Published in final edited form as:

Prev Sci. 2013 February ; 14(1): . doi:10.1007/s11121-012-0298-x.

The Application of Meta-analysis within a Matched-pair Randomized Control Trial: An Illustration Testing the Effects of Communities That Care on Delinquent Behavior

Kathryn C. Monahan,

University of Pittsburgh, 210 South Bouquet Street, Pittsburgh, PA 15260, USA

J. David Hawkins, and

Social Development Research Group, University of Washington, Seattle, WA, USA

Robert D. Abbott

College of Education, University of Washington, Seattle, WA, USA

Kathryn C. Monahan: monahan@pitt.edu

Abstract

Use of meta-analytic strategies to test intervention effects is an important complement to traditional design-based analyses of intervention effects in randomized control trials. In the present paper, we suggest that meta-analyses within the context of matched-pair designs can provide useful insight into intervention effects. We illustrate the advantages to this analytic strategy by examining the effectiveness of the Communities That Care (CTC) prevention system on 8th-grade delinquent behavior in a randomized matched-pair trial. We estimate the intervention effect within each of the matched-pair communities, aggregate the effect sizes across matched pairs to derive an overall intervention effect, and test for heterogeneity in the effect of CTC on delinquency across matched pairs of communities. The meta-analysis finds that CTC reduces delinquent behavior and that the effect of CTC on delinquent behavior varies significantly across communities. The use of meta-analysis in randomized matched-pair studies can provide a useful accompaniment to other analytic approaches because it opens the possibility of identifying factors associated with differential effects across units or matched pairs in the context of a randomized control trial.

Keywords

Delinquent behavior; Prevention; Communities That Care; Meta-analysis

One of the primary goals of prevention science is to identify prevention programs or interventions that are tested and effective. To serve this goal, the lion's share of analytic approaches used by prevention scientists test whether or not the intervention produces a significant difference from a control group or a comparison group. In recent years, prevention science has called for tests of whether the effects of interventions are universal or if the effects of an intervention are stronger among some subpopulations than others (Flay et al. 2005). A less studied, but equally important goal, is to understand why some applications of a prevention system or intervention are more successful than other applications. That is, across a number of organizations, schools, or communities that implement an intervention, there is undoubtedly variation in the success of that intervention. For example, with respect

to community-based prevention programs, it is likely the case that characteristics of a community can impact the strength of the effect of the intervention. Yet, relatively little research seeks to understand how characteristics of the implementation of the intervention lead a program or system to be more or less successful. In the context of a randomized matched-pair design, examining how characteristics of a pair contribute to the success of an intervention is an important step in understanding the effects of the intervention.

One reason that little research investigates heterogeneity within matched pairs with respect to intervention effects is that, analytically, traditional approaches for testing intervention effects cannot estimate a randomly varying intervention effect (i.e., if the effect of an intervention varies across matched pairs). In any study design where the intervention is randomized to a matched pair (a one-to-one match), the level 2 intervention effect is synonymous with the level 3 grouping. For example, in a matched-pair trial of a community-based prevention program, an appropriate hierarchical model would consist of three levels. At level 1 would be youth outcomes, which are viewed as nested within communities (level 2) and intervention and control communities (matched pairs; level 3). The test of the intervention effect is a level 2 predictor of level 1 outcomes. If the study design uses one-to-one matching where one control community is matched to one experimental community, the test of the intervention effect at level 2 is synonymous with the clustering at level 3. That is, when estimating the intervention effect at level 2, the information estimated is identical to the level 3 grouping, meaning that the intervention effect *cannot* be allowed to vary at random across level 3 units (matched pairs). Thus, while hierarchical models are ideal for testing the overall effects of an intervention, they do not allow for investigators to test if the intervention was relatively more or less successful across matched pairs.

One analytic strategy that overcomes this statistical limitation in the context of this commonly used matched-pair design in prevention experiments is to conduct a meta-analysis of matched pairs within the randomized control trial. A meta-analytic strategy allows for an investigation of *intervention effects within each matched pair*, the *overall intervention effect across matched pairs*, and the *variance of the intervention effect across matched pairs*. The ability to examine if the intervention effect varies significantly across matched-pairs has the potential to answer important questions about why an intervention may be more successful in one set of matched pairs compared to other matched pairs. Indeed, testing for significant variance of an intervention effect across matched pairs is an important supplemental analysis to tests of the overall intervention effect. Yet meta-analytic approaches of matched-pair randomized trials that fail to account for the hierarchical nature of the data are incorrectly assessing effect sizes and standard errors. Specifically, each matched pair is not independent of the others, since the effects are drawn from the same randomized control trial. This additional source of variance, called the between-group variance, is important to account for in matched-pair randomized trials in order to accurately reflect the structure of the data and the study design. Although traditionally this was a valid and significant limitation of conducting a meta-analysis of such data, recent advances in statistics have enabled a meta-analysis to take both within-group (i.e., within matched pair) and between-group (i.e., between matched pair) sources of variance into account. One possibility is to account for between sources of variance when deriving an effect size (Dunlap et al. 1996). Alternatively, meta-analysis can be estimated within a hierarchical linear modeling (HLM) framework, which simultaneously takes into account both within-group variance as well as between-group variance (Raudenbush and Bryk 1996). Thus, it is possible to conduct a meta-analysis of matched pairs in a single randomized control trial and to accurately account for variance due to the nested structure of the analytic design.

