



Published in final edited form as:

J Consult Clin Psychol. 2005 December ; 73(6): 1036–1044. doi:10.1037/0022-006X.73.6.1036.

Iatrogenic Effects of Group Treatment for Antisocial Youth

Bahr Weiss,

Vanderbilt University

Annalise Caron,

New York State Psychiatric Institute

Shelly Ball,

Vanderbilt University

Julie Tapp,

Vanderbilt University

Margaret Johnson, and

Vanderbilt University

John R. Weisz

Judge Baker Children's Center, Harvard University

Abstract

It has been argued that group treatment of antisocial adolescents may increase rather than decrease conduct problems. One mechanism that has been suggested to underlie this effect is “deviancy training” wherein during group sessions deviant peers reinforce each other’s antisocial actions and words. These two hypotheses have important implications, and warrant close review at conceptual and empirical levels. In this paper we present such a review. Conceptually, deviancy training potential of treatment sessions appears less significant than the more extensive peer influences outside treatment. Empirical findings previously cited in support of iatrogenic effects appear on close examination to provide little support. Finally, seventeen of eighteen new meta-analytic tests produced results not supportive of iatrogenic or deviancy training effects.

Concerns recently have been increasingly raised (e.g., Arnold & Hughes, 1999) that group interventions with antisocial adolescents may increase rather than decrease antisocial behavior, with a number of potential mechanisms underlying such iatrogenic effects. For instance, association with antisocial youth in group treatment may be stigmatizing, leading to increases in antisocial behavior through influences on self-concept, self-expectations, and personal norms (e.g., De-Haan & MacDermid, 1999). One potential mechanism relatively frequently raised involves “deviancy training” wherein during group sessions, through verbal and non-verbal communication deviant peers may positively reinforce each others’ antisocial behavior, increasing the likelihood of future delinquent acts (Dishion, McCord & Poulin, 1999).

Such concerns about aggregating deviant peers have been raised in the past (e.g., Fo & O'Donnell, 1975) but an unusually thoughtful description of deviancy training's potential iatrogenic effects has been provided by Dishion et al. (1999), who suggested several conditions that may exacerbate such effects. First, effects may be stronger during early adolescence, a period when many individuals' interpersonal focus shifts from family to peers. Second, deviancy training effects during group treatment may be most likely to occur when it is antisocial youth who are aggregated as opposed to, say, depressed youth. Third, deviancy training may be strongest when groups are homogeneous, containing only antisocial adolescents. Inclusion of pro-social youth who attend to more normative behavior should decrease reinforcement for deviant talk and actions. Fourth, Dodge (1999) has suggested that relatively unstructured treatment groups, with relatively little control over the direction of the session, may be more supportive of deviancy training effects. Finally, Arnold and Hughes (1999) have suggested that youth without a history of serious antisocial behavior may be particularly vulnerable to be influenced by deviant peers in group treatment.

In the present paper we conceptually and empirically re-examine two aspects of these hypotheses: (a) that group treatments are iatrogenic, and (b) that deviancy training effects underlie these iatrogenic effects. Because these two hypotheses are closely linked, in this paper we review them concurrently. We first emphasize three points: (a) in our discussion we are considering deviancy training effects as they might occur during group treatment sessions, not as they might occur in other contexts. The evidence for deviancy training effects in general is convincing (Brendgen, Vitaro & Bukowski, 2000); (b) we are not arguing that deviancy training cannot occur during group treatment, but rather that the potential effects of deviancy training during group interventions may be limited, relative to what goes on in neighborhoods, classrooms, etc.; (c) although there are a number of potential mechanisms through which adolescent group treatments could become iatrogenic, in our review we focus on deviancy training because it is the mechanism that has been discussed most frequently in the literature, and because its effects are well established outside of therapeutic contexts (Dishion et al., 1999).

