Participation in Victim-Offender Mediation and the Prevalence and Severity of Subsequent Delinquent Behavior: A Meta-Analysis

William R. Nugent

Mona Williams

Mark S. Umbreit

I. INTRODUCTION

Recently there has been the reemergence of an ancient philosophy of justice that has been called “restorative justice.”¹ In this philosophy, a crime is viewed as an offense against a victim, and the emphasis is on resolving conflict, repairing harm to the victim, holding the offender accountable to the victim, and returning things as much as possible to the way they were before the offense occurred.² Victim-offender mediation is the oldest and most widely practiced expression of restorative justice.³ Victim-offender mediation (VOM) programs have most commonly involved the victims and perpetrators of juvenile property offenses and minor assaults, though there have been increasing efforts to broaden the scope of these programs to include adult offenders and serious violent crimes.⁴ A recent survey found more than 300 VOM programs—also referred to as victim-offender reconciliation programs—in North America and more than 700 in Europe.⁵ In about 80% of VOM programs, the mediator (usually a trained community volunteer) meets initially with crime victims and offenders separately. Following these separate sessions, which help prepare the victims and offenders for

¹Children’s Mental Health Services Research Center, College of Social Work, University of Tennessee.
²College of Social Work, University of Tennessee.
³Center for Restorative Justice & Mediation, School of Social Work, University of Minnesota.
⁶UMBREIT, supra note 2, at 5–6.
subsequent dialogue, there is a mediation session, which is the heart of VOM.6 The goal of the mediation is to create an environment that allows the involved parties to engage in a dialogue in which emotional and informational needs are met and in which a plan for the offender “to make things right,” as much as possible, is developed. About 87% of VOM sessions result in a restitution agreement.7

Ideally, an approach to juvenile justice, regardless of the philosophy upon which it is based, would at minimum not be associated with an increase in either the prevalence or severity of subsequent delinquent behavior, and would at best be associated with a decrease in both.8 There is now a growing body of research on VOM, and a number of studies have investigated the relationship between juvenile participation in VOM and the prevalence and/or severity of recidivism.9 There is a need for a comprehensive synthesis of this research. Meta-analysis is a quantitative approach to aggregating and synthesizing the results from many research studies and determining what the results in toto tell us about the subject matter of the research. A meta-analysis can help us understand why results differ across studies, answering such questions as, “Why do the results of one study show that VOM participants reoffend 20% less than nonparticipants, while a second shows that VOM participants and nonparticipants reoffend at the same rate, while yet a third shows nonparticipants to reoffend less than VOM participants?” The results of a meta-analysis of the research on the relationship between VOM participation and subsequent delinquent behavior would have important implications for the development of juvenile justice programs and would inform policy makers considering restorative justice approaches to juvenile crime.

This Article reports the results of meta-analytic analyses10 of the research on VOM and the subsequent delinquent activity of juveniles. These analyses focused on three primary research questions pertaining to the prevalence and severity of delinquent behavior subsequent to VOM participation: (1) are the differences in reoffense between VOM and non-VOM groups the same across studies?; (2) if not, what factors seem to be responsible for the variation in between-group differences?; and (3) after controlling for factors that explain the significant variation in between-group differences in reoffense across studies, is there any

---

6Galaway, supra note 1, at 668–83.
7Umbreit & Greenwood, supra note 5, at 239.
evidence that VOM participation is related to a decrease in the prevalence or severity of subsequent delinquent behavior? A particularly important issue in these analyses concerned the possibility that between-group differences in reoffense were artifacts of biased VOM and comparison group formation, so a primary emphasis was the assessment and control of methodological differences across the included studies.

II. METHODOLOGY

A. Inclusion Criteria

To be included in this meta-analysis a study had to have: (1) focused exclusively on juveniles; (2) investigated the relationship between participation in VOM and the prevalence or severity of subsequent delinquent behavior; and (3) employed at least a VOM group and one comparison group of juveniles who had not participated in VOM. These inclusion criteria were used so that the meta-analysis could focus on what the extant research tells us about the differences in prevalence and severity of delinquent behavior between juveniles who participated in VOM and those who did not.

B. Selection of Studies

Following procedures described by Sowers, Ellis, and Meyer-Adams, an extensive literature search was conducted for studies of VOM that included a focus on recidivism. A search was conducted through the University of Tennessee’s library internet catalog, KUDZU catalogue, searches on Article First, ERIC, E-subscribe, First Search, General Reference Center Gold, JSTOR, Papers First, PsychInfo, Social Sciences Abstracts, Digital Dissertations & Theses, National Criminal Justice Reference Service, and searches on Proquest. The following key words and phrases were used in these searches: (a) victim-offender mediation; (b) victim-offender reconciliation program; (c) mediation; (d) family group conferencing; (e) restorative justice; and (f) various combinations of names of persons who have conducted research on restorative justice. The list of empirical studies on restorative justice compiled in the Research and Resources Review was also used. The reference lists and bibliographies included in articles, books, book chapters, and program evaluation reports were also searched for studies not found in any of the databases. Contacts were made with authors of

articles and reports on program evaluations of VOM in an effort to find unpublished studies. In addition, a search on the Internet was conducted (Google search engine) using the following key words and phrases: (a) evaluation of victim-offender mediation, and (b) assessment of victim-offender mediation.

Our search identified nineteen studies that investigated VOM and recidivism. Of these studies, fifteen met the selection criteria above. Our sample therefore included a total of fifteen studies of the prevalence of delinquent behavior subsequent to participation in VOM, conducted at nineteen different sites and focusing on nineteen different VOM programs, involving a total of 9307 juveniles. Unfortunately only three of these studies included a focus on the severity of subsequent delinquent behavior. These three studies had data on severity of reoffense for a total of 1003 juveniles. These fifteen studies included six that were published in peer-reviewed journals and books, one unpublished master’s thesis, and eight program evaluations. The studies included in the meta-analysis are listed in appendix A.

C. Outcome Measures

1. Prevalence

The prevalence effect size of focus was the “VOM effect,” defined as the difference in proportion (or percentage) of juveniles in VOM and non-VOM groups who reoffended [VOM proportion (percentage) minus non-VOM proportion (percentage)]. This difference could also be represented as the ratio of the odds of VOM participants reoffending to the odds of non-VOM participants reoffending. Analysis methods were also used that allowed us to focus on the between-group differences in rate of reoffense per unit of exposure to risk of recidivism, as described below. While not a major focus in this meta-analysis, the data analysis methods used also allowed us to explicitly represent the reoffense rates of non-VOM groups as well as the relationship between reoffense rates in non-VOM groups and that in VOM groups.