Ultimately, analyzing intervention effects with a meta-analytic approach could allow researchers to assess why an intervention might be more successful in certain matched pairs

of communities compared to others, yet this technique is not often applied in the context of matched-pair randomized control trials. Given that utilizing this strategy could allow investigators to address important questions about implementation and success of an intervention, it is important for prevention science that researchers take stock of variance in intervention effects within a given sample. One reason that meta-analysis of matched-pair randomized control trials is not commonly applied is that information regarding this approach is generally not published in prevention science outlets. Indeed, to conduct such an analysis would require the use of a number of different primary sources from statistical literature. Because this approach is new to prevention scientists, particularly in terms of application to analyzing intervention effects within a single matched-pair randomized trial, the goal of the present study is to illustrate the usefulness of a meta-analytic technique, providing the conceptual and statistical bases for these models. The present study illustrates the use of meta-analyses for prevention scientists using data from the Community Youth Development Study, which tests the effectiveness of the Communities That Care (CTC) prevention system in preventing delinquent behavior prior to the end of the eighth grade in a randomized matched-pair study of 24 communities.

Communities That Care

The CTC system operates at the community level and guides a coalition or board of community stakeholders to select preventive interventions from a menu of diverse tested and effective prevention programs that address the specific risk factors which assessments show are elevated within the community. Because risk and protective factors for delinquency vary across communities, this prevention system provides a community-specific assessment and a list and description of interventions that have been found in controlled studies to prevent problem behaviors. Communities select from this list programs that address their prioritized risk and protective factors. To capitalize on the heterogeneity across communities, the present study uses a meta-analytic strategy to derive an effect size for each of the matched pairs of communities in the study and to estimate an overall effect size of the CTC prevention system on delinquent behavior across the 12 matched pairs of communities in the study.

The CTC system is guided theoretically by the social development model (SDM) (Catalano et al. 1998; Hawkins and Weis 1985), which posits that social development is a process in which key socialization elements – families, schools, and peers – influence behavior directly and indirectly. Within each unit of socialization, three process variables (opportunities for involvement and interaction, skills for participation, and reinforcement for behavior) determine whether a youth will develop a bond of attachment to, commitment to, and belief in conventional (e.g., nondeviant) or nonconventional (e.g., deviant) society.

The SDM is an integration of elements of control, differential association, and social learning theories. Tests of control theory (Hirschi 1969) have found that ‘attachment’ to family, school, and conventional others; ‘commitment’ to conventional action; and ‘belief’ in the validity and legitimacy of the legal order are elements of a social bond to conventional society. This bond to society prevents delinquency because youth have become bonded to conventional aspects of society. From a control perspective, prevention efforts should seek to strengthen the elements of the social bond (attachment, commitment, and belief) between a youth and conventional society. While control theory specifies the elements of the social bond that protect an individual from becoming delinquent, the theory does not specify how social bonds develop within units of socialization. The SDM uses differential association (Matsueda 1982, 1988; Matza 1969; Sutherland 1973) and social learning theory (Akers 1977; Akers et al. 1979) to identify the process by which social bonds develop and delinquent behavior is extinguished or maintained.

Differential association posits that youth are exposed differentially to opportunities to engage in behavior. Specifically, youths may encounter opportunities for interaction with prosocial others or opportunities to interact with those engaged in antisocial behavior. Within the opportunities described by differential association, social learning theory suggests that behavior is learned when it is rewarded and it is not learned or is extinguished when it is not rewarded or is punished. In the context of a social interaction with prosocial others or others involved in antisocial behavior, the skill of the individual in these interactions influences the reinforcement that individual receives. Positive reinforcement for skillful involvement is expected to create or strengthen a bond to the others with whom the individual interacts. If the individual develops strong bonds to prosocial others, these are expected to lead to conforming or prosocial behavior. If the individual receives positive reinforcement for involvement with those involved in antisocial behavior, this is expected to strengthen bonding to antisocial others and to increase the likelihood of delinquent behavior (see Catalano and Hawkins 1996 for full explication of the SDM).

Thus, the SDM combines principles from control, social learning, and differential association theories to posit that social development is the product of longitudinal socialization of a child in a number of different social contexts (e.g., family, school, peer group, and community), specifically the opportunities encountered by the child in these contexts, the skills the child brings to his or her involvements and interactions in these contexts, and the rewards the child receives from these contexts for his or her interactions and involvement. Within this framework, the specific factors within socialization that contribute to delinquent or nondelinquent behavior vary according to context and developmental age.

Given the diversity in risk and protective factors that predict delinquency or nondelinquency, it is likely that the level of factors that place youths at risk for delinquent behavior in one community may differ from those that are most elevated in another community. The premise of CTC is that prevention efforts should address the specific risk factors within a community that are most prevalent and the protective factors that are most depressed in that community. In addition to targeting specific risk and protective factors within a given community, CTC operates by encouraging community members to utilize the social development strategy (Hawkins 1999), which is essentially the prosocial path outlined in the SDM. The social development strategy states that all individuals who interact with children in a community should seek to increase opportunities for prosocial involvement and interactions; help provide youth with the skills to be successful within these prosocial opportunities; and reward and recognize youth for effort, improvement, and achievement in prosocial interactions and involvements. Using this social development strategy should build bonding to prosocial contexts among children, and, in turn, reduce the likelihood of behavior that violates these prosocial expectations, including delinquent behavior. Thus, CTC operates by targeting elevated risk and protective factors with tested and effective programs while simultaneously providing scaffolding through the social development strategy that encourages positive development among youth.