Conceptual Issues

What we see as a fundamental issue in regards to deviancy training's potential iatrogenic effects is that the reinforcement adolescents potentially may receive for antisocial actions during group treatment may represent only a relatively small increment of peer reinforcement for deviant behavior over and above 'deviancy training' occurring outside treatment. That is, (a) adolescents selected to participate in prevention or treatment groups for delinquent behavior likely are already associating with delinquent youth outside of treatment, or are at high risk for such associations, independent of involvement in the group treatment; and (b) the amount of time spent in treatment sessions represents a small portion of their waking hours.

In regards to the first point, many of the risk factors for delinquency (e.g., having a dysfunctional family) also are risk factors for negative peer relationships (e.g., Fergusson & Horwood, 1999), and negative peer associations are a strong predictor of delinquency (e.g.,

Elliot & Menard, 1996). Thus, if subjects are selected based on the (a) presence of or (b) risk for delinquent behavior, they also likely are (a) already associating with or (b) at risk for future association with deviant peers, irrespective of their participation in treatment. The treatment group may well foster friendships outside the therapy group, but insofar as youth are selected for inclusion into the group treatment based on high risk status, they are likely to associate with delinquent youth in the future, regardless of their participation in the treatment group. In regards to the second point, the majority of adolescents' free time is spent in unstructured activities (Larson, 2001), with one-third of their waking hours spent interacting with peers, providing substantial opportunities for associating with deviant peers (Tarter, 2002). Thus, the amount of time spent in therapy sessions represents a small portion of adolescents' waking hours, providing a relatively small increment for deviancy training.

One might respond to this position by suggesting that it implies that any form of therapy should not have much of an impact on clients' lives since for most forms of therapy, sessions consume only a small portion of clients' time. However, our point is not simply that the time antisocial adolescents spend in therapy sessions represents a small portion of their lives but rather that it represents a small portion of the total time potentially spent in deviancy training. That is, we have posited that adolescents likely spend much more time outside of therapy engaging in the causal activity of interest here, deviancy training, than they do in treatment sessions. In contrast, in regards to therapy more generally, the crucial activities that occur in it are in some sense unique and unlikely to occur outside of the therapy sessions, and they thus represent an addition, hopefully positive, that can impact on clients' lives.

Time spent in a treatment group does differ from time spent in other settings in a number of ways. On the one hand, the likelihood of adults curtailing or challenging deviant actions during groups is high, almost certainly higher than in less therapeutic contexts. On the other hand, the fact that these antisocial discussions or subtle antisocial behavior occur in the context of an adult-sanctioned event could give the reinforcement or other effects more power. For instance, adults could serve as a catalyst for the development of an especially strong antisocial group identity, by providing a direct 'target' against which the antisocial adolescents could demonstrate to their peers their antisocial values. The sense of united overt opposition to adult authority could create a stronger sense of allegiance than would occur outside of therapy. To the best of our knowledge, although there are anecdotal reports about such processes, there has been no direct research comparing the strength of reinforcement or antisocial identity formation within versus outside of treatment groups. This is a significant gap in the literature base, and an important area for future research.

Empirical Issues

Studies Cited as Evidence of Iatrogenic Effects

We next discuss several studies that various authors (e.g., Arnold & Hughes, 1999; Dishion et al., 1999; McCord, Widom & Crowell, 2001) have suggested provide evidence that (a) adolescent peer group treatments may be iatrogenic, and / or (b) that deviancy training may underlie these effects. We do not suggest that these studies are particularly flawed; treatment research is a difficult enterprise and shortcomings can be found in any study, including our

own. Rather, the purpose of this discussion is to consider whether the results from these studies supports these conclusions, or whether there are other equally plausible interpretations of the results. Because of space limitations, we only discuss aspects of the studies most relevant to our questions.