Reoffense has been defined as any offense for which a youth was adjudicated guilty during a one-year period. Similarly, it has been defined as any subsequent offense that was neither dismissed for lack of evidence nor for which the juvenile

---

was exonerated. In contrast, reoffense was defined in broader terms in the remaining eleven studies. In these studies reoffense was defined as any official contact with a law enforcement agency; as any subsequent court contact; or as any record of a rearrest. When the data were analyzed by site, there were nineteen prevalence effect sizes, and when analyzed by study, there were fifteen.

2. Severity

Umbreit compared juveniles’ reoffenses with their original offenses and created three classes of “severity of reoffense:” less serious reoffense, reoffense of same level of severity, and more serious reoffense. Though he did not include them, youths who failed to commit a reoffense could have been included in his “less serious reoffense” category since no reoffense is less serious than any original offense.

Other studies conceptualized a “harm to victim” dimension underlying the ordinal indicator of “severity of reoffense” used in those studies. They conceptualized four levels of “severity of reoffense.” The first level was “no reoffense” (level 0). The next level (level 1) represented new infractions in which there was no victim other than, perhaps, the offender him or herself, such as a curfew violation that was considered a violation of probation. The next higher level (level 2) represented new offenses in which a victim was harmed through her or his property, as in vandalism or theft. The most serious level (level 3) represented reoffenses in which a person was directly harmed, as in assault.

3. Combining Severity Measures

All original offenses in the Nugent and Paddock and Wiinamaki studies were property crimes (level 2), so the level of severity of the original offense was a constant in these two studies. Thus, in these studies a reoffense of severity level 0 or severity level 1 represented a decrease in the severity of juveniles’ subsequent delinquent behavior; a level 2 reoffense represented an infraction of the same severity as the original offense; and a reoffense of severity level 3 represented an increase in severity.

In contrast to the Nugent and Paddock and Wiinamaki studies, Umbreit’s study involved juveniles whose original offenses were either property (level 2) or

---

15UMBREIT, supra note 13, at 116.
16Nugent & Paddock, supra note 13, at 159; Wiinamaki, supra note 13, at 1.
17Nugent & Paddock, supra note 13, at 159.
18Wiinamaki, supra note 13, at 1.
personal (level 3) offenses. His “less serious re-offense” category, therefore, included: (a) youths who had committed, in Nugent and Paddock and Wiinamaki’s scheme, level 2 reoffenses subsequent to level 3 original offenses; (b) juveniles who had committed level 1 reoffenses subsequent to level 3 original offenses; and (c) youths who had committed level 1 reoffenses subsequent to level 2 offenses. Thus, the severity of reoffense classes used by Umbreit were more heterogeneous than the categories created by Nugent and Paddock and Wiinamaki. In order to facilitate a comparison of the relationships between VOM participation and severity of reoffense obtained in these three studies, three classifications of “severity of subsequent delinquent behavior” were created. The first was labeled “delinquent behavior decreased in severity” and included youths whose reoffense was of lesser severity (including those who did not commit a reoffense) than their original offense. This category included juveniles whose reoffenses ranged from levels 0 to 2 in Nugent and Paddock and Wiinamaki’s scheme. The second category was labeled “delinquent behavior remained at same level of severity” and included juveniles who committed a reoffense of the same level of severity as their original offense. This category included youths whose reoffenses were of severity levels 2 or 3 in Nugent and Paddock and Wiinamaki’s scheme. The third category was “delinquent behavior increased in severity” and included juveniles whose reoffenses were of greater severity than their original offenses, and also included youths whose reoffenses were of severity level 3 in Nugent and Paddock and Wiinamaki’s scheme. The heterogeneity of these categories in terms of the severity of the juveniles’ reoffenses in Nugent and Paddock and Wiinamaki’s scheme led to the conceptualization of this “severity of subsequent delinquent behavior” measure as a nominal variable.

D. Explanatory and Moderating Variables: Prevalence

One plausible and important explanation for differences in prevalence rates between VOM and non-VOM juveniles was the creation of groups that were initially nonequivalent in terms of potential for reoffense. Suppose, for example, that VOM and non-VOM groups were created in such a manner that VOM group members were first-time property offenders, while all non-VOM group members had long delinquent histories that included violent offenses. In this case, VOM...

---

19UMBREIT, supra note 13, at 29.  
20Id.  
21Id.  
22Nugent & Paddock, supra note 13, at 155–78; Wiinamaki, supra note 13, at 84.  
23Nugent & Paddock, supra note 13, at 155–78; Wiinamaki, supra note 13, at 84.  
24Nugent & Paddock, supra note 13, at 155–78; Wiinamaki, supra note 13, at 84.  
25Nugent & Paddock, supra note 13, at 155–78; Wiinamaki, supra note 13, at 84.
group members would be less likely to reoffend than non-VOM group members. Regardless of how the two groups were treated after formation, there would be an expected between-group difference in reoffense due entirely to the initial nonequivalence of the groups. An important goal in the meta-analyses was to control, as much as possible, for methodological differences across studies in the manner in which VOM and non-VOM groups had been formed. This would enable the meta-analysis, at least to some extent, to disentangle between-group differences in reoffense rates that were artifacts of biased group formation from group differences due to participation in VOM.

To this end, twelve dichotomously scored items were created, each of which indicated the presence or absence of a methodological feature that would increase the likelihood of the VOM and non-VOM groups being initially equivalent, and each study (and the evaluation conducted at each site in studies with multiple sites) was rated using these twelve items. Total scores on this “methodological quality scale” could range from 0 to 12, with higher scores indicative of studies more likely to have created initially equivalent groups and lower scores indicative of studies that were more likely to have created initially nonequivalent groups. A score of 12 would indicate a randomized experiment, one of the more powerful means for creating initially equivalent groups.

Two raters independently rated each of the studies (and the evaluations conducted at each of the study sites) using the methodological quality scale. In the language of generalizability theory, this was a study-by-rater-by-item G-study, or, for the sites, a study-site-by-rater-by-item G-study. The mean interrater agreement between the two sets of ratings was 91.7% (range 75% to 100%). Generalizability theory (or g-theory) methods were also used to estimate an intraclass correlation for generalizing from the means of the two raters’ averaged item ratings to the same means that would be obtained by having all raters in a universe of raters rate each study using all items in a universe of methodological quality items. This intraclass correlation, called a “generalizability coefficient,” was .92. Consistent with g-theory procedures, in the HGLM analyses discussed below, the mean scores across the averaged (across the two raters) item scores for each study were used as the methodological quality scores. These scores could (and did) range from 0 to 1.

Eight other variables were used as possible factors explaining significant variation in VOM effects: length of time juveniles were tracked for reoffense; type of sample employed; differences between the non-VOM and VOM groups in terms of percentage of juveniles who had committed violent offenses (such as assault); a dummy variable indicating whether or not a study/site was missing data on differences between the non-VOM and VOM groups in terms of percentage of violent offenders; the size of the county served by the VOM program; the definition of reoffense used; whether a study included only juveniles who had been referred to VOM; and the publication status of the study. Tracking periods in the included studies ranged from six to thirty-six months. Longer tracking
periods would be expected to be associated with greater reoffense rates, and possibly with smaller VOM effects if any VOM effect attenuates with time.  

