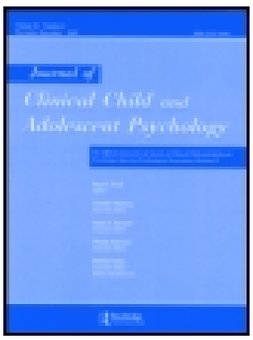

Routledge
Taylor & Francis Group

Journal of Clinical Child Psychology

ISSN: 0047-228X (Print) (Online) Journal homepage: http://www.tandfonline.com/loi/hcap19

Child Sexual Abuse Prevention Programs: A Meta-Analysis

M. Katherine Davis & Christine A. Gidycz

To cite this article: M. Katherine Davis & Christine A. Gidycz (2000) Child Sexual Abuse Prevention Programs: A Meta-Analysis, Journal of Clinical Child Psychology, 29:2, 257-265, DOI: 10.1207/S15374424jccp2902_11

To link to this article: http://dx.doi.org/10.1207/S15374424jccp2902_11

Journal of Clinical Child Psychology
2000, Vol. 29, No. 2, 257–265

Child Sexual Abuse Prevention Programs: A Meta-Analysis

M. Katherine Davis and Christine A. Gidycz
Department of Psychology, Ohio University

Conducted a meta-analytic evaluation of the effectiveness of school-based child abuse prevention programs. Literature searches identified 27 studies meeting inclusion criteria for use in this meta-analysis. The average effect size for all programs studied was 1.07, indicating that children who participated in prevention programs performed 1.07 SD higher than control group children on the outcome measures used in the studies. Analysis of moderator variables revealed significant effects for age, number of sessions, participant involvement, type of outcome measure, and use of behavioral skills training. Most important, programs presented over 4 or more sessions that allowed children to become physically involved produced the highest effect sizes. Although most often used only with younger children, findings suggest that active, long-term programs may be more effective for children of all ages.

Child sexual abuse has received much attention in the past two decades as researchers have begun to document its prevalence. Given a consistent definition of sexual abuse as ranging from fondling to intercourse and perpetrated by an individual at least 5 years older, typical estimates state that 20% of children will be sexually assaulted before the age of 18 (Sandberg, Lynn, & Green, 1994). In response, several programs aimed at teaching children to avoid and report abuse have been implemented. A survey of 400 school districts in the United States found that over 85% had offered a prevention program in the past year, and 64% of these districts mandated programs (Daro, 1994). Telephone interviews of a representative sample of children in the United States revealed that 67% had been exposed to a prevention program (Finkelhor & Dziuba-Leatherman, 1995). The programs to which these children had been exposed were almost exclusively school based and typically took place in late elementary school. Given the pervasiveness of prevention programs and the importance of their mission, it is essential that they be systematically evaluated.

Recent evaluations examining several child abuse prevention programs have suggested that children both enjoy them and show significant improvements in abuse-related knowledge and skills following participation (e.g., Binder & McNiel, 1987). Knowledge content generally involves information on body ownership and the touch continuum, and targeted skills include recognizing abusive situations, saying no, and telling a trusted adult about the abuse (e.g., Saslawsky & Wurtele, 1986; Tutty, 1992). Knowledge and skill gains are typically maintained at follow-up 2 to 12 months later (e.g., Fryer, Kraizer, & Miyoshi, 1987b; Saslawsky & Wurtele, 1986).

Although some basic concepts concerning prevention strategies are common to all school-based programs, the programs vary widely on a number of dimensions including format, the age of the target audience, the occupation of the program leader, and the length of the program. Although it is hypothesized that many of these variables could influence program effectiveness, researchers are only beginning to investigate the impact of these key components.

One variable that has been examined is program format. Wurtele, Marrs, and Miller-Perrin (1987) found that behavioral skills training, including active skills rehearsal, shaping, and reinforcement, produced significantly higher knowledge and skill gains than did simple modeling of skills by a presenter. Other studies have corroborated that more active programs that include role plays and participant rehearsal are more effective than less active modes of presentation (e.g., films or lectures) in improving prevention knowledge and skills of children (Blumberg, Chadwick, Fogarty, Speth, & Chadwick, 1991; Wurtele, Saslawsky, Miller, Marrs, & Britcher, 1986).

Age is a second component examined by researchers. It has been documented consistently that older children are better able to learn prevention concepts than younger children (e.g., Blumberg et al., 1991; Conte, Rosen, Saperstein, & Shermack, 1985), but there appears to be no difference in the medium that works best; that is, children of all ages appear to learn best from programs that include active behavior skills components.

Although data are beginning to emerge regarding the important aspects of prevention programs, narrative

Requests for reprints should be sent to Christine A. Gidycz, Department of Psychology, Ohio University, 231 Porter Hall, Athens, OH 45701. E-mail: Gidycz@oak.cats.ohiou.edu

reviews in the area of child abuse prevention are limited in their ability to integrate the literature. It remains difficult to determine how much benefit children are receiving from abuse prevention programs or just which program components are critical. Variables such as the occupation of the program leaders, the length of the program, and the type of design utilized in the evaluation study are hypothesized to affect program effectiveness (Finkelhor, Asdigian, & Dziuba-Leatherman, 1995) but have not been assessed in any systematic manner. Previous narrative reviews have examined only a limited number of variables that might affect program outcomes.

