sample size must be interpreted with some caution. Many unpublished studies were retrieved for this meta-analysis which should diminish the influence of any publication bias and, additionally, small sample studies may show larger effect sizes for reasons other than publication bias (e.g., they may be better implemented).

Characteristics of the juvenile samples. There are many characteristics of the participating juveniles that might relate to their responsiveness to intervention. For inclusion in meta-analysis, we are restricted to those reported widely enough to support analysis—mainly basic demographic characteristics. The following variables were available in this category:

- Mean age of the study sample at the time of intervention; overall mean = 15.5
- Gender mix of the sample (4-point scale from “no males” to “all males”); 87% were all male or mostly male, only 4% were all female
- Predominant ethnicity of the sample (Anglo, mixed, minority); 43% were mixed, 34% Anglo, and 23% minority
- Delinquency risk for the sample. This composite variable summed a 4-point scale on the proportion of the sample with prior offenses and a 4-point categorization of the recidivism rate of the control sample after adjustment for methodological variables. The resulting risk scores ranged from 2 (no juveniles with priors; control recidivism <.30) to 8 (all juveniles with priors; control recidivism >.80) with an overall mean of 5.4.
- History of aggressive/violent behavior for the sample, a 4-point scale ranging from none of the juveniles to all of the juveniles. The overall mean was 1.9; for only 6% of the samples were most (>50%) or all of the juveniles reported to have prior offenses or behavior indicating aggression or violence.

Juvenile justice supervision and control. The juveniles represented in the research were in varying stages of penetration into the juvenile justice system. These different levels constitute rather different contexts for intervention and, as such, may influence the effectiveness of the respective interventions. The levels of juvenile justice supervision and control were represented as a set of dummy coded variables for the following categories:

- No supervision—juveniles were in the community with no juvenile justice supervision (prevention programs); 18% of the study samples
- Diversion—juveniles had been diverted to community treatment after law enforcement or juvenile court contact; 27% of the study samples
- Probation or parole—juveniles were in the community but under the supervision of the juvenile court or parole officers; 33% of the study samples
- Incarceration—juveniles were in a juvenile correctional institution; 22% of the study samples

Type of intervention. There is no established taxonomy for characterizing interventions for juvenile offenders. The scheme used here was developed inductively from the program and service descriptions appearing in research reports, guided by prior research and the conceptual distinctions mentioned earlier that were of specific interest for the analysis. A hierarchical organization was used with the top level differentiating broad treatment philosophies and particular types of intervention nested underneath. Neither the broad philosophies nor the particular intervention types were treated as mutually exclusive categories. Any given intervention might be coded in more than one category if it presented major treatment components that spanned multiple categories. For example, a boot camp program with a paramilitary disciplinary regimen and daily group counseling sessions would be coded for both discipline and group counseling.

The seven intervention philosophies that were identified from the research studies are listed below with the specific program types listed afterward where applicable. The number of study samples represented in each category and subcategory is shown in parentheses.

- *Surveillance* (N = 17). Interventions in this category are based on the idea that closer monitoring of the juvenile will inhibit reoffending. The main program of this sort is intensive probation or parole oriented toward increasing the level of contact and supervision. Such programs also sometimes include additional services but the surveillance component refers only to the monitoring; any major services were coded in one of the categories below.