The “type of sample” was a dichotomous variable indicating whether the sample of juveniles included only property offenders (“type of sample” = 0) or included both property and violent offenders (“type of sample” = 1). Studies in which one of the groups had a significantly greater proportion of juveniles originally charged with violent offenses might have biased group differences in reoffense since a juvenile who commits a violent offense may be at a greater risk of subsequent delinquent behavior. Efforts were made to reconstruct the between-group differences in percentage of violent offenders. This variable (symbolized by δ in the remainder of this Article) was defined as the percentage difference between non-VOM and VOM groups in terms of violent offenses; that is, non-VOM group’s percentage of violent offenders minus VOM group’s percentage of violent offenders. There was not enough information to reconstruct δ for two of the studies (a total of 13% of the studies or 10.5% of the study sites).

Expectation maximization (EM) procedures were used to impute these missing values from item scores on the methodological quality scale, and an indicator variable (0 = study/site had no missing value, 1 = study/site had missing value) was used to represent missing data on this variable. This missing data variable was used to determine whether there was a systematic relationship between missing data on this variable and study outcomes. Analyses gave no evidence of any relationship between missing data on δ and outcomes.

The population size of the county served by a VOM program might be associated with the magnitude of VOM effects. Juveniles in counties with larger populations may be at greater risk of delinquent behavior because of access to peer groups and exposure to other environmental factors, associated with larger population areas, that are related to delinquent activity. A dichotomous variable was used to identify the population size of the county served by the VOM program (0 = population less than 100,000; 1 = population greater than 100,000).

The definition of reoffense used in a study could also impact the reoffense rates observed and possibly the magnitude of VOM effects. The broader definitions used in eleven of the studies may have led to greater observed reoffense rates than the narrow definition used in the Nugent and Paddock, Winamaki, Umbreit, and Schneider studies by identifying juveniles as “reoffenders” who were ultimately cleared of the offense they had been suspected

---


of committing. A dichotomous variable was used to identify the definition of reoffense used in a particular study (0 = narrow definition; 1 = broader definition).

The samples used in a few of the studies contained only VOM-referred juveniles. In these studies, all involved juveniles had been selected for VOM participation by juvenile court and/or VOM program staff, and the non-VOM group was comprised of those VOM-referred juveniles who, although referred, did not participate for reasons such as parent saying “no.” A sample such as this would potentially be biased towards lower reoffense rates (and possibly smaller VOM effects) since the selection process used by juvenile court or VOM staff could lead to referrals only of juveniles less likely to reoffend (for example, juveniles without a prior offense history). A dichotomous variable entitled “all VOM referrals” was created and scored 0 for studies in which the sample of juveniles was not all VOM referrals and 1 for those in which the sample was all VOM referrals. The publication status of a study was also used as a potential explanatory variable. Researchers have observed that the magnitude of effect sizes in studies published in peer-reviewed journals tend to be different than those found in studies that are either unpublished (such as dissertations) or that are published in non-peer-reviewed sources (such as books). A dichotomous variable (0 = not published in peer-reviewed source; 1 = published in peer-reviewed source) was used to represent publication status.

E. Explanatory and Moderating Variables: Severity

The very small number of studies in which severity of reoffense was investigated made it impossible to investigate the relationship between specific explanatory variables and severity of subsequent delinquent behavior. We therefore only investigated the link between VOM participation and severity of subsequent delinquent activity and tested the plausibility that the results were homogeneous across studies. The homogeneity of results would be consistent with successful replication of the results across the included studies. Participation in VOM was a dichotomous variable (nonparticipation = 0, participation = 1).

F. Data Analysis Methods

1. Prevalence

Multiple analytic methods were employed, each with a somewhat different conceptualization of, and based on somewhat different assumptions about, the data. This would enable us to look for consistency, or the lack thereof, across the results from these different procedures. Consistency of results would suggest that the results were insensitive to the analysis method employed and reduce the likelihood that the results were artifacts of an interaction between data and data
Therefore, graphical and binomial, Bernoulli, and Poisson hierarchical generalized linear model (HGLM) techniques were used to analyze the prevalence results across sites and studies. The binomial and Bernoulli models are reminiscent of logistic regression models in which the probability of an event occurring, such as reoffense, can be modeled as a function of explanatory variables such as VOM participation and how VOM and non-VOM groups were formed.

In contrast, the outcome in a Poisson model is the number of events of a certain kind (such as reoffenses) that occur per unit “exposure” to a certain risk. Suppose that there are two groups of juveniles, one in which there are fifty youths and a second with 100 juveniles. Suppose further that the first group is tracked for reoffense for six months and that seventeen youths from this group reoffend, while the juveniles in the second group are observed for twelve months and fifty-one of them reoffend during this period. If only the numbers of reoffenders are considered, without taking into account the different numbers of youths in the two groups or the different periods of time the juveniles are observed for reoffense, then it appears that the second group has reoffended at an incidence three times as great as that of the first group: fifty-one reoffenders versus seventeen. The size of the group and the duration that they are observed for reoffense are both indicators of the groups’ “exposure” to risk of reoffense. The size of the group represents one exposure variable, specifically the number of juveniles in the environment that are presented with the opportunities to reoffend that are endemic to their environment. The duration they are observed for possible reoffense represents a second exposure dimension—the amount of time each juvenile has been exposed to the environmental opportunities for engaging in new delinquent behavior.

These two dimensions of exposure to reoffense risk can be combined by multiplication into a variable with the metric “juvenile-month,” a unit indicating one juvenile exposed to the risk of reoffense for one month. Then the numbers of youths reoffending in a particular group can be divided by this product, creating a rate of reoffense per unit exposure that takes into account the exposure variables. The rate variable for the group with seventeen reoffenders would be: \( \frac{17 \text{ reoffenders}}{50 \text{ juveniles} \times 6 \text{ months}} = 0.057 \text{ reoffenders per juvenile month} \), while for the second group, it would be: \( \frac{51}{100} \times 12 = 0.043 \text{ reoffenders per juvenile month} \). It thus becomes apparent that the second group is actually reoffending at a rate only about three-fourths as great as the first per unit exposure. This is because the second group has had a much greater exposure to


\[39\] Bryk et al., supra note 10, at 124–27.

\[40\] Id.
reoffense than the first. This approach therefore allows explicit consideration of the differences between the groups in numbers of youths and tracking periods of different lengths. The rate of reoffense per unit exposure can then be modeled as a function of explanatory variables, such as participation in VOM and how VOM and non-VOM groups were formed.