In contrast to narrative reviews, meta-analysis allows a reviewer to examine a body of literature quantitatively and to identify existing relations even when the outcomes of single studies differ in magnitude, significance levels, or even the direction of effects. Two previous meta-analyses of child sexual abuse prevention programs have been published. Berrick and Barth (1992) examined overall program effectiveness and assessed whether participant age and program evaluation source (internal vs. external) were related to effectiveness. This analysis found a mean effect size of .90, indicating that knowledge levels of children increased by close to 1 *SD* following participation in a prevention program. No significant differences were found among children of different ages or between studies conducted by internal versus external evaluators. This meta-analysis provided some encouraging data regarding the effectiveness of prevention programs in increasing children's knowledge. The analysis was limited by its small sample size, however, which may have lacked the power to identify existing moderating effects. Many potentially important moderating variables were not investigated.

Rispens, Aleman, and Goudena (1997) conducted a second meta-analysis and found an overall postintervention effect size of .71. These authors did examine several additional variables for moderating effects, but they identified and analyzed only 60% of the available empirical studies of school-based child abuse prevention programs. Failure to include all available literature in a meta-analysis may result in biased findings, thus it is important to replicate the findings of these researchers using every available study.

This study is a meta-analysis assessing the overall effectiveness of child sexual abuse prevention programs and the relation between program variables and knowledge gains. The goals of this meta-analysis included: (a) to find an overall effect size, (b) to determine what effect methodological problems have on effect size, (c) to establish whether publication bias exists in the literature, and (d) to explore potential moderator variables. Based on the data available from previous narrative reviews, it was hypothesized that older children would learn more prevention concepts than younger children and that programs involving active participation would be most effective for children of all ages.

Method

Sample

Searches of computer databases, article references, and relevant journals were conducted in an attempt to locate all research evaluating child abuse prevention programs. Computer searches included the following databases: PsycInfo, Medline, ERIC, HealthStar, and Dissertation Abstracts. The reference sections of identified articles were then examined for studies missed in the computer databases. Finally, the most recent volumes of relevant journals (e.g., *Child Abuse & Neglect*) were searched manually.

To be included in the proposed meta-analysis, studies were required to: (a) evaluate a school-based prevention program focusing on sexual abuse, (b) examine children aged 3 to 13, (c) rate effectiveness using a knowledge-based or behavioral outcome measure, (d) include a control group comparison, and (e) be presented in English. All studies that met the inclusion criteria were examined, regardless of methodological quality. With these inclusion criteria, 27 studies with a total of 8,115 participants were located. Twenty-one of the identified studies were published journal articles, three were unpublished doctoral dissertations, two were unpublished internal program evaluations, and one was an unpublished master's thesis.

Data Coding

The studies selected for use in the analysis were coded according to several variables: (a) mean age of children in the study; (b) mode of presentation (e.g., written material, puppets, role plays); (c) level of child participation (physical participation, verbal participation, no participation); (d) qualifications of instructor; (e) sex of instructor; (f) length of program in hours; (g) number of sessions into which the program was divided; (h) time period between program and posttest; and (i) type of outcome measure used. Methodological quality was operationalized by type of control group, random assignment of participants, research design, interviewer status (was the individual administering outcome measures informed or uninformed concerning participants' group membership), and number of items on the outcome measure. Interrater reliability was good (correlations and κ values > .85) for all but one of the variables examined. The one variable with an insufficient interrater agreement level (κ = .664) had only three discrepant codings. Those studies were reexamined, and a coding decision was reached.

Journal of Clinical Child Psychology 2000.29:257-265.

Journal of Clinical Child Psychology 2000.29:257-265.

Examination of Effect Size and Moderator Variables

Effect sizes and their variances were calculated as *d* values for posttest comparisons of control and treatment groups (Hunter & Schmidt, 1990). This statistic is computed as the difference between the two group means divided by the pooled within-group standard deviation. The population treatment effect (*D*) was estimated by averaging individual effect sizes (*d*s) weighted according to the sample size of each study. One effect size for each individual construct represented in each study (verified by tests of homogeneity; Hedges & Olkin, 1985) was used in these calculations.

Moderator variables were examined by breaking the studies into subsets according to the proposed variable and conducting a meta-analysis on each subset separately. Once the subset effect sizes were calculated for each variable, homogeneity tests were run to determine whether the search for moderators should continue. For linear and continuous variables, a significant correlation (two-tailed) with the *d* statistic signified a moderating effect. Categorical moderators were examined using multiple regression.[1]

Examination of Publication Bias

Results of unpublished studies (e.g., unpublished theses and dissertations) were compared with the results of published articles to determine whether a publication bias existed for this body of literature. In addition, a file-drawer analysis (Hunter & Schmidt, 1990) was conducted. These analyses are based on the hypothesis that studies with nonsignificant findings may not be published; they may instead end up filed or thrown away. The file-drawer analysis measured the number of missing studies averaging null results that would be necessary to lower the mean effect size to .50, a moderate effect size, or to .20, a small effect size (Cohen, 1988).