- *Deterrence* (N = 15). Interventions in this category attempt to deter the youth from reoffending by dramatizing the negative consequences of that behavior. The prototypical program of this sort is prison visitation—“scared straight” type programs in which juvenile offenders are exposed to prisoners who graphically describe the aversive nature of prison conditions.
- *Discipline* (N = 22). The theme of these interventions is that youth must learn discipline to succeed in life and avoid reoffending and that, to do so, they need to experience a structured regimen that imposes such discipline on them. The main programs of this sort are paramilitary regimens in boot camps.
- *Restorative programs* (N = 41). Programs of this sort aim to repair the harm done by the juvenile’s delinquent behavior by requiring some compensation to victims or reparations via community service. They may also involve some form of direct reconciliation between victims and offenders. Two different intervention types appear in the research, sometimes combined in the same program:
 - Restitution (N = 32). Offenders provide financial compensation to the victims and/or perform community service.
 - Mediation (N = 14). Offenders apologize to their victims in spoken or written form and may meet with them under supervision. These interventions typically also include a restitution component.
- *Counseling and its variants* (N = 185). This diverse and popular program approach is characterized by a personal relationship between the offender and a responsible adult who attempts to exercise influence on the juvenile’s feelings, cognitions, and behavior. Family members or peers may also be involved and the peer group itself may take the lead role in the relationship. The major variants on this intervention approach that appear in sufficient numbers in the research to warrant separate consideration are the following:
 - Individual counseling (N = 12)
 - Mentoring by a volunteer or paraprofessional (N = 17)
 - Family counseling (N = 29)
 - Short term family crisis counseling (N = 13)
 - Group counseling led by a therapist (N = 24)
 - Peer programs in which the peer group plays much of the therapeutic role; for example, guided group interaction programs (N = 22)
 - Mixed counseling—combinations of any of the above but especially individual, group, and/or family (N = 39)

- Mixed counseling with supplementary referrals for other services, a common form for diversion programs (N = 29)
- *Skill building programs* (N = 169). These programs provide instruction, practice, incentives, and other such activities and inducements aimed at developing skills that will help the juvenile control their behavior and/or enhance their ability to participate in normative prosocial functions. The main forms of these programs are the following:
 - Behavioral programs—behavior management, contingency contracting, token economies, and other such programs that reward selected behaviors (N = 30)
 - Cognitive-behavioral therapy (N = 14)
 - Social skills training (N = 18)
 - Challenge programs—interventions that provide opportunities for experiential learning by mastering difficult or stressful tasks (N = 16)
 - Academic training—for example, tutoring and GED programs (N = 41)
 - Job related interventions—vocational counseling and training, job placement (N = 70)
- *Multiple coordinated services* (N = 138). Programs in this category are not organized around a primary service type or a combination of a few such service types but, rather, are designed to provide a package of multiple services which may be basically similar for all the participating juveniles or may be individuated with different juveniles receiving different services. The primary intervention forms of this type are the following:
 - Case management—a designated case manager or case team develops a service plan for each juvenile, arranges for the respective services, and monitors progress (N = 58)
 - Service broker—referrals are made for the service or services deemed appropriate for each juvenile with a relatively minimal role for the broker afterwards (N = 49)
 - Multimodal regimen—a multimodal curriculum or coordinated array of services is provided to all participating juveniles, often occurring in a residential setting (N = 32)

Amount and quality of service. The reports for most studies did not provide much detail about the operational aspects of the intervention. However, information about two broad dimensions—quantity and quality—was extracted as follows.

- *Amount of service.* This is described by two variables, one indicating the approximate duration of service (number of weeks from start to end) and the other indicating the approximate number of total contact hours. Because of the limited reporting of this information, both of these are rather crudely measured and mainly represent the intended quantity of service rather than what was actually received by the participants. Different reports provided information in different forms—total hours or number of sessions, frequency of contact, hours per week, period over which service was delivered, and the like—but nearly all reported something. Missing values on the duration and total contact variables were imputed using maximum likelihood estimates based on what information was available and the type of intervention.
- *Quality of implementation.* This dimension was characterized, albeit crudely, with a composite variable that combined two correlated features. One was a simple coding of whether the study report indicated any problems with the implementation of the program—high dropout rates, staff turnover, poorly trained personnel, and the like—which was coded simply “yes” (39% overall), “maybe” (15%), and “no” (46%). The other was a 4-point scale describing the extent to which the researcher was involved in the delivery of the intervention, ranging from a research role only (54% overall) through planning and supervising the intervention (24%) and delivering the treatment directly (4%). This continuum mainly differentiated research and development programs from those implemented in routine practice. For present purposes, it was taken as a proxy for the extent to which attention was given to specifying the intended intervention, implementing it as intended, monitoring the implementation, and taking corrective action when problems arose—all of which were mentioned as aspects of service delivery in some study reports.