2. Severity

Multinomial logistic regression (MLR) procedures allow the researcher to determine the extent to which the probability of falling into one of several distinct categories, such as “delinquent behavior remained at same level of severity,” is associated with any of a number of predictor variables, such as participation in VOM. MLR procedures were used to test the extent to which the findings on the relationship between VOM participation and the severity of juveniles’ subsequent delinquent behavior was homogeneous across the Umbreit, Nugent and Paddock, and Wiinamaki studies. The data from the Umbreit, Nugent and Paddock, and Wiinamaki studies on VOM participation and severity of subsequent delinquent behavior were combined and a set of two dummy variables were created to represent the distinctions between these three studies. A test of statistical significance of the set of dummy variables would assess the equivalence of the three studies in terms of the combined VOM and non-VOM groups’ reoffense rates in each severity category. An interaction set was also created that was the product of the dichotomous VOM participation variable and the dummy variable set. A test of statistical significance of this interaction set would assess the homogeneity of the relationship between VOM participation and severity of subsequent delinquent behavior across the three studies, controlling for any differences between studies in the combined VOM and non-VOM groups’ reoffense rates in each severity category. Statistically nonsignificant results in this latter test would be consistent with homogeneity of the relationship between VOM participation and severity of subsequent delinquent behavior across the studies and with the successful replication of results.

III. PREVALENCE RESULTS

A. Sample Sizes and Sample Characteristics

The nineteen sites (and fifteen studies) gave a combined sample of 9307 juveniles with data on prevalence of reoffense. The characteristics of the samples

---

32Id.
The VOM effects, represented as VOM percentage of reoffenders minus non-VOM percentage, and approximate 95% confidence intervals for these effects, for each of the included studies (fifteen effect sizes, top figure) and study sites (nineteen effect sizes, bottom figure) are shown in Figure 1. Also shown in this figure is a point estimate and confidence interval (far right) for the results if all studies are pooled. As shown in this figure, eleven of the fifteen studies (fifteen of nineteen sites) show results in which the VOM effect is negative, indicating that VOM participants reoffended at a lower rate than nonparticipants. The obtained differences were statistically significant in seven of these studies (four sites). Non-VOM groups had lower reoffense rates (positive VOM effects) in four of the studies (four sites), and these differences were statistically significant in only one of these studies (one site).

**Figure 1.** Differences between VOM and non-VOM reoffense rates for fourteen studies in meta-analysis (top graphic; differences shown are VOM percentage reoffending minus non-VOM percentage reoffending). Vertical bars are approximate 95% confidence intervals for between groups’ differences. Circle at far right is estimated between group difference (and approximate 95% confidence interval) for all studies combined. Bottom graphic shows same information for differences between VOM and non-VOM reoffense rates by eighteen sites.
C. VOM Effects as a Function of Definition of Reoffense

Figure 2 shows plots of these VOM effects versus definition of reoffense. The top figure shows these results by study, and the lower figure by site. There is a clear relationship between variation in VOM effect sizes and definition of reoffense, with greater dispersion among studies using the broader definition.

**Figure 2.** Estimated VOM effects (differences shown are VOM percentage reoffending minus non-VOM percentage reoffending) as a function of definition of reoffense. Top graphic shows this information by study, while bottom graphic shows this breakdown by site.
There were significant methodological differences between studies using the narrow and broad definitions. The mean score on the methodological quality scale for sites using the narrow definition was .56 ($SD = .22$; range $.38 - 1.0$), and .19 ($SD = .19$; range $0 - .5$) for sites using the broader definition, a statistically significant difference ($t(17) = 4.0, p = .001$). Further, all studies (and sites) that used the narrow definition of reoffense had non-VOM and VOM groups that were matched in terms of percentages of violent offenders, so $\delta = 0$. While the mean $\delta$ across studies that used the broader definition was quite small (-1.1%) the variability was large ($SD = 15.3$; range -42% to +20.3%). These results suggested the possibility that the definition of reoffense was an important variable related to outcomes. Thus, individual HGLM analyses were conducted on the data from the sites using the narrow and broad definitions of reoffense, and the results from these analyses were used to create the HGLM model for the combined data.

**D. HGLM Results for Sites using Narrow Definition of Reoffense**

The results of fitting an unconditional binomial model (i.e., no predictor variables were in the HGLM model) produced an estimate of 34.3% ($t(7) = -3.04, p = .02$) for the mean reoffense rate across non-VOM groups and suggested that there was significant variability across sites in the reoffense rates of non-VOM participants (estimated parameter variance = .30; $\chi^2(7) = 33.9, p < .001$). The results also produced an estimate of 23.7% ($t(7) = -3.79, p = .008$) for the mean reoffense rate across VOM groups, indicating a mean VOM effect of -10.6%. The results very strongly suggested that the VOM effect was homogeneous across these sites (estimated parameter variance = .0036, $\chi^2(7) = 1.12, p > .95$), implying that for the sites using the narrow definition of reoffense all the differences in magnitude of VOM effects was due to sampling variability, thus
requiring no predictors of VOM effects. The results of exploratory visual and HGLM analyses suggested that only the score on the methodological quality scale was a significant predictor of non-VOM groups’ reoffense rates for these sites.

E. HGLM Results for Sites using Broad Definition of Reoffense

The results of fitting an unconditional binomial model produced an estimate of 32.3% ($t(10) = -5.46, p < .001$) for the mean non-VOM group reoffense rate and suggested that there was significant variability in the reoffense rates of non-VOM participants (estimated parameter variance = .155; $\chi^2(10) = 237.6, p < .001$). The results also produced an estimate of -7% ($t(10) = -1.21, p = .25$) for the mean VOM effect. Since this estimated mean VOM effect was statistically nonsignificant, however, it suggested that the mean VOM effect across these sites was not statistically different from zero. The results also suggested that the VOM effect was not homogeneous across these sites (estimated parameter variance = .76, $\chi^2(10) = 92.5, p < .001$). The results of exploratory visual and HGLM analyses suggested that δ, the type of sample used, and the publication status of the results from the site were likely significant predictors of non-VOM groups’ reoffense rates. Furthermore, the results suggested that δ and the score on the methodological quality scale were potentially significant predictors of the magnitude of VOM effects, and other results suggested that there were no significant predictors remaining among the left out potential explanatory variables.

F. HGLM Results for Full Data by Sites

The results of fitting an unconditional HGLM binomial model to the combined data from all sites suggested that the parameter variance in non-VOM groups’ reoffense rates was .18 ($\chi^2(18) = 245.6, p < .001$), clearly suggesting significant systematic variation in reoffense rates for the non-VOM groups. The estimated VOM effect parameter variance was .40 ($\chi^2(18) = 97.5, p < .001$), results clearly suggesting that the magnitude of the VOM effects was heterogeneous. The estimated mean non-VOM group reoffense rate parameter was -.71 ($t(18) = -6.4, p < .001$), indicating a mean reoffense rate across non-VOM groups of about 33%. The estimated mean VOM effect parameter was -.39 (odds ratio = .68; $t(18) = -2.35, p = .03$), indicating a mean VOM group reoffense rate of 25%, and, therefore, a mean VOM effect of -8% and a mean reduction in reoffense rates associated with VOM participation of about 24% relative to non-VOM groups’ rates.