Adolescent Transitions Program—The Adolescent Transitions Program (Dishion & Andrews, 1995; reviewed by Dishion et al., 1995) contained three group treatments (Parent group only; Adolescent group only; Parent group and adolescent group) utilizing therapy approaches generally considered to be efficacious (Dishion & Andrews, 1995; Kazdin & Weisz, 1998). Post-treatment assessments showed a few positive treatment effects (e.g., negative parent-child interactions decreased) but for teacher-reported externalizing problems only the “Parent only” condition showed a treatment effect, which was marginally significant, and no significant intervention effects were reported for parent-report externalizing problems (Dishion & Andrews, 1995) for any condition. Two apparent iatrogenic treatment effects for self-reported smoking of tobacco were found at post-treatment, with the “Adolescent only,” and “Parent and adolescent” groups showing increased rates of self-reported smoking relative to the control groups; the “Parent only” condition did not differ from the control groups. However, carbon monoxide breath assessments, generally considered a more reliable indicator of smoking than self-reports (e.g., Wills & Cleary, 1997), were in the same direction as self-reports but were not significant (personal communication, T. Dishion, April 7, 2003).

It was at the one year follow-up that apparent iatrogenic treatment effects became most evident. For the “Adolescent only” group there was a significant iatrogenic effect for teacher reported externalizing problems and for the “Parent and adolescent” group there was a marginally significant iatrogenic effect; there were no significant treatment effects on parent-reported externalizing problems for any of the three groups (Dishion & Andrews, 1995). At the one-year follow-up, the iatrogenic effect on self-reported smoking remained significant for the “Parent and adolescent” but not the “Adolescent only” condition; group differences on CO breath assessments were not significant (personal communication, T. Dishion, April 7, 2003).

Somewhat similar results were found at the two and three year follow-up (Poulin, Dishion & Burraston, 2001), for which the “Adolescent only” and “Parent and adolescent” groups were combined into one group. With teacher-reported delinquency as the dependent variable, there was no treatment effect for the combined group at post-treatment, a significant iatrogenic treatment effect at the one-year follow-up, and marginally significant ($p < .10$) iatrogenic treatment effects at the two- and three-year follow-ups. For self-reported smoking as the dependent variable, iatrogenic treatment effects were significant at post-treatment and the one- and three-year follow-ups, and marginally significant at the two-year follow-up. Parent report data were not reported, nor were CO smoking data.

These various results were interpreted as supportive of an iatrogenic treatment effect for the adolescent peer group component, with deviancy training the likely mechanism (Dishion et al., 1999). We believe that there are a number of aspects of the results that are not supportive of these interpretations. First, support for the iatrogenic treatment effect is based on a

number of marginally significant effects, which is particularly problematic given the tests evaluated a post-hoc hypothesis generated from the data. Second, the “Parent-only” condition showed no significant treatment effects on parent-reported externalizing problems at either post-treatment or follow-up, and the effect of the “Parent-only” condition on teacher-reported externalizing problems was reported as marginally significant at post-treatment and non-significant at the one year follow-up; parent data for two and three year follow-ups were not reported. As the authors note in their introduction (Dishion & Andrews, 1995), and as other authors have concluded (e.g., Kazdin & Weisz, 1998), parent-training interventions such as implemented here generally are effective. These are experienced, skilled clinical investigators and it is unclear why the “Parentonly” intervention showed limited effectiveness at best. Perhaps there were problems with the control groups; the original control group was supplemented because of concerns that this control condition (which involved access to treatment materials and “self-study”) may have actually been an effective treatment whereas the second control condition was not randomly assigned. Regardless, at the very least, the limited effectiveness of the “Parent-only” treatment suggests that based on the results of this study, conclusions regarding iatrogenic treatment effects may be limited to non-efficacious treatments.

Finally, in regard to the iatrogenic effect on smoking, the breath CO data, generally taken as a measure free of self-report bias (e.g., Wills & Cleary, 1997), trended in the same direction as the self-report data but did not show significant effects. The self-report findings, which represented the strongest iatrogenic treatment effect, were not verified by significant effects for the more valid breath CO data. Overall, then, taken from a different perspective, we would interpret this study as not providing strong evidence of iatrogenic treatment effects, particularly given the number of marginally significant, post-hoc effects upon which the conclusions rest.