RESULTS

Major Factors Influencing Recidivism Effects

The major categories of moderator variables described above are (a) the characteristics of the study methods, (b) the characteristics of the juvenile samples, (c) the level of juvenile justice supervision, and (d) the treatment philosophy on which the intervention is based. A random effects multiple regression analysis was conducted to examine the relative influence of each of these factors on the recidivism effect sizes. The methodological variables were included primarily to statistically control for variation in effect sizes associated with study methods. In preliminary analyses, the method variables from the list presented above that did not show statistically significant

relationships at the level of $p = .10$ or less were dropped stepwise from the analysis with the exception of the basic design variables relating to randomization.

The results of this regression analysis are presented in Table 1, which shows several influential relationships. First, it indicates that methodological differences among the studies account for a significant portion of the variation in recidivism effect sizes. This highlights the importance of controlling for the method variables before drawing conclusions about the influence of the substantive features of the intervention. Second, above and beyond the methodological differences, the characteristics of the juvenile samples and the general intervention philosophy are significantly associated with recidivism outcomes when the other variables in the analysis are held constant.

Finally (and quite notably, with the other variables controlled), there is no significant relationship in this overall analysis between recidivism effects and

Table 1: Regression results for the prediction of recidivism effect sizes from the major moderator variables.

	B ^a	β^b	Q-Added ^c
Constant	-.1393		
Method controls			51.4 (df = 6) $p < .001$
Arrest frequency recidivism	-.0443*	-.08	
Sample size	-.0001*	-.12	
Design: Matched control	.0018	.01	
Design: Unmatched control	.0369*	.11	
Design: Randomized ^d			
Initial differences	.5985*	.16	
Journal/book publication	.0344*	.10	
Juvenile samples			98.6 (df = 5) $p < .001$
Mean age	.0046	.05	
Gender mix	-.0152 [†]	-.07	
Ethnicity	.0062	.03	
Delinquency risk	.0434*	.41	
Aggressive history	-.0358*	-.12	
Juvenile justice supervision			5.4 (df = 3) $p = .14$
No supervision ^d			
Diversion	.0195	.06	
Probation/parole	.0020	.01	
Incarceration	-.0314	-.08	
Intervention philosophy			15.6 (df = 7) $p < .03$
Surveillance	-.0207	-.03	
Deterrence	-.0619	-.06	
Discipline	-.0932*	-.11	
Restorative	-.0030	-.01	
Counseling	.0132	.04	
Skill building	.0072	.02	
Multiple services	.0093	.03	

[†] $p < .10$, * $p < .05$

(a) Unstandardized regression coefficients. (b) Standardized regression coefficients. (c) Q-test for the variance added by each group of variables. (d) Omitted as the reference category in a group of dummy codes.

the level of juvenile justice supervision. That is, assuming juveniles of similar characteristics (i.e., risk, age, gender, ethnicity) and similar intervention approaches, the effects of those interventions are not significantly different whether the youth is treated in the community, after diversion, while on probation or parole, or while incarcerated. Juvenile justice supervision is related to risk of course, with higher risk samples more likely to have higher levels of supervision ($r = .45$ with supervision coded from 1 = no supervision to 4 = incarceration). This presumably reflects considerations of public safety subsequent to apprehension for an offense as well as some matching of the punitive consequences with the severity of the offense. With risk statistically controlled, however, this analysis shows no significant influence of the level of supervision on later recidivism.

Within the category of juvenile characteristics, the results in Table 1 reveal that two variables have especially strong relationships to recidivism effects. The largest relationship by far was with overall delinquency risk, with larger effect sizes (greater recidivism reductions) associated with higher risk juveniles. In addition, juvenile samples with aggressive/violent histories showed smaller recidivism effects. There was also a marginally significant relationship for the gender mix of the sample—those with higher proportions of males had less favorable recidivism outcomes. The mean age and ethnic mix of the sample were not related to the magnitude of the recidivism effects.