---

33BRYK ET AL., supra note 10, at 124–27.
The analyses of data by definition of reoffense had suggested that the component of the HGLM model for non-VOM groups’ reoffense rates should include $\delta$, type of sample used (both implied by results from broad definition), and score on the methodological quality scale (implied by results from narrow definition). The analyses by definition had also suggested that variation in the magnitude of VOM effects for data from sites where the narrow definition was used was random and, therefore, that there was no relationship between any of the explanatory variables and VOM effect sizes. In contrast, the analyses of the data from the sites where a broad definition was used suggested that VOM effect sizes were moderated by the score on the methodological quality scale and by $\delta$. These results suggested that for all data combined, the portion of the HGLM model that explains the magnitude of VOM effects should reduce to one with no predictors when using a narrow definition, and that sites using the broader definition should include both the methodological quality score and $\delta$. These considerations suggested, for explaining the magnitude of VOM effects, an interaction model of the form:

$$VOM\, effect = VOM\, effect\, parameter + (definition\, of\, reoffense \times score\, on\, methodological\, quality\, scale) + (definition\, of\, reoffense \times \delta) + random\, error\, (equation\, 1)$$

where definition of reoffense is the dichotomous variable indicating the definition of reoffense used at a particular site. This model would reduce to,

$$VOM\, effect = VOM\, effect\, parameter + random\, error\, (equation\, 2)\, for\, sites\, using\, the\, narrow\, definition.$$  

The omnibus multivariate test the HGLM model discussed by the HGLM model immediately above was statistically significant ($\chi^2(5) = 96.3, p < .000001$), and an omnibus test of all potential explanatory variables left out of this model produced nonsignificant results ($\chi^2(14) = 19.8, p > .10$). The final results for this model accounted for the majority of the variation in non-VOM groups’ reoffense rates (87%; estimated residual parameter variance = .023, $\chi^2(15) < 45.4, p = .001$), and for all of the significant variation in VOM effects (94.5%; estimated residual parameter variance = .022, $\chi^2(16) = 21.0, p = .18$). These results suggest that the residual parameter variance for VOM effects could be set to zero.

The results for this model with residual VOM effect parameter variance set to zero are shown in the left side of Table 1. The results of this analysis suggested that the odds of VOM participants reoffending were slightly more than half (.57)
as great as were the odds of nonparticipants reoffending when the narrow
definition of reoffense is used.

Table 1. Results of HGLM Binomial Model Analysis of Prevalence Data from
All Sites Combined (left side, normal font), and of a Three-Level Bernoulli Model
(right side, italicized). All results are Population Average Results.

Non-VOM groups’ reoffense rates

<table>
<thead>
<tr>
<th>variable</th>
<th>B</th>
<th>t(15)</th>
<th>p</th>
<th>B</th>
<th>t(11)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept level 2</td>
<td>-1.34</td>
<td>-8.1</td>
<td>&lt; .001</td>
<td>-1.43</td>
<td>-10.2</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>δ</td>
<td>.018</td>
<td>3.12</td>
<td>.008</td>
<td>.019</td>
<td>5.04</td>
<td>.008</td>
</tr>
<tr>
<td>methodological score</td>
<td>1.09</td>
<td>4.09</td>
<td>.001</td>
<td>1.07</td>
<td>5.23</td>
<td>.001</td>
</tr>
<tr>
<td>sample</td>
<td>.37</td>
<td>2.63</td>
<td>.019</td>
<td>.48</td>
<td>4.07</td>
<td>.019</td>
</tr>
</tbody>
</table>

VOM Effects

<table>
<thead>
<tr>
<th>variable</th>
<th>B</th>
<th>t(31)</th>
<th>p</th>
<th>B</th>
<th>t(12)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept level 2</td>
<td>-5.6</td>
<td>-6.33</td>
<td>&lt; .001</td>
<td>-5.4</td>
<td>-5.29</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>definition of reoffense X</td>
<td>-0.39</td>
<td>-7.59</td>
<td>&lt; .001</td>
<td>-0.4</td>
<td>-7.56</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>δ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>methodological score</td>
<td>1.0</td>
<td>3.82</td>
<td>.001</td>
<td>1.12</td>
<td>3.24</td>
<td>.008</td>
</tr>
</tbody>
</table>

The results also suggested that as the methodological quality of a study
approached that of a randomized experiment (assuming a narrow definition of
reoffense and a mixed sample of property and violent offenders), the expected
results would show a reduction in reoffense for VOM participants of about 26%
relative to nonparticipants’ reoffense rates.

These results further suggested that, when the broader definition of reoffense
is used, the magnitude of VOM effects depends upon δ and on the score on the
methodological quality scale. This relationship is graphically illustrated in the
upper graph in Figure 3. This figure was drawn so as to, as much as possible,
represent regions covered by joint observations, thereby helping to prevent hidden
extrapolations.35

35Id.
Figure 3. Graphic illustration of results of HGLM binomial model for VOM effects (top graph) and plot of residuals from predicting VOM effects from all methodological variables (bottom graph).

As can be seen in this graph, as the percentage of violent offenders in the VOM group becomes increasingly greater than the percentage in the non-VOM group (i.e., δ in Figure 3 changes from a positive to a negative number and then becomes a larger and larger negative number), the VOM effect changes from one showing VOM participants to have lower reoffense rates than nonparticipants to one showing nonparticipants to have lower reoffense rates than VOM participants.
Vice versa, as the percentage of violent offenders in the non-VOM group becomes increasingly greater than the percentage in the VOM group (i.e., $\delta$ in the figure becomes a larger and larger positive number), the VOM effect shows VOM participants to be reoffending at decreasing rates relative to nonparticipants. Finally, as the score on the methodological quality scale increases, given a particular value of $\delta$, the magnitude of the odds ratio increases. It is important to note in this figure that the maximum methodological quality score for sites using the broad definition was only .5. The dashed line shows the apparent constant VOM effect across the observed values of scores on the methodological scale for narrow definition sites.

G. Results for Alternate Analysis Approaches

Meta-analysts have noted\textsuperscript{36} that meta-analyses that use multiple within-study outcomes as independent effect sizes, such as was done in the above analyses by site, can be adversely affected by complex interdependencies among within-study effect sizes. In order to test whether interdependencies between site outcomes within studies could have adversely affected the above findings, a three level HGLM analysis, using a Bernoulli model, of the full data by study, as opposed to site, was conducted. The results, shown in the right side of Table 1 in italics, were essentially the same as those from the Binomial analysis conducted by site. These findings suggested that the results were not adversely affected by interdependencies among effect sizes within studies.