Results

Summary Statistics

From a total of 8,115 participants, 73 separate effect sizes were calculated.[2] One interview outcome measure assessing assertiveness resulted in an effect size of zero (Downer, 1984), but the remaining 72 effect sizes were all positive, ranging from 0.10 to 2.93. The average effect size (*D*), weighted for sample size, was 1.07, with a variance of 0.76 after adjusting for sampling error. By Cohen's (1988) classification, 1.07 would be considered a large effect.

One study (Nemerofsky, Carran, & Rosenberg, 1994) was extremely influential in the results of this meta-analysis due to its unusually large effect size (d = 2.93) and its very large sample size (N = 1,339). The sample size caused the results of this study to be weighted quite heavily. This in combination with the magnitude of its effect size allowed the study to have a large impact on the overall results. With the study by Nemerofsky et al. (1994) removed from the analysis, the average effect size dropped to 0.81 (variance = 0.26), which is still considered a large effect (Cohen, 1988). Because of the undue influence of this one study, analyses are presented both with and without its inclusion where its removal altered the findings substantially.

Publication Bias

The observed correlation between effect size and publication status was .342 (p = .003). A subset examination of the publication status variable indicated that studies that were unpublished yielded a higher mean effect size (D = 1.252) than those studies that were published (D = 1.005). To further examine publication bias, a file-drawer analysis was completed. This analysis revealed that 56 studies with results supporting the null hypothesis (no significant program effect) would be necessary to lower the calculated effect size from large to moderate (.50). To lower the effect size to small (.20), 249 studies with null findings must exist.

Analysis of Methodological Quality

The relation between methodological quality and effect size was assessed using multiple regression. The results suggested that type of control group, research design, interviewer status, random assignment, and number of items on the outcome measure were together significant predictors of effect size, accounting for 43.5% of the total variance in d, $F(5, 37) = 5.708$, $p < .0005$. In the presence of the other variables entered into the regression, interviewer status (informed vs. uninformed concerning participants' group membership) accounted for a significant proportion of the variance in d, $F(1, 37) = 11.203$, $p < .005$. Subset analyses (Table 1) found higher mean effect sizes when studies (a) used interviewers who were informed about the status of participants, (b) did not use random assignment of participants, and (c) used a wait-list control instead of an unrelated alternate program that would control for amount of experimenter contact. These findings indicate that studies with poorer methodology found higher mean effect sizes. The exceptions to this trend were the design variable and the number of items on the outcome

[1]For a more complete explanation of the statistical methodology, see Davis (1997).

[2]A listing of studies, study characteristics, and effect sizes is available in Davis (1997).

measure. Studies that utilized pretests to examine initial control group and experimental group equivalence were considered superior methodologically; these studies found higher effect sizes than those using only posttests. Also, studies with more items on outcome measures found higher effect sizes.

Analysis of Potential Moderator Variables

Mean age, level of participation, and number of sessions all showed significant correlations to the *d* value ($p < .015$). For this reason, the subset effect sizes for all of these variables were examined (see Table 1). Mean age was divided into three groups for the purposes of subset analysis. The weighted mean effect size was highest for the youngest group (*M* age = 3 to 5 years, *D* = 2.143), followed by age group two (*M* age = 5.1 to 8 years, *D* = 1.243), and age group three (*M* age > 8 years, *D* = .770). Removing the study by Nemerofsky et al. (1994) lowered the effect size for the youngest age group to .937, and the other age groups maintained the same effect sizes.

Child participation was analyzed at three different levels: physical participation (e.g., role plays), verbal participation (e.g., discussion), and no participation by the children (e.g., film or lecture). The mean weighted effect size was similar for programs with some type of child participation (physical, *D* = 1.202; verbal, *D* = 1.226) but much lower for programs with no participation (*D* = .453). Without the study by Nemerofsky et al. (1994), the effect size for verbal participation dropped to .657, resulting in a large difference between physical and verbal participation.

The variable for the number of sessions involved in the presentation of a prevention program was divided into three subsets: programs administered in one session, those taught in two to three sessions, and those presented in more than three sessions. A linear trend was found in the mean effect size, with *D* increasing along with the number of sessions from .598 in the one-session programs to 2.153 in programs with more than three sessions. The same trend remained with the removal of the Nemerofsky et al. (1994) study, although the effect size for programs with more than three sessions lowered to 1.536. To ensure that the linear trend was not due to increased time spent learning prevention, a partial correlation between *D* and number of sessions controlling for hours spent in the program was run. Even with program length controlled for, session number was significantly correlated with *D* ($p = .05$).