CTC is a prevention system that provides training, materials, and technical assistance to mobilize a coalition of community members to adopt methods based on the advances of prevention science (Coie et al. 1993; Mrazek and Haggerty 1994) to address adolescent drug use and delinquency. Coalitions of key stakeholders within the community are formed and the CTC Youth Survey is administered to adolescents in the community. The CTC Youth Survey assesses multiple risk and protective factors predictive of adolescent problem behaviors, including adolescent substance use and delinquency. The CTC coalition uses the results of the CTC Youth Survey to identify two to five risk factors that are elevated as well as protective factors that are depressed in the community. The coalition then uses the CTC

Prevention Strategies Guide to choose and implement developmentally appropriate tested, effective preventive interventions that seek to change these specific risk and protective factors. Collaborating organizations in the community are trained to implement the new policies and deliver the new prevention programs selected by the CTC coalition. The CTC coalition monitors implementation of all new prevention programs. Thus, the CTC coalition of community members determines the type and number of tested and effective programs that are implemented in their community and encourages the use of the social development strategy to promote prosocial bonding.

The CTC system is expected to produce community-level changes in prevention service system characteristics, including greater adoption of science-based prevention programs, increased collaboration among service providers, increased use of tested and effective prevention programs, good implementation of tested and effective prevention programs, and adoption of the social development strategy by those in a variety of roles in the community who interact with children by promoting opportunities for prosocial involvement and interaction, teaching skills for prosocial participation, recognizing skillful prosocial involvement, and as a result strengthening bonding to prosocial others. Changes in prevention service systems are expected to produce reductions in the risk factors and increase the protective factors targeted by a specific CTC coalition which, in turn, should reduce problem behaviors among youth. According to CTC's theory of community-level change, it should take between 2 and 5 years to observe community-level changes in targeted risk and protective factors, and 4 to 10 years to observe community-level changes in adolescent problem behaviors (Hawkins and Catalano 2009).

To date, the CTC system has been implemented in several countries (United States, United Kingdom, the Netherlands, Canada, Cyprus, Germany, and Australia). CTC has been placed in the public domain by the Center for Substance Abuse Prevention of the federal Substance Abuse and Mental Health Services Administration. All materials for CTC are available on the internet (<http://www.communitiesthatcare.net>). Nonrandomized evaluations of CTC have indicated that the prevention system helps communities to develop more effective prevention service systems and that CTC can reduce levels of risk exposure and adolescent drug use within a community (Feinberg et al. 2007; Greenberg et al. 2005). The Community Youth Development Study (CYDS) is the first community-randomized trial of CTC (Fagan et al. 2008; Hawkins et al. 2008). The CYDS study was designed to determine whether CTC reduces levels of risk, increases levels of protection, and reduces the incidence and prevalence of tobacco, alcohol, and other drug use and delinquency in adolescence. The CYDS study includes a repeated cross-sectional design and a longitudinal panel of individuals followed annually (Brown et al. 2009; Hawkins et al. 2009). The present study focuses on the longitudinal panel of youth in intervention and control communities followed from fifth through eighth grade (Brown et al. 2009).

Previous analyses of the CYDS have found that the CTC system has been successfully implemented with fidelity in intervention communities (Quinby et al. 2008) and found that experimental communities implementing the CTC system were more likely to adopt a science-based approach to prevention and have higher levels of community collaboration than control communities (Brown et al. 2007). Studies have also documented that experimental communities use tested and effective preventive programs (Fagan, Hanson et al. 2008). Furthermore, 3 years after implementation, hypothesized effects of CTC on targeted risk factors and the incidence of delinquent behavior were observed (Fagan et al. 2008). Four years after implementation, the CTC program was associated with lower incidence of alcohol use, binge drinking, smokeless tobacco use, and delinquency in experimental communities compared with control communities (Hawkins et al. 2009).

Present Study

The present study examines the impact of the CTC system on the variety of delinquent behaviors in the eighth grade. In previous analyses of this same data, Hawkins and colleagues (2009) tested the effect of CTC on delinquent behavior in the eighth grade, accounting for fifth-grade levels of delinquency as well as other covariates (e.g., age, sex, parental education, race, ethnicity, religious attendance, and rebelliousness). Hierarchical linear modeling was used to model the nesting of the data (children within communities, intervention or control community, matched community pairs). Controlling for fifth-grade individual and community characteristics, analyses estimated the effect of the CTC intervention on delinquent behavior, accounting for within- and between-community sources of variance. Results indicated that the CTC communities had significantly lower levels of delinquent behavior compared to matched communities randomly assigned to the control condition (Hawkins et al. 2009).

In this report, we illustrate the use of meta-analysis within a single matched-pair randomized control trial from an HLM perspective as a complement to other methods of testing intervention effects in prevention science. In contrast to the previous analysis of the effects of the CTC prevention system on delinquency in the eighth grade, the present paper: (a) estimates the effect of the intervention within each of the matched-pair communities, (b) combines effect sizes across matched pairs to derive an overall effect of the CTC system on delinquent behavior, and (c) tests if the effect size varies significantly across matched pairs.

Method

Communities in the CYDS were selected from 41 communities in the states of Colorado, Illinois, Kansas, Maine, Oregon, Utah, and Washington that participated in an earlier study, the Diffusion Project (Arthur et al. 2005), a naturalistic study of the diffusion of science-based prevention strategies. The drug abuse prevention agencies within these states identified 20 communities that the agency perceived to be trying to implement risk- and protection-focused prevention services. These 20 communities were then matched, within state, on population size, racial and ethnic diversity, economic indicators, and crime rates, to comparison communities that were not perceived to be using a risk- and protection-focused approach to prevention services (see Table 1 for key community demographics by matched pair). The 20 community pairs were then recruited to participate in the Diffusion Study. In one instance, two comparison communities were identified and this resulted in a total of 41 communities (i.e., one community had two matched-comparison communities). Over the course of the 5 years of the Diffusion Project, in 13 of the 20 pairs neither community in the pair advanced in their use of science-based prevention to the point of utilizing tested, effective preventive interventions to address prioritized community risks (Arthur et al. 2003). As such, these 13 pairs of communities were deemed eligible for inclusion in the CYDS study. Twelve of these matched-pair communities were recruited for the CYDS. One community from each matched pair was assigned randomly by a coin toss to either the intervention (CTC) or control condition.