The data were also analyzed using a Poisson model in which the data were conceptualized as count data indicating the number of juveniles in a group of a specific size who reoffended during a specific period of time in which they were exposed to the risk of reoffending. The exposure variable in this analysis, as illustrated in earlier discussion on Poisson models, was the number of juveniles in a specific non-VOM or VOM group multiplied by the duration of the follow-up period for that group, creating an exposure variable with the metric of “juvenile-months.” The same explanatory variables included in the model whose results are shown in Table 2 were used in the Poisson model in an effort to replicate the findings from the binomial and Bernoulli models. Successful replication would suggest that the overall findings pertaining to the magnitude of VOM effects were relatively insensitive to the data analysis method used.

The results of fitting this Poisson model are shown in the left side of Table 2. The results for the magnitude of VOM effects are consistent with those in Table 1. These results suggested that all of the significant parameter variation in the VOM effects, in this case differences in reoffense rates per unit exposure between VOM and non-VOM groups, is explained by the two interaction variables in the

\textsuperscript{36}GLASS ET AL., supra note 10, at 218.
VOM effect portion of the model (estimated residual parameter variance = .02, \( \hat{\chi}^2(15) = 17.7, p > .25 \)).

**Table 2.** Results of HGLM Analysis of Prevalence Data using a Poisson Model (Population Average Results). Results controlling for definition of reoffense are shown in the left side of the table, and those ignoring the definition of reoffense are shown in the right side (*in italics*).

### Non-VOM groups’ reoffense rates

<table>
<thead>
<tr>
<th>variable</th>
<th>B</th>
<th>t(15)</th>
<th>p</th>
<th>B</th>
<th>t(15)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ</td>
<td>.0003</td>
<td>.031</td>
<td>&gt;.50</td>
<td>.0006</td>
<td>.06</td>
<td>&gt;.50</td>
</tr>
<tr>
<td>methodological score</td>
<td>.009</td>
<td>.02</td>
<td>&gt;.50</td>
<td>-.056</td>
<td>-.12</td>
<td>&gt;.50</td>
</tr>
<tr>
<td>sample</td>
<td>-.076</td>
<td>-.32</td>
<td>&gt;.50</td>
<td>-.13</td>
<td>-.52</td>
<td>&gt;.50</td>
</tr>
</tbody>
</table>

### VOM Effects

<table>
<thead>
<tr>
<th>variable</th>
<th>B</th>
<th>t(16)</th>
<th>p</th>
<th>B</th>
<th>t(16)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept level 2</td>
<td>-.40</td>
<td>-5.31</td>
<td>&lt;.001</td>
<td>-.508</td>
<td>-4.03</td>
<td>.001</td>
</tr>
<tr>
<td>definition of reoffense X</td>
<td>-.028</td>
<td>-5.60</td>
<td>&lt;.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>δ</td>
<td>.75</td>
<td>3.87</td>
<td>.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>duration of follow-up</td>
<td>.014</td>
<td>2.91</td>
<td>.011</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These results suggested that the reoffense rate per juvenile-month for VOM participants was only about two-thirds (67%) as great as the rate for nonparticipants. Stated another way, the results suggested that the expected rate of reoffense per unit exposure would be 33% lower for VOM participants relative to nonparticipants. The results further suggested that the explanatory variables for predicting the non-VOM groups’ reoffense rates only accounted for about 14% of the significant parameter variation and that the Poisson model implied different predictors of non-VOM groups’ reoffense rates than did the binomial model or the three-level HGLM Bernoulli model.

Overall, these findings suggested that the results of this meta-analysis as they pertain to the magnitude of VOM effects may be relatively insensitive to the type of data analysis used. In all analyses, three methodological variables—δ, the definition of reoffense, and the score on the methodological quality scale combined in an interaction regression model—accounted for all of the variation in magnitude of VOM effects. However, the results regarding the non-VOM
groups’ reoffense rates appeared to be somewhat sensitive to the data analysis method used.

H. Results Ignoring Definition of Reoffense

In the analyses above, a major emphasis was placed on the differences that appeared to be associated with the two alternate definitions of reoffense. It may be argued that there is little substantive difference between the two definitions, and that the differences in dispersion of VOM effects associated with the definitions shown in Figure 2 are actually a function of the relatively small sample size and of the fact that studies in which the narrow definition was used were of significantly higher methodological quality than the studies in which the broader definition was used. Analyses were conducted to assess the plausibility of this argument. If the differences in dispersion were due entirely to methodological factors, then the differences in dispersion should disappear once methodological factors were controlled.

The graph in the lower half of Figure 3 shows the residuals from predicting VOM effect sizes (binomial model) for the combined data from all methodological variables. As can be seen in this graph, there is still considerably greater variation in residual effect sizes for broad definition studies than there is for narrow definition studies. Using HGLM methods, analyses of the data from the studies using the broader definition suggested that, after controlling for all variables related to methodological quality, significant residual variation remained in magnitude of VOM effects \( \text{estimated residual parameter variance} = .14, \chi^2(5) = 27.6, p < .001 \). In contrast, analyses of the data from narrow definition sites suggested that, after controlling for all variables related to methodological quality, the residual parameter variation in magnitude of VOM effects was random \( \text{estimated residual parameter variance} = .0004, \chi^2(4) = 0.99, p > .50 \). These results were inconsistent with the argument that there is no meaningful difference between the narrow and broad definitions of reoffense.

However, the analyses were redone ignoring the difference in definitions of reoffense so as to assess how different the results of the meta-analysis might be if the assumption were made that the two different definitions of reoffense were, for all practical purposes, the same. The results of these reanalyses are shown in Table 3 (binomial and Bernoulli models) and in the right side of Table 2 (Poisson model). As can be seen in Table 3, the predictors of non-VOM groups’ reoffense rates remained the same, while \( \delta \) and duration of follow-up were significant predictors of magnitude of VOM effects.
Table 3. Results of HGLM Binomial Model Analysis, ignoring definition of reoffense, of Prevalence Data from All Sites Combined (left side, normal font), and of a Three-Level Bernoulli Model (right side, italicized). All results are Population Average Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>t(15)</th>
<th>p</th>
<th>B</th>
<th>t(11)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-VOM groups’ reoffense rates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>intercept level 2</td>
<td>-1.3</td>
<td>-7.68</td>
<td>&lt;.001</td>
<td>-1.4</td>
<td>-9.92</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>δ</td>
<td>.017</td>
<td>2.89</td>
<td>.012</td>
<td>5.07</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>methodological score</td>
<td>1.03</td>
<td>3.81</td>
<td>.002</td>
<td>5.01</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>sample</td>
<td>.33</td>
<td>2.26</td>
<td>.039</td>
<td>3.69</td>
<td>.004</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VOM Effects</th>
<th>B</th>
<th>t(31)</th>
<th>p</th>
<th>B</th>
<th>t(12)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept level 2</td>
<td>-.66</td>
<td>-5.11</td>
<td>&lt;.001</td>
<td>-.65</td>
<td>-3.80</td>
<td>.003</td>
</tr>
<tr>
<td>δ</td>
<td>-.039</td>
<td>-7.78</td>
<td>&lt;.001</td>
<td>-.04</td>
<td>-7.45</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>duration of follow-up</td>
<td>.016</td>
<td>3.06</td>
<td>.005</td>
<td>.018</td>
<td>2.26</td>
<td>.043</td>
</tr>
</tbody>
</table>

In both the binomial and Bernoulli models, δ and the duration of follow-up accounted for all significant variation in magnitude of VOM effects. The results for the binomial model are shown in the upper graph in Figure 4.