To measure the significance of type of outcome measure, mode of presentation, and instructor qualifications, regression analyses were used. The first regression examined the type of outcome measure as a predictor of effect size. The three dichotomous variables representing type of outcome measure were use of a questionnaire, use of a vignette interview, and use of a behavioral measure. These variables were significant predictors of effect size when entered into a regression in one block, accounting for 16.2% of the variance in *d*, $F(3, 69) = 4.458$, $p = .0064$. Use of a behavioral measure was the only variable found to be a significant predictor of effect size with all three variables in the equation, $F(1, 69) = 13.337$, $p < .001$. Subset analyses revealed that assessments using behavioral observation (in vivo simulation) or questionnaires detected greater effects than those using vignette interviews. Studies with in vivo simulations had a weighted mean effect size of 1.191, those using a questionnaire found a mean effect size of 1.138, and those with interviews had a mean effect size of .617. Removing the Nemerofsky et al. (1994) study from this regression did not significantly alter the findings. Without this extreme case, the

Table 1. *Average Effect Sizes (*D*) for Subgroups*

Subgroup	D^a	D^b
Interviewer Status = Uninformed	0.655	0.655
Interviewer Status = Informed	2.020	0.967
Random = Yes	0.725	0.725
Random = No	1.131	0.826
Item Number < 20	0.862	0.862
Item Number > 20	1.455	0.619
Type of Control = Alternate Program	0.727	0.727
Type of Control = Wait List	1.198	0.825
Design = Pre- and Posttests	1.249	0.886
Design = Posttest Only	0.601	0.601
Published	1.005	0.632
Not Published	1.252	1.252
Number of Sessions = 1	0.598	0.598
Number of Sessions = 2 or 3	0.663	0.663
Number of Sessions > 3	2.153	1.536
Active Physical Participation	1.202	1.202
Active Verbal Participation	1.226	0.657
No participation	0.453	0.453
Mean Age = 3 to 5	2.143	0.937
Mean Age = 5.01 to 8	1.243	1.243
Mean Age = 8.01 to 12	0.770	0.770
Behavioral Outcome Measure	1.191	1.191
Questionnaire Outcome Measure	1.138	0.838
Interview Outcome Measure	0.617	0.617
Did Not Use Written Material	0.878	0.878
Used Written Material	1.296	0.709
Did Not Use Role Play or Drama	1.226	0.688
Used Role Play or Drama	0.910	0.910
Did Not Use Discussion or Lecture	0.718	0.718
Used Discussion or Lecture	1.094	0.821
Did Not Use Video or Film	1.183	0.704
Used Video or Film	0.923	0.923
Did Not Use Puppet Show	1.084	0.805
Used Puppet Show	0.898	0.898
Did Not Use Behavioral Training	1.024	0.663
Used Behavioral Training	1.210	1.210
Did Not Use Dolls	1.081	0.817
Used Dolls	0.741	0.741

[a]Effect sizes with Nemerofsky et al. (1994) included. [b]Effect sizes with Nemerofsky et al. (1994) removed.

Journal of Clinical Child Psychology 2000.29:257-265.

effect size for studies using a questionnaire dropped to .838.

The seven dichotomous variables representing mode of presentation were use of written material, use of role plays or drama, use of discussion or lecture, use of a video or film, use of puppet shows, use of behavioral skills training, and use of dolls or stuffed animals. When all seven dichotomous variables were entered into a regression in one block, they did not significantly predict effect size, $F(7, 65) = 1.245$, *ns*. In the presence of the other variables, however, the use of puppet shows was a significant predictor, $F(1, 65) = 6.41$, $p < .05$. Studies using puppet shows had a significantly lower mean effect size ($D = .898$) than those not using puppet shows ($D = 1.084$). This finding reversed when the Nemerofsky et al. (1994) study was removed. Without this study, mode of presentation did approach significance as a predictor of effect size, $F(7, 64) = 2.101$, $p = .056$. The use of puppet shows remained a significant predictor, $F(1, 64) = 11.127$, $p < .01$, but in this analysis studies using puppet shows had a significantly higher mean effect size ($D = .898$) than those not using puppet shows ($D = .805$). The use of behavioral skills training was also found to be a significant predictor in this analysis, $F(1, 64) = 4.133$, $p < .05$; studies with behavioral skills training had a mean effect size twice as high as those without such training ($D = 1.210$ and $D = .663$, respectively).

Eight dichotomous variables represented the use of an instructor with one of the following qualifications: teacher, parent, mental health professional, college student, high school student, volunteer, police officer, or researcher. When all eight of these variables were entered into a regression in one block, they did not significantly predict effect size, $F(7, 43) = 1.558$, *ns*. Furthermore, no single instructor qualification was significantly correlated with effect size. The same results were found with removal of the Nemerofsky et al. (1994) study. This variable was, therefore, not considered as a moderator.