For communities in the intervention condition, CTC training and implementation began in the summer of 2003. Intervention communities received six CTC trainings delivered over 6 to 12 months by certified CTC trainers. Community leaders were oriented to the CTC system and identified or created a community coalition of diverse stakeholders to implement CTC. Coalition members were trained to utilize survey data of students collected in 1998, 2000, and 2002 in the Diffusion Project (Arthur et al. 2005) to (a) prioritize risk and protective factors to be targeted by prevention actions, (b) choose tested and effective prevention policies and programs that address the community's targeted risk and protective factors, (c) implement these interventions with fidelity, and (d) monitor implementation and

outcomes of newly installed prevention programs. Because the CYDS study was initially funded by a 5-year grant, CTC communities in CYDS were asked to focus their prevention planning efforts on programs for youths aged 10 to 14 (approximately Grades 5 through 9) and their families and schools so that effects on delinquency and drug use could be observed within the grant period. CYDS implementation staff provided technical assistance to the communities through weekly phone calls, emails, and a minimum of one site visit per year to each CTC community.

By June of 2004, coalitions in intervention communities had selected prevention programs to address their prioritized risk factors and had created plans to implement these programs with fidelity. Across the 12 intervention communities, 13 different tested and effective prevention programs were selected for implementation in the 2004–2005 school year, 16 programs were selected for implementation during the 2005–2006 school year, and 14 programs were selected for implementation during the 2006–2007 school year. These included school-based programs (All-Stars, Life Skills Training, Lion's Quest Skills for Adolescence, Project Alert, Olweus Bullying Prevention Program, and Program Development Evaluation Training), community-based, youth-focused programs (Participate and Learn Skills, Big Brothers/Big Sisters, Stay Smart, and academic tutoring), and family-focused programs (Strengthening Families 10–14, Guiding Good Choices, Parents Who Care, Family Matters, and Parenting Wisely) (Fagan, Hanson et al. 2008). Each year, community coalitions implemented from one to five of these programs to address their own community-targeted risk and protective factors. On average, three programs were implemented per community in each year. The new programs were implemented by local providers, including teachers for school programs, health and human service workers for community-based, youth-focused, and family-focused programs, and community volunteers for tutoring programs and Big Brothers/Big Sisters.

Each of the programs selected for implementation within the intervention communities was selected because it had been found to be effective in at least one well-controlled trial in preventing substance use (tobacco, alcohol, or other drug use) or delinquent behavior among youths in Grades 5 through 9. For this trial, alcohol policy changes (e.g., tax increases, social host liability, key registration) were not included in the menu of tested programs since the focus of the study was on youth in Grades 5 through 8, and these interventions have not been found to be effective in changing the behavior of youth in this age range (Spath et al. 2008). However, policies and changes in policies related to substance use and delinquent behavior were monitored in both the intervention and control communities through the duration of the study period (see Hawkins et al. 2008 for more information on study design).

Participants and Procedures

Data on adolescent substance use and delinquent behavior were obtained from annual surveys of a panel of public school students who were in the fifth grade during the 2003–2004 school year in the 24 CYDS communities (see Brown et al. 2009 for a complete description of the longitudinal panel methodology). Recruitment of students began in the fall of 2003 by mailing information packets and making in-person calls to each school district superintendent and elementary and middle school principals within the 24 CYDS communities, asking for their commitment to participate in the study and outlining the requirements of involvement in the coming year. As a result, 28 of 29 school districts (88 schools) agreed to participate. All students in fifth-grade classrooms during the 2003–2004 school year in these schools were eligible to participate in the study (see Fig. 1 for flow of participant recruitment and enrollment).

During the second wave of data collection (Grade 6), an effort was made to recruit additional students who were not surveyed in Grade 5. During Grades 5 and 6, parents of

4,420 students (76.4 % of the eligible student population) consented to participate in the study. Final consent rates did not differ by intervention condition (consent rates were 76.1 % for the intervention and 76.7 % for the control communities). Eleven percent ($n=404$) of the students consented in Wave 1 were ineligible for participation in Wave 2 because they moved out of the school district before participating in the study for one semester ($n=388$), did not remain in the grade cohort (i.e., skipped or were held back a grade; $n=4$), were in foster care and did not have consent from state authorities to participate ($n=7$), or were unable to complete the survey on their own due to severe learning disabilities ($n=5$). Thirteen of the originally consented students were absent during the scheduled dates of data collection and were not available for initial surveying. The final active longitudinal panel consisted of 4,407 students (2,194 girls, 2,213 boys; 55 % from intervention communities) in 77 elementary and middle schools in Grade 6 (41 schools in intervention communities and 36 schools in control communities). Students who remained in intervention or control communities for at least one semester were tracked and surveyed annually, even if they left the community. Retention in the study was excellent, and 96 % of the students in the panel completed the survey in the eighth grade (Wave 4).

At each wave of data collection, students completed the Youth Development Survey (YDS) (Social Development Research Group 2005–2007), a self-administered, paper-and-pencil questionnaire designed to be completed in a 50-minute classroom period. To ensure confidentiality, identification numbers, but no names or other identifying information, were included on the surveys. Parents of the students provided written informed consent for their children's participation in the study; students read and signed assent statements indicating that they were fully informed of their rights as research participants. Upon completing the surveys, students received small incentive gifts worth approximately \$5 to \$8 (Brown et al. 2009; Hawkins et al. 2008). These procedures were approved by the Human Subjects Review Committee of the University of Washington.