Figure 4. Graphic illustration of HGLM binomial model for VOM effects ignoring definition of reoffense (top graph) and of HGLM Poisson model for VOM effects ignoring definition of reoffense (bottom graph).
As can be seen in this figure, the magnitude of the VOM effect diminishes (assuming $\delta = 0\%$, a mixed sample, and that the methodological quality approaches that of a randomized experiment) from an odds ratio of about .57 at six months follow-up, indicating that VOM participants are nearly 50% less likely to reoffend than nonparticipants six months after court involvement, to an odds-ratio of about .83 at thirty months, indicating that at thirty months postcourt involvement, VOM participants are about 20% less likely to reoffend than nonparticipants. The results from the Bernoulli model applied to the data by study as opposed to by site, shown in italics in the right side of Table 3, produced results that were essentially the same.

The results in the right side of Table 2 from the Poisson model are shown in the graph in the lower half of Figure 4. As can be seen in Figure 4, these results were similar in form to those from the binomial and Bernoulli models and suggested that the magnitude of VOM effects attenuates across time and that $\delta$ and the duration of follow-up accounted for all of the significant variation in magnitude of VOM effects. These results suggested that, assuming that $\delta = 0\%$, a mixed sample, and that the methodological quality of a study approached that of a randomized experiment, the magnitude of the VOM effect diminishes from one showing that VOM participants reoffended at a rate per juvenile-month that is 33% less than that of nonparticipants at six months follow-up to one showing VOM participants reoffending at a rate per juvenile-month at about 10% lower than nonparticipants at thirty months.

IV. SEVERITY RESULTS

The results of the multinomial logistic regression analysis of the severity of subsequent delinquent behavior data from the Umbreit, Nugent and Paddock, and
Wiinamaki studies were consistent with the homogeneity of the relationship between VOM participation and severity of subsequent delinquent behavior across these studies. The test of homogeneity of the relationship between VOM participation and severity of subsequent delinquent behavior was statistically nonsignificant ($\chi^2(4) = 5.8, p = .21$). The relationship between VOM participation and severity of reoffense (controlling for any differences between studies in combined non-VOM and VOM severity rates within categories) was statistically significant ($\chi^2(2) = 10.8, p = .001$). The severity of subsequent delinquent behavior rates for juveniles are shown in Table 4 for each of the three categories of severity of subsequent delinquent behavior. Also shown in Table 4 are approximate 99% confidence intervals for the differences in delinquent behavior rates between VOM and non-VOM groups in each severity category. If VOM participation reduced the severity of subsequent delinquent behavior, then there should be: (a) a significantly higher percentage of VOM than non-VOM juveniles in the “delinquent behavior decreased in severity” category; (b) a significantly lower percentage of VOM than non-VOM juveniles in the “delinquent behavior remained at same level of severity” category; and (c) a significantly lower percentage of VOM than non-VOM juveniles in the “delinquent behavior increased in severity” category.

Table 4. Numbers and percentages of non-VOM and VOM juveniles in each of the three categories of severity of subsequent delinquent behavior for the three studies which investigated this outcome, and approximate 99% confidence interval for the difference VOM percentage minus non-VOM percentage.

<table>
<thead>
<tr>
<th></th>
<th>Non-VOM</th>
<th>VOM</th>
<th>Approximate 95% C.I. (directional) for difference (VOM percentage minus non-VOM percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delinquent behavior decreased</td>
<td>427 (82.1%)</td>
<td>428 (88.6%)</td>
<td>2.5% to 10.5%</td>
</tr>
<tr>
<td>in severity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delinquent behavior remained</td>
<td>66 (12.7%)</td>
<td>32 (6.6%)</td>
<td>-16.1% to 3.9%</td>
</tr>
<tr>
<td>at same level of severity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delinquent behavior increased</td>
<td>27 (5.2%)</td>
<td>23 (4.8%)</td>
<td>-10.8% to 10.0%</td>
</tr>
<tr>
<td>in severity</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As can be seen in Table 4, the difference between VOM and non-VOM groups in reoffense rates for the “delinquent behavior decreased in severity” category was in the expected direction and statistically significant. The difference
between VOM and non-VOM groups in reoffense rates in the “delinquent behavior remained at same level of severity” category was in the expected direction, but statistically nonsignificant. The difference between VOM and non-VOM groups in reoffense rates in the “delinquent behavior increased in severity” category was opposite the expected direction and statistically nonsignificant. The small numbers of youths in these latter two categories made these low-power tests. These results are shown graphically in Figure 5. In this figure, the squares are point estimates of the differences between VOM and non-VOM groups in percentages of reoffenders in each severity category, and the vertical bars are approximate 95% confidence intervals for the differences.

**Figure 5.** Graphic illustration of results for severity of reoffense. Squares are point estimates of the difference “VOM percentage of juveniles in severity category minus non-VOM percentage,” and vertical bars are approximate 95% confidence intervals for this difference.

---

**V. DISCUSSION AND CONCLUSION**

The results of this meta-analysis strongly suggested that the definition of reoffense used in each study was a critically important methodological factor when looking at prevalence outcomes. The narrow definition of reoffense—a subsequent charge for which the juvenile was adjudicated guilty—may be a conservative indicator of a juvenile’s involvement in new delinquent activity. It would appear to be conservative in that the adjudication of guilt implies that the youth acknowledged responsibility for the act and/or that there was enough evidence indicative of the juvenile’s involvement to justify a determination of
guilt. Thus, use of the narrow definition may decrease the number of false positives.

This is in contrast to the broader definitions of reoffense used in many of the included studies. The fact of a new contact with the police or of a new arrest does not necessarily indicate that a new delinquent offense has occurred. Juveniles (minorities in particular), especially those who have a known history of delinquency, may come to the attention of law enforcement for a variety of reasons, and may be singled out as a suspect in a crime and yet not have committed the delinquent act of which they are suspected. The fact of a new law enforcement contact, even a new arrest, does not imply responsibility for an offense with the same probative force that an adjudication of guilt does. However, the broad definition may correctly identify some juveniles as reoffenders who would be missed by the narrow definition since there will be youths who are guilty of an offense for which they are arrested but never adjudicated guilty. The broad definition of reoffense may therefore decrease to some extent the number of false negatives, but at the cost of increasing the number of false positives. All things considered, the narrow definition may be a more valid archival indicator of delinquent behavior than the broader definition.