Correlations Among Moderator Variables

High correlations were found among most of the moderators identified; significant findings are described here. Because of the number of correlations run, a conservative significance level was used ($p \leq .005$). As the mean age of the children involved in the study increased, programs used fewer sessions to teach the material ($r = -.399$) and used lower levels of participation ($r = .421$). As the number of sessions used in a program increased, the evaluation study was more likely to be unpublished ($r = .630$), and the program was more likely to use active participation ($r = -.442$). Unpublished studies were less likely to use random assignment ($r = .386$) and uninformed interviewers ($r = .753$) or to have the control group participate in an alternate group activity ($r = .323$).

To better understand the relations between age and the other predictors, correlations were run using pairwise comparisons of the age groups. Correlations using only the younger two age groups indicated that children aged 5 to 8 were more likely than children under age 5 to be administered a behavioral outcome measure ($r = .453$). Correlations comparing children under 5 to children over 8 indicated that younger children were more likely to be allowed to actively participate in their programs ($r = .456$) and had a greater number of sessions ($r = -.625$). Correlations using the older two age groups showed that children aged 5 to 8 were more likely than children over 8 to have more sessions ($r = -.349$) and to be evaluated with a behavioral outcome measure ($r = -.526$).

Discussion

The goals of this study were to use meta-analysis to evaluate school-based sexual abuse prevention programs. The overall weighted mean effect size derived from 27 control-group studies was found to be 1.07. This indicates that on average, after controlling for sample size, children who participated in prevention programs performed 1.07 *SD* higher than the control group children on outcome measures used in the studies. The calculated effect size dropped to .81 when one outlying study (Nemerofsky et al., 1994) was removed from the data set. Both numbers are considered to be large effects (Cohen, 1988). Importantly, these effect sizes can only be interpreted as improvements in the prevention-related knowledge and skills of children who participated in programs. No identified studies have examined the effects of prevention programs on abuse prevalence, so it cannot be assumed that children participating in abuse prevention programs are at a lower risk for sexual abuse.

In general, studies with poorer methodology found larger effect sizes, implying that the calculated mean effect size may be artificially inflated by studies using poor methods. Specifically, interviewers who were aware of the group status of study participants found larger effect sizes, demonstrating the power of expectancy effects. Interestingly, the one study that was an outlier (Nemerofsky et al., 1994) was rated as having poor methodological quality and, thus, it may be that the high effect size obtained by this study is due at least in part to the methodological problems.

An examination of moderator variables identified several program and participant characteristics related to effect size. Programs that allowed physically active participation and made use of behavioral skills training such as modeling, rehearsal, and reinforcement pro-

Journal of Clinical Child Psychology 2000.29:257-265.

duced the largest changes in performance level. Additionally, programs that broke down the presentation of material into more than three sessions were much more effective than those programs that lasted for only one, two, or three sessions. This was not due to increased amount of time spent learning about prevention; rather, learning in smaller increments appeared to be beneficial. Children in preschool and early elementary school learned the most from prevention programs. The oldest children, those in late elementary school or early middle school, improved the least with respect to the control group. The manner in which children's learning was evaluated was also of importance in determining effect size; tests of behavioral change found a larger difference between the treatment and control groups than any other outcome measure. These moderating effects are further explored in what follows.

Child Participation Level

Child participation level was one important variable predicting knowledge gain; programs that allowed the children to be active participants found effect sizes three times as high as those with no participation. Without the Nemerofsky et al. (1994) study, a large difference emerged between physical and verbal participation. Although the difference in effect sizes was not large between studies with passive participants and studies where participants were involved in verbal discussion (change in the mean effect size = .24), the shift to active physical participation in a program doubled the effect size. Similarly, behavioral skills training increased the learning of prevention material; studies utilizing behavioral skills training found effect sizes twice as high as those with no behavioral skills training.

Program participation and behavioral skills training may be helpful both because the involvement increases attention level and because the children can practice skills with instructive supervision. Physical participation may have some advantage over verbal participation in teaching prevention skills to children due to their stage of cognitive development. Children have difficulty using symbolic representations in abstract thought; thus they are less able to logically think through hypothetical situations. Physical involvement makes the information and skills being taught more concrete and salient for the children. Additionally, the use of rehearsal and reinforcement in the behavioral skills training may increase the motivation of children to learn the information provided and increase the likelihood that children will use the skills in the future. These results are consistent with meta-analyses in the child treatment literature that have found greater effect sizes for behavioral techniques than for nonbehavioral techniques in treating children and adolescents (Kazdin, 1994).

Number of Program Sessions

Studies of programs with more than three sessions found effect sizes from two to three times those of programs presented in one, two, or three sessions. This effect does not seem to be due to increased time spent learning about prevention, as the number of total hours spent in the program did not correlate significantly with effect size. Furthermore, the number of sessions was significantly correlated with effect size even after controlling for the variance accounted for by number of hours. This result may be explained by the *spacing effect*, which is the tendency for spaced presentations of information to produce greater learning than massed presentations of the same total length (Dempster & Farris, 1990). Perhaps dividing a program into more segments, thereby shortening the length of each, allows children to maintain their attention for the entire period and increases the amount of repetition of the material, thus leading to greater retention of the material.