The significant differences among the interventions representing different philosophies shown in Table 1 involved only one individually significant relationship—all else being equal, programs emphasizing discipline showed notably smaller recidivism effects than the other approaches. The regression coefficients indicate the relative effectiveness of these different thematic groupings with positive signs generally indicating above average effects and negative signs indicating below average ones. Counseling interventions had the largest effects on recidivism followed by multiple services, skill building, restorative programs, surveillance, deterrence, and discipline. This relative comparison, however, does not tell us how large (or small) the recidivism effects were for each of these intervention types. To determine that, we need to examine the covariate adjusted effect sizes for each of these categories—that is, the mean phi coefficient for each when all the other variables in the equation are held constant.

Table 2 reports such covariate adjusted phi coefficients generated as predictions from the regression model of Table 1 under the following assumptions: (a) the method variables are assumed optimal for the recidivism measure (12-month rearrest), the design (randomized), and the initial treatment-control differences (none) and are assumed average for sample size and publication source; (b) the juvenile characteristics are assumed average (mean age, gender mix, ethnicity, risk, and aggressive history); (c) juvenile justice supervision is assumed to be probation/parole; and (d) the intervention philosophy is assumed

Table 2: Covariate adjusted mean recidivism effect sizes for the different intervention philosophies.

Intervention Philosophy	Mean Phi Coefficient ^a	Recidivism Rate ^b	Percentage Difference ^c
Counseling	.066	.43	-13%
Multiple services	.062	.44	-12%
Skill building	.060	.44	-12%
Restorative	.050	.45	-10%
Surveillance	.032	.47	-6%
Deterrence	-.009	.51	+2%
Discipline	-.040	.54	+8%

(a) Covariate adjusted as described in the text. (b) Recidivism rate for the intervention group that corresponds to the effect of the given phi coefficient on a control recidivism rate of .50. (c) Recidivism reduction (or increase) for the intervention group compared to a control group with a .50 recidivism rate.

to be each of the seven types in turn. The overall median recidivism rate for the control groups on probation in these studies was .49 which, to simplify comparisons, has been rounded off to .50. Table 2 shows the mean recidivism rate for the intervention group that corresponds to each covariate adjusted phi coefficient and the percent reduction from a .50 control recidivism rate that it represents.

What Table 2 reveals is that the mean covariate adjusted phi coefficient effect sizes for recidivism represent reductions of 10–13% from a mean control group recidivism rate of .50 for counseling, multiple services, skill building, and restorative interventions. Because of the nature of these interventions and their positive effects on recidivism, we will call these the therapeutic intervention types. Especially notable is that, when the characteristics of the methods and the juvenile participants are held constant, the differences in effectiveness among these therapeutic intervention types are negligible—other than restorative programs showing slightly (and nonsignificantly) smaller effects. Among the nontherapeutic intervention types, surveillance shows positive effects on recidivism, but they are about half the magnitude of those for the therapeutic interventions. Deterrence and discipline, on the other hand, show net negative effects—that is, the intervention groups have higher recidivism rates than the control groups. For deterrence that effect is small (virtually zero), but for discipline it is larger—an average 8% increase in recidivism.

Differential Effectiveness Within the Therapeutic Intervention Approaches

The analysis shown in Table 1 provides an overview of the major factors influencing the effects of interventions on recidivism but it leaves much unexplained.

Statistically, that regression model accounts for 38% of the between-study variation in recidivism effects, leaving 62% unaccounted for. From the standpoint of assessing the effectiveness of different intervention philosophies, there was much more variation within the various approaches than there was between them. There were insufficient numbers of studies to support further analysis of the interventions in the nontherapeutic categories but ample numbers of restorative, counseling, skill building, and multiple service programs. Separate random effects regression analyses were therefore conducted for each of these intervention categories. These analyses included the same method control variables that were used in the overall analysis (shown in Table 1) except for the restorative program group where the smaller number of studies required a reduced set. They also included the same descriptive variables for the characteristics of the juvenile sample and the level of juvenile justice supervision.