Tested prevention programs were implemented in CTC communities beginning in the summer and fall of 2004. Data were collected annually (in fifth, sixth, seventh, and eighth grade). The fourth wave of data was collected in the spring of 2007, approximately 2.67 years after the prevention programs chosen by the intervention communities were first implemented.

Measures

Delinquent Behavior—Individuals were asked how many different types of delinquent behavior they had engaged in during the past month. In the fifth grade, youth reported on four different types of delinquent behavior (e.g., stealing, property damage, shoplifting, attacking someone with intention of hurting them). In the eighth grade, youth self-reported on these original four items, as well as five more serious types of delinquent behavior (carrying a gun to school, beating up someone, stealing a vehicle, selling drugs, and being arrested). The number of different types of delinquent behavior a youth endorsed was calculated at each age. These variety scores ranged from 0 to 4 in the fifth grade and 0 to 9 in the eighth grade, with higher scores indicating greater delinquency. Variety scores are commonly used to assess criminal activity (Hindelang et al. 1981) and are advantageously less subject to issues of memory recall, particularly when the delinquent activity is frequent.

Control Variables—A number of characteristics were used as control covariates in the analyses: age at time of the Grade 6 survey, sex, race (dichotomized as White or not White, ethnicity (dichotomized to reflect being Hispanic or not), parental education (ranging from grade school or less to a graduate or professional degree; range=1 to 6), attendance of religious services in Grade 5 (*never to once a week or more*; range=0 to 4), and a three-item

scale measuring rebelliousness in Grade 5 (e.g., “*I like to see how much I can get away with;*” scores ranged from 0 [very false] to 4 [very true]; $\alpha=.69$).

Plan of Analyses

Although the proportion of individuals with missing data in the present study was quite small (from 0 % to 2.7 % for delinquency items in Grade 8), any amount of missing data can bias results. Consequently, missing data were dealt with via multiple imputation (Schafer and Graham 2002). Using NORM version 2.03 (Schafer 2000), 40 separate data sets including data from all four waves were imputed separately by intervention condition. Imputation models included student and community characteristics, drug use and delinquent behavior outcomes, and community memberships.

After imputing missing data, we calculated the adjusted means of delinquent behavior within each community, adjusting for age, sex, race, ethnicity, parental education, religious attendance, rebelliousness, and delinquency in the fifth grade. Adjusted means were calculated as the predicted mean for each community at the average of all covariates (i.e., average parental education, religious attendance, baseline delinquency). Subsequently, the standardized mean difference in delinquent behavior was calculated using these adjusted means in each of the matched-pair communities (i.e., delinquent behavior in experimental communities compared to control communities):

$$ES = \frac{\bar{X}_{\text{experimental adjusted}} - \bar{X}_{\text{control adjusted}}}{S_{\text{pooled}}}, \quad (1.1)$$

where the effect size (ES) is the difference between the mean of the experimental and the control community divided by the pooled standard deviation of the experimental and control community:

$$S_{\text{pooled}} = \sqrt{\frac{(n_{\text{experimental}} - 1)s_{\text{experimental}}^2 + (n_{\text{control}} - 1)s_{\text{control}}^2}{(n_{\text{experimental}} - 1) + (n_{\text{control}} - 1)}}, \quad (1.2)$$

where $s_{\text{experimental}}^2$ is the standard deviation for the experimental group squared, s_{control}^2 is the standard deviation for the control group squared, and $n_{\text{experimental}}$ and n_{control} refer to the total number of individuals in the matched experimental and control groups, respectively. Because the standardized mean difference estimate can be upwardly biased when based on small samples, and some communities in the matched pairs had relatively small sample size (five communities had less than 75 participants in the longitudinal panel), we used the correction recommended by Hedges (1981) for small-sample bias. This unbiased effect size was calculated within each pair:

$$ES' = \left[1 - \frac{3}{4N - 9} \right] ES, \quad (1.3)$$

$$SE_{ES'} = \sqrt{\frac{n_{\text{experimental}} + n_{\text{control}}}{n_{\text{experimental}} n_{\text{control}}} + \frac{(ES')^2}{2(n_{\text{experimental}} + n_{\text{control}})}}, \quad (1.4)$$

where N is the total sample size ($n_{\text{experimental}} + n_{\text{control}}$), ES is the biased standardized estimate of the mean difference, and $n_{\text{experimental}}$ and n_{control} are the number of subjects in the experimental and control groups, respectively. Next, each effect size is weighted to the proportion of its variance:

$$w_{ES} = \frac{1}{SE^2} = \frac{2n_{\text{experimental}}n_{\text{control}}(n_{\text{experimental}}+n_{\text{control}})}{2(n_{\text{experimental}}+n_{\text{control}})^2 + n_{\text{experimental}}n_{\text{control}}(ES')^2}, \quad (1.5)$$

Weighting each effect size as the inverse of the variance around the effect size essentially weights matched-pairs with larger numbers of adolescents more heavily than samples based on smaller numbers of youth. This is because the variance around the effect size is impacted by the proportion of individuals within each matched pair (see Formula 1.4). Thus, smaller communities have larger standard errors (and variance) and are weighted less in deriving the overall effect size across matched pairs compared to larger communities. Because our effect sizes are calculated as the difference between the experimental group and the control group (experimental – control), negative effect sizes reflect that the experimental group has *lower* levels of delinquent behavior relative to the matched control, while positive effect sizes reflect that the experimental group has *higher* levels of delinquent behavior relative to the matched control.