Use of Puppet Shows

A final program characteristic that emerged as a significant moderator was the use of puppet shows. Interestingly, the direction of this effect was dependent on the presence of the Nemerofsky et al. (1994) study. With this study, programs with puppet shows had a lower mean effect size than programs without them; without the study, programs with puppet shows had a higher mean effect size. The mutability of this finding may be influenced by the small number of studies that used puppet shows, thus the significance of this variable should be viewed with caution.

Type of Outcome Measure

Along with identifying important program characteristics, the results of this meta-analysis suggested that type of outcome measure is related to effect size. Specifically, the use of behavioral observation to measure outcome significantly predicted higher effect sizes. Only two studies using behavioral outcome measures were included in the analysis, however, so the generalizability of these results is questionable, and conclusions must be tentative.

Higher effect sizes found using a behavioral outcome measure may be the result of several factors related to both the measure itself and the type of program for which it was used. Behavioral outcome measures involved in vivo simulation of a potentially dangerous stranger situation. Some time following program presentation, participants were approached by a stranger and asked to accompany this person to a more isolated location. Typically, this outcome measure was scored

Journal of Clinical Child Psychology 2000.29:257-265.

as simply a pass if the child did not agree to go with the stranger and a fail if the child did agree to go. This type of outcome measure had very few items and tested a narrow range of learning for which children were specifically trained. This may be responsible for the increased effect sizes found using this type of measure.

Programs evaluated using a behavioral measure differed from other programs as well. These programs typically focused on strangers as potential perpetrators and taught specific rules for dealing with strangers. This concrete approach to prevention may be easier for a child to learn than abstract concepts concerning known-adult abuse. Wurtele, Kast, Miller-Perrin, and Kondrick (1989) found that children taught specific rules concerning touching were more successful in distinguishing appropriate and inappropriate touches than were children taught the more abstract touch continuum. Additionally, situations with strangers are less ambiguous than those with known adults. The majority of abuse, however, is perpetrated by known adults.

Age of Participants

A final component important in predicting program effect size was child age. The relation was such that studies with older children had the lowest effect sizes. This did not support the hypothesis that older children would learn the most from prevention programs (e.g., Blumberg et al., 1991). Low effect sizes obtained for older children may have resulted from ceiling effects; this older group may have already known some prevention concepts and therefore been able to perform well on the outcome measure. Alternately, the lowered effect sizes found in studies evaluating older children may be due to the types of programs these children received. Studies evaluating children with a mean age higher than 8 found the lowest results but were also less likely to include active participation, behavioral skills training, and more than three sessions of instruction. These results imply that future programs for older children should alter format to match that of programs for the younger children. Importantly, the finding that younger children show more gains following prevention programs is consistent with the results of the Rispens et al. (1997) meta-analysis. In that analysis, however, effect size discrepancies were found to disappear at follow-up evaluation, indicating that younger children do not retain the learned information as well as older children.

Publication Bias

A publication bias was found in this body of literature, but it was in the opposite direction of what is typical. Rosenthal (1991) reported that dissertations and theses yielded effect sizes 40% smaller than those of other studies. In the analysis reported here, however, unpublished studies had a mean effect size 25% larger than that of studies that were published. A likely explanation for the high effect sizes of unpublished studies is the intercorrelation found between publication status and methodological quality. Specifically, unpublished studies had poorer designs, which may explain their status as unpublished as well as their inflated effect sizes.

Future Directions

Although this analysis has focused on the positive outcomes of prevention programs, it is important to note the possibility of negative reactions to the abuse-related material. In surveys, a moderate proportion of children and parents have reported increases in child anxiety and fear of adults (Finkelhor & Dziuba-Leatherman, 1995). Interestingly, parents who rated their children as more worried following prevention programs also rated the programs themselves as more helpful; perhaps increased anxiety is seen as effective in helping children to avoid victimization. Future meta-analyses should examine the evidence for negative consequences of prevention programs.

Possible limitations of this study include small sample size and inclusion of research with poor methodologies. Sample size was a problem in this meta-analysis not only because the area of child abuse prevention is relatively new, but because many of the evaluation studies that exist were unusable. A primary problem in this literature is the lack of control groups in much of the research; approximately 25 additional studies were located that could not be used in this meta-analysis because they did not have control groups. Additionally, in the analysis of moderators, the sample size decreased even further due to missing information; many studies failed to report information critical to literature integration such as cell sizes, design characteristics, and specific program content. This lack of information made it even more difficult to compare the already diverse array of existing prevention programs.

This meta-analysis incorporated all studies meeting inclusion criteria regardless of methodological quality so that all available information pertaining to program effectiveness could be utilized. This may have resulted in a somewhat inflated mean effect size, as studies with poorer methodological quality were found to have larger effect sizes. It was difficult to justify removing studies due to methodological quality, however, as the large majority of the studies were rated poorly on at least one methodological variable. Future evaluations should use uninformed interviewers, active control groups, and random assignment of participants (not just classrooms).