In place of the variables representing the intervention philosophies in Table 1, the analysis within each philosophy type differentiated the specific interventions associated with that philosophy. For example, the analysis of restorative programs examined the relative effects of interventions with a primary restitution component and those with a primary mediation component. In addition, the variables for amount and quality of service were added to the analysis (described earlier in the section on moderator variables). These regression models each accounted for a significant portion of the respective between-study effect size variance (restorative, 46%; counseling, 59%; skill building, 65%; and multiple services, 64%) but also left significant residuals of unexplained variation.

The results of these multiple regression analyses are presented in two different segments. Table 3 reports the standardized regression coefficients for the variables describing the juvenile samples, level of juvenile justice supervision, and service amount and quality. Those coefficients can be read as semipartial correlation coefficients that indicate the strength of the relationship between each predictor variable and the recidivism effect sizes. Table 4 shows the mean covariate adjusted phi coefficient for each of the intervention types along with the corresponding recidivism rate for the intervention condition and the reduction it represents from a .50 control group recidivism rate.

For the intervention types representing each of the four broad treatment philosophies, Table 3 shows, once again, that it is the delinquency risk of the juveniles that is most strongly and consistently related to positive recidivism effects. Effects for juveniles with aggressive/violent histories, however, tend to be smaller—though the coefficients are statistically significant only for counseling interventions and, marginally, for skill building ones. There was also a pattern of larger recidivism effects for older juveniles except for multiple coordinated services.

Table 3: Standardized regression coefficients showing the relationship between recidivism effect sizes and the intervention characteristics common to all studies.

Intervention Characteristics	Intervention Philosophy Category			
	Restorative β^a	Counseling β^a	Skill Building β^a	Multiple β^a
Juvenile samples				
Mean age	.25	.16*	.17 [†]	.02
Gender mix	.12	-.15*	-.07	-.04
Ethnicity	-.22	.12 [†]	.10	.06
Delinquency risk	.21	.46*	.31*	.41*
Aggressive history	-.03	-.23*	-.13 [†]	-.12
Juvenile justice supervision				
No supervision ^b				
Diversion		-.10	.28*	.03
Probation/parole	.10	-.07	.00	-.19
Incarceration		-.29*	.10	.14
Service amount and quality				
Duration of service ^c	-.31	-.02	.09	.11
Total hours of service ^c	.51*	-.06	-.03	-.03
Quality of implementation	.21	.13 [†]	.25*	.18*

[†] $p < .10$, * $p < .05$

(a) Standardized regression coefficient for the relationship of each individual predictor to the effect size. (b) Omitted from the analysis as the reference category for a set of dummy codes. (c) Centered for each intervention type within these categories to represent variation around the mean for that specific intervention.

Note: Method control variables were included in analysis but are not shown.

Aside from delinquency risk, the largest and most consistent relationship with recidivism effects is the quality of program implementation with, of course, higher quality associated with bigger effects on recidivism. Curiously, variation in the duration and total hours of service was not generally related to the effects. These variables were centered on the mean values for each specific intervention type, so the indication here is that (given the mean amount of service overall for that type) variation around that mean within the range represented in these studies did not make much difference to the outcome.

Some of the other relationships shown in Table 3 are more specific to the interventions within a particular treatment philosophy. The level of juvenile justice supervision was generally not related to recidivism effects except that counseling interventions appear less effective for incarcerated juveniles and skill building interventions were more effective for diversion cases. Gender and ethnicity were largely unrelated to the effects, though counseling was less effective for largely male juvenile samples and somewhat more effective for largely minority samples.

The covariate adjusted phi coefficients in Table 4 that estimate the mean recidivism effects for each intervention type within each treatment philosophy are remarkably similar. And, indeed, the Q statistics that test for those

Table 4: Covariate adjusted mean recidivism effect sizes for the different types of interventions within each treatment philosophy.