If it is correct that the narrow definition is a more valid indicator of delinquent behavior, then the results of this meta-analysis, given that the studies using the narrow definition had the highest methodological quality and that all significant variation in magnitude of VOM effects appeared to be explained by methodological factors, are consistent with the intriguing possibility that VOM participation may cause a decrease in delinquent behavior. In terms of effect size, the reduction in reoffense may be as great as 26% relative to non-VOM participants. This estimated effect size is nearly three times as large as the mean reduction in reoffense associated with interventions for general delinquents found in a recent meta-analysis\(^\text{37}\) and about twice as large as the mean reduction associated with programs for serious delinquents reported in a second meta-analysis.\(^\text{38}\) However, only future experimental research on the effects of VOM participation on reoffense can clarify the question of causality.

Both of the definitions of reoffense used in the included studies are indicators of specific phenomena in their own right. The narrow definition is indicative of delinquent activity that leads juveniles to be charged and adjudicated guilty. The broader definition may be indicative of delinquent activity that leads to contact with law enforcement personnel and rearrest, without necessarily leading to an


\(^{38}\)See Mark W. Lipsey & David B. Wilson, Effective Intervention for Serious Juvenile Offenders: A Synthesis of Research, in Serious and Violent Juvenile Offenders, supra note 27, at 313, 318 (finding 12% reduction).
adjudication of guilt. Viewed in this light, the results of the meta-analysis suggested, first of all, that participation in VOM is clearly associated with a decrease in subsequent delinquent behavior that leads to an adjudication of guilt. The relationship between VOM participation and subsequent delinquent behavior under the broad definition of reoffense is less clear. The methodological quality of the studies in which the broad definition of reoffense was used was quite low. The mean methodological quality scale score of .19 for these studies indicated that precious few methodological features were employed in order to create initially equivalent groups. This clearly suggested that these studies were potentially plagued by biased group formation. This makes it difficult to interpret the above results for studies using the broader definition. Hence, future research in which researchers decide to use the broader definition of reoffense very clearly needs to employ random assignment and other procedures in order to create unbiased groups.

As noted earlier, it may be argued on conceptual and/or other grounds that the difference between the narrow and broad definitions of reoffense are minor and of small consequence, and that both definitions are equally valid indicators of delinquent activity. The results of the meta-analysis in which the difference between the definitions of reoffense were ignored also suggested that there is a relationship between participation in VOM and subsequent delinquent behavior. The results of these analyses suggested that all significant variation in magnitude of VOM effects could be explained by two variables, δ and duration of follow-up, and that after controlling for these factors there was a clear relationship between VOM participation and subsequent delinquent behavior. These results suggested that, as the methodological quality of a study approached that of a randomized experiment, the expected results would show VOM participants, at six months after their involvement in VOM, to be reoffending at a rate nearly 27% lower than that of nonparticipants, and that this reduction in reoffense attenuates to about 9% at thirty months after court involvement. These results were both consistent with, and divergent from, those obtained when explicitly representing the definition of reoffense in the analyses. They were consistent in that there was a clear relationship between VOM participation and lower reoffense and that methods factors accounted for all of the significant variation in VOM effects. This consistency suggests that the finding of a relationship between VOM participation and subsequent delinquent behavior does not depend upon any claim that the narrow definition of reoffense is a more valid indicator of delinquent behavior than is the broad definition. The results are divergent, however, in that the form of the relationship between VOM participation and subsequent delinquent behavior seems to depend upon the issue of the relative validity of narrow and broad definitions.

The finding that definition of reoffense may be a methodological factor of major significance has important implications for the measurement of “subsequent delinquent behavior” in future research on VOM. Future research needs to employ
more than just archival measures of delinquent behavior. All the studies in this meta-analysis used juvenile court records, law enforcement records, and/or VOM program records. Use of such archival data, whether the narrow or broader definitions are used, may result in underestimates of the extent of juvenile involvement in delinquent behavior since juveniles may commit delinquent offenses for which they are neither apprehended nor adjudicated guilty. The use of archival data constrains the researcher to adopt a definition of “subsequent delinquent behavior” or “reoffense” consistent with the archival source. As the above results suggested, the definition adopted may have significant impact on the magnitude of the VOM effects observed. To remedy these potential problems, alternate approaches to measuring delinquent behavior subsequent to VOM participation need to be used. These include self report methods, like those used by Elliott, Huizinga, and Morse; and multi-informant approaches such as the various report forms of the Child Behavior Checklist. The use of these alternate approaches to measuring delinquent behavior, particularly in a multimethod approach to measuring the delinquent behavior, can improve the validity of the measurement in future research on VOM.

Precious few studies have been done that investigated the link between VOM participation and severity of reoffense. The results of the analyses of the severity outcomes were consistent with homogeneity of results across studies, and hence, with successful replication of findings that VOM participants tend to commit reoffenses of lesser severity than nonparticipants. These meta-analytic findings should be considered to be very preliminary given the exceedingly small number of studies involved.

The findings of this meta-analysis lend strong support to the efforts of those who have sought to influence juvenile court policies to facilitate the implementation of VOM programs in communities throughout the United States. The results clearly imply that VOM participation is associated with a reduction in delinquent behavior and that this outcome may be robust to data analysis methods and to how “recidivism” is defined. The results also suggest, however tentatively, that VOM participants also tend to commit reoffenses of lesser severity. These findings clearly support the development of VOM programs across the country and the adoption of juvenile justice policies that emphasize restorative justice approaches. These results also help focus the debate on how to deal with youths involved in delinquent behavior. There have been calls for harsher treatment of juvenile offenders as a way of curtailing delinquent behavior. The proponents of such approaches to juvenile crime argue that this “just deserts”

philosophy will serve to deter juveniles from engaging in delinquent activities. Recent research suggests that this notion may be in error and could even lead to the institutionalization of juvenile justice practices that create more serious offenders. The results of this meta-analysis suggest that approaches to juvenile justice based on a restorative justice philosophy may hold great promise for the development of juvenile justice practices that lead to more positive outcomes for victims, juveniles, and for the general public.

APPENDIX A


EDWARD DICK, VICTIM OFFENDER RECONCILIATION PROGRAM OF MENDOCINO COUNTY (1999).


SHIRLY LEE, SANTA CLARA CTY. PROB. DEP’T, VICTIM OFFENDER MEDIATION PROGRAM EVALUATION (1999).


DAVID SHICHOR ET AL., VICTIM OFFENDER MEDIATION IN ORANGE COUNTY, CALIFORNIA (2000).