Journal of Clinical Child Psychology 2000.29:257-265.

A shortcoming of all the evaluation studies utilized in this meta-analysis is that they did not objectively assess whether abuse prevention was actually achieved. The majority of the investigations evaluated program effectiveness using attitude and knowledge scales. Indeed, the effect size calculated in this meta-analysis can only provide a measure of knowledge and skill gains in program participants. Although the effect size is large, it does not indicate that program participants are less likely to be sexually abused. We cannot assume that knowledge gains will translate into behavioral differences. If programs are to achieve their goal of prevention, individuals' behavior when confronted with a possibly abusive situation must change.

To address the issue of behavioral change, some programs assess skill gains by having children respond to touch discrimination vignettes or participate in role plays (e.g., Blumberg et al., 1991; Hazzard, Webb, Kleemeier, Angert, & Pohl, 1991) and, thus, have documented that children can learn several strategies to use when in dangerous situations. Additionally, to determine actual behavior change in dangerous situations, some researchers have used in vivo simulations with strangers and judged treatment effectiveness based on children's responses to the stranger (Fryer, Kraizer, & Miyoshi, 1987b; Poche, Yoder, & Miltenberger, 1988). These studies found that children's risky behavior was significantly diminished following program participation. Survey data have found evidence that material learned in prevention programs is used in response to actual threats in children's lives (Finkelhor et al., 1995). Children who have participated in prevention programs report using primary prevention strategies when confronted by the threat of victimization.

These studies reflect efforts to document behavioral changes resulting from prevention program participation. This work provides a valuable contribution to the literature on program effectiveness, but the crucial question remains unanswered: Do prevention programs actually reduce the prevalence of child sexual abuse? Future studies must further these efforts to examine the practical benefits of prevention programs. Although the high overall effect size found in this meta-analysis is encouraging, it is essential that existing programs begin to document their impact on the actual rates of sexual victimization.

References

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

*Anderson-Varney, T. (1988). *The evaluation of a personal safety curriculum for preschoolers.* Unpublished master's thesis, Michigan State University, East Lansing.

Berrick, J. D., & Barth, R. P. (1992). Child sexual abuse prevention: Research review and recommendations. *Social Work Research and Abstracts, 28,* 6–15.

Binder, R., & McNiel, D. (1987). Evaluation of a school-based sexual abuse prevention program: Cognitive and emotional effects. *Child Abuse & Neglect, 11,* 497–506.

Blumberg, E., Chadwick, M., Fogarty, L., Speth, T., & Chadwick, D. (1991). The touch discrimination component of sexual abuse prevention training. *Journal of Interpersonal Violence, 6,* 12–28.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Conte, J., Rosen, C., Saperstein, L., & Shermack, R. (1985). An evaluation of a program to prevent the sexual victimization of young children. *Child Abuse & Neglect, 9,* 319–328.

*Crowley, K. (1989). *Evaluation of good touches/bad touches: A program to prevent child sexual abuse in school-age children.* Unpublished doctoral dissertation, Pace University, New York.

Daro, D. (1994). Prevention of child sexual abuse. *The Future of Children, 4,* 198–223.

Davis, M. K. (1997). *Child sexual assault prevention programs: A meta-analysis.* Unpublished master's thesis, Ohio University, Athens.

Dempster, F., & Farris, R. (1990). The spacing effect: Research and practice. *Journal of Research and Development in Education, 23,* 97–101.

*Dhooper, S., & Schneider, P. (1995). Evaluation of a school-based child abuse prevention program. *Research on Social Work Practice, 5,* 36–46.

Downer, A. (1984). *Evaluation of talking about touching.* Seattle, WA: Committee for Children.

Finkelhor, D., Asdigian, N., & Dziuba-Leatherman, J. (1995). The effectiveness of victimization prevention instruction: An evaluation of children's responses to actual threats and assaults. *Child Abuse & Neglect, 19,* 141–153.

Finkelhor, D., & Dziuba-Leatherman, J. (1995). Victimization prevention programs: A national survey of children's exposure and reactions. *Child Abuse & Neglect, 19,* 129–139.

*Fryer, G., Kraizer, S., & Miyoshi, T. (1987a). Measuring actual reduction of risk to child abuse: A new approach. *Child Abuse & Neglect, 11,* 173–179.

Fryer, G., Kraizer, S., & Miyoshi, T. (1987b). Measuring children's retention of skills to resist stranger abduction: Use of the simulation technique. *Child Abuse & Neglect, 11,* 181–185.

*Harvey, P., Forehand, R., Brown, C., & Holmes, T. (1988). The prevention of sexual abuse: Examination of the effectiveness of a program with kindergarten-age children. *Behavior Therapy, 19,* 429–435.