After calculating the unbiased effect size within each of the 12 matched-pair communities, we conducted a fixed effects analysis to determine the average effect size across all matched pairs. In a matched-pair design, there are two sources of variance: variance *within* pairs and variance *across* pairs. Consequently, it was necessary to take into account both types of variance when conducting the meta-analysis. There are two techniques to conduct this analysis. Researchers can calculate the variance of the effect size using within-pair variance and across-pair variance (Dunlap et al. 1996). Alternatively, researchers can conduct analyses in a hierarchical modeling framework, where effect sizes (Level 1) are viewed as being nested within pairs (Level 2), and between-pair variance is accounted for at Level 2. In the present analyses, we used hierarchical linear modeling (HLM) to conduct the meta-analysis.

Using (HLM, version 6.0) (Raudenbush et al. 2004), we estimated the average effect size across all 12 matched pairs. In the first model, we did not allow the intercept (mean effect size) to vary. Level 1 is defined as:

$$d_j = \delta_j + e_j, \quad (1.6)$$

Where d_j is an estimate of δ_j . For matched pairs $j=1, \dots, J$ and level 2 is defined as:

$$\delta_j = \gamma_0, \quad (1.7)$$

In the second model, we tested if there was significant heterogeneity in the effect size by allowing the intercept (mean effect size) to vary across matched pairs.

Importantly, because the unbiased estimate of effect size and standard error of the effect size were calculated within each matched-pair community, we treated Level 1 variance as known, specifying that the model utilize our calculations of the variance around the effect size. The advantage of using an HLM perspective is that both within-community and between-community variance is accounted for when deriving the average effect size across matched pairs of communities. Analyses were run in each of the 40 imputed data sets and aggregated together by Rubin's rules (Rubin 1987).

Results

Unbiased estimates of effect sizes were calculated and corrected for small sample bias within each matched-pair community. Figure 2 presents a box and whisker plot that illustrates the mean effect size of CTC on delinquency within each matched-pair. Within most communities, the experimental community had significantly lower levels of delinquency in the 8th grade compared to its matched pair. A few experimental communities were no different from their respective matched pair with respect to 8th grade delinquency, and one experimental community had higher levels of eighth grade delinquency compared to its matched pair. Subsequently, we estimated the overall effect of CTC on delinquency across the matched pairs.

In the first model, we did not allow the effect size to vary across community. When these 12 effect sizes were aggregated across communities, accounting for within- and between-community variance, the average effect size was -0.31 (Cohen's d ; $SE=0.14$; $p=0.04$), indicating that experimental communities had significantly lower levels of delinquent behavior compared to controls. Converting this effect size to the common language effect size (CLES= 0.59), if a person from an experimental and control site were chosen at random, 59 out of 100 times, the individual from the intervention site would have lower delinquency than the individual from the control site (McGraw and Wong 1992).

Finally, we tested if there was significant heterogeneity across communities in the effect of CTC on delinquent behavior (e.g., we allowed the intercept to vary at random across community matched-pairs). In this model, the overall effect size of the intervention decreased slightly ($d=-0.28$, $SE=0.14$; $p=0.08$). Importantly, we found that the variance around the effect size was significantly different from zero ($\chi^2(11)=188.15797$, $p<0.01$), indicating that the strength of the CTC intervention varied across communities. Thus, our meta-analysis indicates that CTC communities report significantly lower levels of antisocial behavior, but it also found that the strength of the intervention effect varied across matched pairs.

Discussion

An important contribution of the present study is that it documents how meta-analytic techniques can be applied to randomized matched-pair study designs. While the results of the present study are similar to those of other analyses of the CTC data (e.g., Hawkins et al. 2009), analyses that aggregate findings across all communities lose an important piece of information: that the effectiveness of an intervention may vary across matched pairs. Indeed, the present study finds evidence of just that. The meta-analytic technique used in the present paper allows documentation of the consistency—or lack thereof—of the intervention effect across matched pairs randomly assigned to condition. From a prevention science perspective, utilizing meta-analysis within a single matched-pair trial provides an important complement to other analytic approaches that may be used to test intervention effects (i.e., HLM pre-posttests of intervention effect).

There are notable advantages to other data analytic approaches, and the present paper does not suggest that meta-analytic techniques should replace other established methods of testing intervention effects. Rather, we argue that meta-analyses allows for an evaluation of how intervention effects may vary across matched pairs and can be an important supplement to analyses of intervention effects. Just as prevention scientists often test if the intervention effect is universal or stronger among a target subpopulation, it is important to investigate if the intervention itself is more or less successful in some implementations than others. Using a meta-analytic approach across multiple outcomes of an intervention could allow

researchers to investigate why some matched pairs may be more successful at creating effects on some outcomes, but not others. Thus, prevention scientists conducting analyses on matched pairs should consider meta-analysis as a useful tool to help understand the effectiveness and heterogeneity in the effectiveness of the intervention. An important next step for researchers will be to identify sources of this heterogeneity. In doing so, an approach that carefully considers the theory of change of a given intervention may provide useful, as well as barriers to intervention implementation or demographics of a given intervention administration that may contribute to the relative success, or failure, of an intervention. Indeed, identifying characteristics that contribute to the success of an intervention allow for an iterative process where subsequent implementations of an intervention can be primed for greater success.