Intervention Type (N)	Mean Phi Coefficient ^a	Recidivism Rate ^b	Percentage Difference ^c
Restorative programs (B = .017, p = .76)			
Restitution (32)	.045	.46	-9 %
Mediation (14)	.062	.44	-12 %
Counseling (Q = 2.9(7) p = .89)			
Individual (12)	.024	.48	-5%
Mentoring (17)	.108	.39	-21%
Family (29)	.065	.44	-13%
Family crisis (13)	.061	.44	-12%
Group (24)	.109	.39	-22%
Peer (22)	.022	.48	-4%
Mixed (39)	.081	.42	-16%
Mixed with referrals (29)	.040	.46	-8%
Skill building programs (Q = 5.0(5) p = .41)			
Behavioral (30)	.109	.39	-22%
Cognitive-behavioral (14)	.133	.37	-26%
Social skills (18)	.066	.43	-13%
Challenge (16)	.061	.44	-12%
Academic (41)	.051	.45	-10%
Job related (70)	.028	.47	-6%
Multiple coordinated services (Q = 4.4(2) p = .11)			
Case management (58)	.100	.40	-20%
Service broker (49)	.049	.45	-10%
Multimodal regimen (32)	.013	.49	-3%

(a) Covariate adjusted for the variables related to method, juvenile samples, and juvenile justice supervision shown in Table 1 and for any co-occurring services from other treatment philosophies. (b) Recidivism rate for the intervention group that corresponds to the effect of the given phi coefficient on a control recidivism rate of .50. (c) Recidivism reduction (or increase) for the intervention group compared to a control group with a .50 recidivism rate.

differences (and the single regression coefficient that does that for restorative programs) are not statistically significant. Overall, then, all the intervention types within each philosophy have statistically similar recidivism effects. However, the modest number of studies within some of these intervention categories, combined with the random effects analysis model, do not provide much statistical power for differentiating the effects. The point estimates for a few interventions do describe notably larger effects than for others—effects which, if reliable, would have great practical significance. Among the counseling approaches, the mean recidivism reductions for mentoring and group counseling were greater than 20%. Individual counseling, peer-oriented group counseling, and mixed counseling with referrals showed distinctly smaller effects. Among the skill building approaches, behavioral and cognitive-behavioral interventions appear especially effective while job-related ones were less effective.

For multiple coordinated services, case management showed relatively high mean recidivism reductions while multimodal regimens showed small effects—the smallest of any of the therapeutic interventions in any category.

CONCLUSION

Perhaps the most striking finding of the analyses reported here is how few of the factors examined were significantly related to the effects of the interventions on recidivism. One important factor that showed little relationship to the effectiveness of the respective interventions was the level of juvenile justice supervision applied to the participating youth. With risk and other characteristics of the juveniles controlled, most of the interventions were about equally effective irrespective of whether the youth was under no official supervision, had been diverted, was under probation supervision, or was in a juvenile custodial facility. The main exceptions were somewhat reduced effectiveness of counseling approaches with incarcerated juveniles and somewhat greater effectiveness of skill building approaches as prevention programs for juveniles in the community without juvenile justice supervision.

The relative robustness of intervention effects across the levels of penetration into the juvenile justice system gives reassuring support to the view that effective treatment is not highly context dependent. In particular, it is encouraging that good programs can be effective within institutional environments where there is more potential for adverse effects through, for instance, greater association with antisocial peers (Dodge, Dishion, & Lansford, 2006). At the same time, it must be remembered that higher degrees of juvenile justice control are associated with higher risk juveniles who pose greater danger to the community. While there is no indication in the analyses presented here that such control in itself has any favorable effects on subsequent recidivism, it may well inhibit offending during the period of supervision and, indeed, clearly must do so for incarcerated juveniles. Juvenile justice control also constitutes the punitive component of the juvenile justice system and, as such, provides the sanctions that associate negative consequences with bad behavior. Control and treatment thus provide two orthogonal axes for the juvenile justice system, with gradations in the level of control addressing the offense and its immediate aftermath while treatment addresses the potential for further offenses after supervision is lifted.

Further indication of the different nature of control and treatment appeared in the finding that interventions that embodied “therapeutic” philosophies, such as counseling and skills training, were more effective than those based on strategies of control or coercion—surveillance, deterrence, and discipline. Indeed, from the perspective of this broad overview, that difference was one of only three characteristics that clearly distinguished more effective from less effective interventions. The second of these was that interventions applied to

juveniles with higher levels of delinquency risk were more effective, though that effect was offset somewhat if the juveniles had aggressive/violent histories. Third, interventions that were implemented with high quality were more effective; given that and an overall average amount of service, further variations in quantity did not matter greatly. With few exceptions, intervention was equally effective for younger and older juveniles, males and females, and minorities and whites.