Hazzard, A., Webb, C., Kleemeier, C., Angert, L., & Pohl, J. (1991). Child sexual abuse prevention: Evaluation and one-year follow-up. *Child Abuse & Neglect, 15,* 123–138.

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis.* New York: Academic.

*Herman, P. (1987). *Evaluation of a sexual abuse curriculum/prevention program.* Unpublished doctoral dissertation, Washington State University, Seattle.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis.* Newbury Park, CA: Sage.

Kazdin, A. (1994). Psychotherapy for children and adolescents. In A. Bergin & S. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (pp. 543–594). New York: Wiley.

*Kolko, D., Moser, J., & Hughes, J. (1989). Classroom training in sexual victimization awareness and prevention skills: An extension of the red flag/green flag people program. *Journal of Family Violence, 4,* 25–45.

*Kolko, D., Moser, J., Litz, J., & Hughes, J. (1987). Promoting awareness and prevention of child sexual victimization using

Journal of Clinical Child Psychology 2000.29:257-265.

the red flag/green flag program: An evaluation with follow-up. *Journal of Family Violence, 2,* 11–35.

*Kraizer, S. (1991). *The Safe Child Program for the prevention of child abuse: Development and evaluation of a school-based curriculum.* Unpublished doctoral dissertation, The Union Institution, Cincinnati, Ohio.

*Nelson, D. (1985). *An evaluation of the student outcomes and instructional characteristics of the "You're in charge" program.* Unpublished manuscript, Utah State Office of Education, Salt Lake City.

Nemerofsky, A., Carran, D., & Rosenberg, L. (1994). Age variation in performance among preschool children in a sexual abuse prevention program. *Journal of Child Sexual Abuse, 3,* 85–102.

*Oldfield, D., Hays, B., & Megel, M. (1996). Evaluation of the effectiveness of Project Trust: An elementary school-based victimization prevention strategy. *Child Abuse & Neglect, 20,* 821–832.

*Peraino, J. (1990). Evaluation of a preschool antivictimization prevention program. *Journal of Interpersonal Violence, 5,* 520–528.

*Poche, C., Yoder, P., & Miltenberger, R. (1988). Teaching self-protection to children using television techniques. *Journal of Applied Behavior Analysis, 21,* 253–261.

*Ratto, R., & Bogat, G. A. (1990). An evaluation of a preschool curriculum to educate children in the prevention of sexual abuse. *Journal of Community Psychology, 18,* 289–297.

Rispens, J., Aleman, A., & Goudena, P. (1997). Prevention of a child sexual abuse victimization: A meta-analysis of school programs. *Child Abuse and Neglect, 21,* 975–987.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (rev. ed.). Newbury Park, CA: Sage.

Sandberg, D., Lynn, S., & Green, J. (1994). Sexual abuse and revictimization: Mastery, dysfunctional learning, and dissociation. In S. Lynn & J. Rhue (Eds.), *Dissociation: Clinical and theoretical perspectives* (pp. 242–267). New York: Guilford.

Saslawsky, D., & Wurtele, S. (1986). Educating children about sexual abuse: Implications for pediatric intervention and possible prevention. *Journal of Pediatric Psychology, 11,* 235–245.

Tutty, L. (1992). The ability of elementary school children to learn child sexual abuse prevention concepts. *Child Abuse & Neglect, 16,* 369–384.

*Wolfe, D., MacPherson, T., Blount, R., & Wolfe, V. (1986). Evaluation of a brief intervention for educating school children in awareness of physical and sexual abuse. *Child Abuse & Neglect, 10,* 85–92.

*Wurtele, S. (1990). Teaching personal safety skills to four-year-old children: A behavioral approach. *Behavior Therapy, 21,* 25–32.

*Wurtele, S., Gillispie, E., Currier, L., & Franklin, C. (1992). A comparison of teachers vs. parents as instructors of a personal safety program for preschoolers. *Child Abuse & Neglect, 16,* 127–137.

*Wurtele, S., Kast, L., & Melzer, A. (1992). Sexual abuse prevention education for young children: A comparison of teachers and parents as instructors. *Child Abuse & Neglect, 16,* 865–876.

Wurtele, S., Kast, L., Miller-Perrin, C., & Kondrick, P. (1989). A comparison of programs for teaching personal safety skills to preschoolers. *Journal of Consulting and Clinical Psychology, 57,* 505–511.

Wurtele, S., Marrs, S., & Miller-Perrin, C. (1987). Practice makes perfect? The role of participant modeling in sexual abuse prevention programs. *Journal of Consulting and Clinical Psychology, 55,* 599–602.

Wurtele, S., Saslawsky, D., Miller, C., Marrs, S., & Britcher, J. (1986). Teaching personal safety skills for potential prevention of sexual abuse: A comparison of treatments. *Journal of Consulting and Clinical Psychology, 54,* 688–692.

Manuscript received October 23, 1997
Final revision received July 6, 1999

Journal of Clinical Child Psychology 2000.29:257-265.