In addition to arguing for the advantages of meta-analysis in matched-pair randomized trials as a supplemental analyses, the results of the present study provide additional support for and extend previous work with the CTC system (Hawkins et al. 2009), suggesting that this community-based prevention system can effectively reduce delinquent behavior at the community level. Furthermore, to the extent that there is heterogeneity in communities in targeted risk and protective factors, the present study supports the use of an intervention framework that flexibly serves the needs of a given community. Determining whether the impact of CTC on delinquent behavior continues throughout mid and late adolescence remains an important next aim of the CYDS project, particularly since delinquent behavior tends to peak in mid adolescence (Farrington 1996).

One important limitation of the present study, and an important consideration for any prevention scientists who apply meta-analysis to their own work, is that the number of matched pairs will impact the specificity of the derived effect size. In our analyses of the CTC effect, we were limited to 12 matched-pair communities that vary in size. Although we utilized corrections to account for variability in sample size and weighted effect sizes on the basis of the standard error, more communities would have enabled greater specificity in estimating the effect size of the CTC intervention. Studies with more matched pairs would be able to derive an effect size with a smaller error and more specificity. Similarly, having more matched pairs would enable researchers to more thoroughly investigate sources of heterogeneity in the effect size across intervention and control sites.

The present study documents that using meta-analytic techniques in randomized control studies can provide useful information about community heterogeneity in outcomes. This heterogeneity can then be explored in relation to other community characteristics to better understand the factors that predict variability in the effectiveness of the intervention. This approach challenges prevention scientists to move beyond estimating the overall effectiveness of a given intervention to understand factors that lead to the strength of the effect of the intervention on outcomes.

Acknowledgments

Funding Note This study was supported by research grant R01 DA015183-03 from the National Institute on Drug Abuse (with cofunding from the National Cancer Institute, the National Institute of Child Health and Human Development, the National Institute of Mental Health, and the Center for Substance Abuse Prevention).

References

- Akers, RL. Deviant behavior: A social learning approach. 2. Belmont, CA: Wadsworth; 1977.
- Akers RL, Krohn M, Lanza-Kaduce L, Radosevich M. Social learning and deviant behavior: A specific test of a general theory. *American Sociological Review*. 1979; 44:636–655. [PubMed: 389120]

- Arthur, MW.; Ayers, CD.; Graham, KA.; Hawkins, JD. Mobilizing communities to reduce risks for drug abuse: A comparison of two strategies. In: Bukoski, WJ.; Sloboda, Z., editors. *Handbook of drug abuse prevention. Theory, science and practice*. New York: Kluwer Academic/Plenum; 2003. p. 129-144.
- Arthur, MW.; Glaser, RR.; Hawkins, JD. Steps towards community-level resilience: Community adoption of science-based prevention programming. In: Peters, RD.; Leadbeater, B.; McMahon, RJ., editors. *Resilience in children, families, and communities: Linking context to practice and policy*. New York: Kluwer Academic/Plenum; 2005. p. 177-194.
- Brown EC, Graham JW, Hawkins JD, Arthur MW, Baldwin MM, Oesterle S, et al. Design and analysis of the Community Youth Development Study longitudinal cohort sample. *Evaluation Review*. 2009; 33:311–334. [PubMed: 19509119]
- Brown EC, Hawkins JD, Arthur MW, Briney JS, Abbott RD. Effects of Communities That Care on prevention services systems: Outcomes from the community youth development study at 1.5 years. *Prevention Science*. 2007; 8:180–191. [PubMed: 17602298]
- Catalano, RF.; Arthur, MW.; Hawkins, JD.; Berglund, L.; Olson, JJ. Comprehensive community and school based interventions to prevent antisocial behavior. In: Loeber, R.; Farrington, DP., editors. *Serious and violent juvenile offenders: Risk factors and successful interventions*. Thousand Oaks, CA: Sage; 1998. p. 248-283.
- Catalano, RF.; Hawkins, JD. The Social Development Model: A theory of antisocial behavior. In: Hawkins, JD., editor. *Delinquency and crime: Current theories*. New York: Cambridge University Press; 1996. p. 149-197.
- Coie JD, Watt NF, West SG, Hawkins JD, Asarnow JR, Markman HJ, et al. The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist*. 1993; 48:1013–1022. [PubMed: 8256874]
- Dunlap WP, Cortina JM, Vaslow JB, Burke MJ. Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*. 1996; 1:170–177.
- Fagan AA, Hanson K, Hawkins JD, Arthur MW. Bridging science to practice: Achieving prevention program implementation fidelity in the Community Youth Development Study. *American Journal of Community Psychology*. 2008; 41:235–249. [PubMed: 18302016]
- Fagan, AA.; Hawkins, JD.; Catalano, RF. Using community epidemiologic data to improve social settings: The Communities That Care prevention system. In: Shinn, M.; Yoshikawa, H., editors. *Toward positive youth development: Transforming schools and community programs*. New York: Oxford University Press; 2008. p. 292-312.
- Farrington, DP. The explanation and prevention of youthful offending. In: Hawkins, JD., editor. *Delinquency and crime: Current theories*. New York: Cambridge University Press; 1996. p. 68-148.
- Feinberg ME, Greenberg MT, Osgood D, Sartorius J, Bontempo D. Effects of the Communities That Care model in Pennsylvania on youth risk and problem behaviors. *Prevention Science*. 2007; 8:261–270. [PubMed: 17713856]
- Flay BR, Biglan A, Boruch RF, Castro FG, Gottfredson D, Kellam S, et al. Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*. 2005; 6:151–175. [PubMed: 16365954]
- Greenberg, MT.; Feinberg, ME.; Gomez, BJ.; Osgood, DW. Testing a community prevention focused model of coalition functioning and sustainability: A comprehensive study of Communities That Care in Pennsylvania. In: Stockwell, T.; Gruenewald, P.; Toumbourou, JW.; Loxley, W., editors. *Preventing harmful substance use: The evidence base for policy and practice*. New York: John Wiley & Sons; 2005. p. 129-142.
- Hawkins JD. Preventing crime and violence through Communities That Care. *European Journal on Criminal Policy and Research*. 1999; 7:443–458.
- Hawkins, JD.; Catalano, RF. *Communities that care community board orientation: Participant's guide*. Seattle: Social Development Research Group, School of Social Work, University of Washington; 2009.