In relation to the Andrews et al. (1990) principles of effective correctional treatment, these results clearly support the risk principle. The alignment of the findings with the need and responsivity principles is not so clear. The negligible and even negative effects found for surveillance, deterrence, and discipline strategies are consistent with those principles. Among the therapeutic interventions, however, the Andrews framework emphasizes cognitive-behavioral, social learning, and (more generally) skill building interventions. The analyses presented here did indeed show that such interventions were effective—in fact, the largest mean effect sizes for any intervention type appeared for cognitive-behavioral therapy. But other types of interventions also ranked near the top in effectiveness, notably mentoring and group counseling. It may well be that these derive their effectiveness by targeting criminogenic needs with change strategies that are responsive under the Andrews et al. definition. However, most research reports provide limited information about the interventions used and we were not able to differentiate the targeted needs and change strategies any more specifically than is represented in the rather generic categories of intervention types described above.

The apparent effectiveness of interventions in the generic categories used in the analyses reported here presents a contrast with the conventional way of thinking about programs for juvenile offenders. These analyses controlled for differences in study methods, juvenile characteristics, level of juvenile justice supervision, and quality of implementation in an attempt to make a fair comparison of the inherent effectiveness of the different intervention types net of these other influences. It is notable both that so many showed rather impressive mean effects—equivalent to recidivism reductions of 20% or more—and that there were so few statistically reliable differences among them in those effects. The random effects analysis model used here is not generous about conferring statistical significance and may understate meaningful differences among the intervention types. Nonetheless, any allowance for error in the effect estimates leads to the conclusion that many of the different types of intervention have very similar positive effects on recidivism.

Current perspectives on evidence-based practice, in contrast, focus on the distinctive character of brand name “model” programs and the research findings for those specific programs. Various agencies and groups have put forward lists of programs supported by research that meets their criteria for demonstrating effectiveness—for example, the Blueprints for Violence Prevention, the

National Registry of Evidence-based Programs and Practices (NREPP), and the Office of Juvenile Justice and Delinquency Prevention's Model Programs Guide. Implicit in that perspective is the assumption that effectiveness depends upon following one of relatively few program recipes that have the capability to work. The interventions represented in the program categories used in this analysis, however, include very few model programs or even named programs of any generally recognized sort. Moreover, there is quite a bit of variability within these categories in the way the programs of that intervention type are designed and implemented. Nonetheless, the findings presented here indicate that the average program of this rather variable generic sort can be quite effective if implemented well and targeted on high risk offenders. It does not take a magic bullet program to impact recidivism, only one that is well made and well aimed.

Most of the programs in actual use in the juvenile justice system are not name brand model programs; they are more like the programs represented in this meta-analysis. But if the effectiveness of generic programs is good news for the world of juvenile justice practice, the bad news is the extent to which those effects depend on high quality implementation directed toward high risk offenders. The quality standard set by the programs included in this meta-analysis, and reflected in the main index of quality of implementation used in the analysis, is that of a research or demonstration program in which the researcher is involved in supervising and perhaps even delivering the intervention. In such circumstances, which described 46% of the studies in the meta-analysis, attaining high fidelity to the program-as-intended is an objective of the research process—the efficacy of the intervention can only be tested if what is actually implemented represents the intervention well. Particular care is taken, therefore, to ensure that the treatment providers are properly trained and supervised, that the service delivery is monitored, and that corrective actions are taken when quality falls off. Practitioners who wish to use generic counseling, skill building, or multiservice programs to attain the magnitude of recidivism reductions shown in this meta-analysis will also have to match at least the average quality with which those programs were implemented. There was no indication in the meta-analysis, or our reading of the studies that went into the meta-analysis, to suggest that this could not be attained in routine practice. But there is also little reason to believe that it is actually attained in everyday practice. In an earlier analysis with this same database (Lipsey, 1999), the mean recidivism effect size for routine practice programs was found to be about half that of the research and demonstration programs. And that was for the routine practice programs that were selected for evaluation, a group itself likely to be above average. There thus may be relatively few major factors that must be optimized for a program to be effective in substantially reducing recidivism, but there is little reason to believe that current practice is anywhere near optimal.