- Hawkins JD, Catalano RF, Arthur MW, Egan E, Brown EC, Abbott RD, et al. Testing Communities That Care: The rationale, design and behavioral baseline equivalence of the Community Youth Development Study. *Prevention Science*. 2008; 9:178–190. [PubMed: 18516681]
- Hawkins JD, Oesterle S, Brown EC, Arthur MW, Abbott RD, Fagan AA, et al. Results of a type 2 translational research trial to prevent adolescent drug use and delinquency: A test of Communities That Care. *Archives of Pediatric and Adolescent Medicine*. 2009; 163:789–798.
- Hawkins JD, Weis JG. The social development model: An integrated approach to delinquency prevention. *Journal of Primary Prevention*. 1985; 6:73–97. [PubMed: 24271382]
- Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*. 1981; 6:107–128.
- Hindelang, MJ.; Hirschi, T.; Weis, JG. Measuring delinquency. Beverly Hills, CA: Sage; 1981.
- Hirschi, T. Causes of delinquency. Berkeley, CA: University of California Press; 1969.
- Matsueda RL. Testing control theory and differential association: A causal modeling approach. *American Sociological Review*. 1982; 47:489–504.
- Matsueda RL. The current state of differential association theory. *Crime and Delinquency*. 1988; 34:277–306.
- Matza, D. Becoming deviant. Englewood Cliffs, NJ: Prentice Hall; 1969.
- McGraw KO, Wong SP. A common language effect size statistic. *Psychological Bulletin*. 1992; 111:361–365.
- Mrazek, PJ.; Haggerty, RJ., editors. Reducing risks for mental disorders: Frontiers for prevention intervention research. Washington, DC: National Academy Press; 1994. Committee on Prevention of Mental Disorders, Institute of Medicine.
- Quinby RK, Fagan AA, Hanson K, Brooke-Weiss B, Arthur MW, Hawkins JD. Installing the Communities That Care prevention system: Implementation progress and fidelity in a randomized controlled trial. *Journal of Community Psychology*. 2008; 36:313–332.
- Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis models. 2. Thousand Oaks, CA: Sage; 1996.
- Raudenbush, SW.; Bryk, AS.; Cheong, YF.; Congdon, RT, Jr. HLM 6: Hierarchical linear and nonlinear modeling. Lincolnwood, IL: Scientific Software International; 2004.
- Rubin, DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
- Schafer, JL. NORM for Windows 95/98/NT. University Park, PA: Center for the Study and Prevention through Innovative Methodology at Pennsylvania State University; 2000.
- Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychological Methods*. 2002; 7:147–177. [PubMed: 12090408]
- Social Development Research Group. Community Youth Development Study, Youth Development Survey [Grades 5–7]. Seattle: Social Development Research Group, School of Social Work, University of Washington; 2005–2007.
- Spoth R, Greenberg M, Turrise R. Preventive interventions addressing underage drinking: State of the evidence and steps toward public health impact. *Pediatrics*. 2008; 121:S311–S336. [PubMed: 18381496]
- Sutherland, EH. Development of the theory. In: Schuessler, K., editor. Edwin H. Sutherland on analyzing crime (1942). Chicago: University of Chicago Press; 1973. p. 13-29.

Fig. 1.
Flow of study communities and participants

Fig. 2.
Mean unbiased effect size of CTC on eighth grade delinquency within each matched-pair and overall effect

Table 1

Community demographics by matched pair from year closest to matching

Matched pair	1990 Population	1990 % White	1995 UCR	1994 FLE	1994 % unemployed
Pair 1: A	2098	98.47	34.1	42.9	N/A
Pair 1: B	1154	93.85	54.8	42.1	N/A
Pair 2: A	25840	87.06	55.9	69.8	7.5
Pair 2: B	15418	88.12	56.3	87.4	N/A
Pair 3: A	39308	91.46	77.6	94.4	5.3
Pair 3: B	25512	89.29	85	85	5.5
Pair 4: A	17767	98.51	64	62.1	N/A
Pair 4: B	21129	79.32	95.4	91.6	N/A
Pair 5: A	4022	99.28	42.5	28.5	4.8
Pair 5: B	7972	99.21	93	65.2	7.9
Pair 6: A	8317	95.67	35.6	67.7	7.3
Pair 6: B	5436	99.15	24.2	41.2	6.2
Pair 7: A	9422	96.38	42.1	50.8	N/A
Pair 7: B	3651	94.41	73.8	49.2	N/A
Pair 8: A	10950	97.25	65.5	61	N/A
Pair 8: B	7535	96.92	57.9	48	N/A
Pair 9: A	13887	93.48	51.6	61	N/A
Pair 9: B	8712	92.06	85.5	53.1	N/A
Pair 10: A	15644	94.75	53.1	57.1	N/A
Pair 10: B	24063	94.52	42.3	54.4	N/A
Pair 11: A	16565	93.76	54.8	52.8	N/A
Pair 11: B	17710	95.31	57.5	46.8	N/A
Pair 12: A	1691	98.11	74.5	57.1	N/A
Pair 12: B	3738	80.77	80	53.2	N/A

UCR Uniform Crime Reports Standard Score from the FBI. FLE Free lunch eligibility