REFERENCES

- Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Newark, NJ: Lexis/Nexis.
- Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, 28, 369–404.
- Cook, T. D. (1993). A quasi-sampling theory of the generalization of causal relationships. *New Directions for Program Evaluation*, 57, 39–82.
- Dodge, K. A., Dishion, T. J., & Lansford, J. E. (Eds.). (2006). *Deviant peer influences in programs for youth: Problems and solutions*. New York: Guilford.
- Dowden, C., & Andrews, D. A. (1999). What works for female offenders: a meta-analytic review. *Crime and Delinquency*, 45, 438–452.
- Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent reoffending: A meta-analysis. *Canadian Journal of Criminology*, 42, 449–467.
- Gendreau, P. (1996). The principles of effective intervention with offenders. In A. T. Harland (Ed.), *Choosing correctional interventions that work: Defining the demand and evaluating the supply* (pp. 117–130). Newbury Park, CA: Sage.
- Gendreau, P., Smith, P., & French, S. A. (2006). The theory of effective correctional intervention: Empirical status and future directions. In F. T. Cullen, J. P. Wright, & K. R. Blevins (Eds.), *Taking stock: The status of criminological theory: Vol. 15. Advances in criminological theory* (pp. 419–446). New Brunswick, NJ: Transaction.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- 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. (2001). A meta-analytic examination of youth delinquency, family treatment, and recidivism. *Canadian Journal of Criminology*, 43, 237–253.
- Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In T. D. Cook, H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, & F. Mosteller (Eds.), *Meta-analysis for explanation: A casebook* (pp. 83–127). New York: Russell Sage Foundation.
- Lipsey, M. W. (1999). Can rehabilitative programs reduce the recidivism of juvenile offenders? An inquiry into the effectiveness of practical programs. *Virginia Journal of Social Policy & the Law*, 6, 611–641.
- Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The Annals of the American Academy of Political and Social Science*, 587, 69–81.
- 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, & J. E. Lansford (Eds.), *Deviant peer influences in programs for youth: Problems and solutions* (pp. 162–184). New York: Guilford Press.
- Lipsey, M. W., & Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annual Review of Law and Social Science*, 3, 279–320.
- Lipsey, M. W., & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D. P. Farrington (Eds.), *Serious*

- and violent juvenile offenders: Risk factors and successful interventions* (pp. 313–345). Thousand Oaks, CA: Sage.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Littell, J. H., Popa, M., & Forsythe, B. (2005). *Multisystemic therapy for social, emotional, and behavioral problems in youth aged 10–17*. Retrieved January 22, 2009 from http://www.campbellcollaboration.org/review_list/index.php
- MacKenzie, D. L., Wilson, D. B., & Kider, S. B. (2001). Effects of correctional boot camps on offending. *Annals of the American Academy of Political and Social Science*, 578, 126–143.
- McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of *r* and *d*. *Psychological Method*, 11, 386–401.
- Nugent, W. R., Williams, M., & Umbreit, M. S. (2003). Participation in victim-offender mediation and the prevalence of subsequent delinquent behavior: A meta-analysis. *Utah Law Review*, 137, 137–166.
- Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2003). Scared straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review of the randomized experimental evidence. *Annals of the American Academy of Political and Social Science*, 589, 41–62.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. New York: Wiley.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Wilson, D. B., Mitchell, O., & MacKenzie, D. L. (2006). A systematic review of drug court effects on recidivism. *Journal of Experimental Criminology*, 2, 459–487.
- Wilson, S. J., Lipsey, M. W., & Soydan, H. (2003). Are mainstream programs for juvenile delinquency less effective with minority youth than majority youth? A meta-analysis of outcomes research. *Research on Social Work Practice*, 13, 3–26.