Experiment in Juvenile Court study—A study cited by McCord et al. (2001) as “... grounds for believing that in some circumstances such settings [i.e., peer group-based interventions] may exacerbate problem behaviors among young adolescents” (p. 135) was the “Experiment in Juvenile Court” study. Berger, Crowley, Gold, Gray and Arnold (1975) evaluated a juvenile court program wherein adjudicated adolescents were randomly assigned to: (a) a “volunteer probation officer” (VPO) program, wherein the volunteer’s goal was to develop a healthy, adaptive relationship with the delinquent adolescent; (b) group counseling, led by a volunteer; (c) academic tutoring, provided by a volunteer; or (d) a probation services-as-usual control group. There were four primary outcome measures: (a) frequency of delinquent acts based on self-report (F-SR); (b) severity of delinquent acts based on self-report; (c) frequency of police contact (FPC); and (c) extent of involvement with the court system (ranging from no involvement to incarcerated). There were three main effects for type of program, all negative: (a) the Tutoring group deteriorated relative to the Control group on F-SR; and (b) the VPO group and (c) the Group Counseling group deteriorated relative to the control on FPC.

It is clear that the results provide no support for the efficacy of the volunteer programs. However, the results also provide little support for the hypothesis that peer-group based interventions result in negative outcomes. Of the three programs only one, the Group

Counseling condition, involved aggregation of antisocial youth. Yet for this condition (a) only 1 of 4 comparisons showed a significant effect, and (b) the non-aggregating conditions showed the same number of negative effects, which suggests that volunteers in general may be ineffective or harmful with antisocial adolescents. Even if Group Counseling produced negative outcomes, it would be inappropriate to consider this an iatrogenic treatment effect. Rather, the outcome in this condition should be viewed as indicating that when (a) volunteers with no formal training (b) receive little training in regards to counseling interventions (two meetings providing information on the juvenile court and its services, juvenile laws, and the desirable kinds of relationships between volunteers and the adolescents) and (c) attempt to implement group counseling with antisocial youth and (d) receive minimal supervision during the process (monthly in-service meetings that discussed child abuse, handling drug overdoses, transactional analysis, and questions raised by the volunteers), they not surprisingly may produce null or negative effects.

Cambridge-Somerville Youth Study—The Cambridge-Somerville Youth Study Evaluation (e.g., McCord, 1992) has been cited relatively frequently as an example of iatrogenic treatment effects (e.g., Arnold & Hughes, 1999), and was reviewed in detail by Dishion et al. (1999). In this study, begun in 1939, pairs of “difficult” and “average” children were matched on a large set of variables, and randomly assigned within pairs to treatment or control conditions. Treatment was individualized, with each child receiving different mixtures of mentoring, academic tutoring, “psychiatric attention,” etc. Mentors encouraged the boys to participate in pro-social community activities such as sports or getting a job. An evaluation shortly after the end of treatment found no differences between the treatment and control groups.

Approximately thirty years after the end of treatment, a follow-up study (McCord, 1978) was conducted that produced a number of findings that have been interpreted as support for iatrogenic treatment effects of adolescent group treatment. The first was that at this follow-up, a higher than chance proportion of children assigned to the treatment condition had “undesirable outcomes,” which Dishion et al. (1999) interpret as support for an iatrogenic group treatment effect. Although this may represent a negative effect of assignment to the treatment condition, we believe that it provides little or no support for either the hypothesis that group treatment is iatrogenic or that deviancy training underlies the iatrogenic effects. First, very little of the intervention these youth received involved group activities; Dishion et al. (1999) note that “the iatrogenic effects of the CSYS program do not appear to be attributable to an emphasis on encouraging boys to participate in group activities” (p. 760). Second, most of the “treatment” in this study appeared to consist of mentoring and case management, which by current standards probably would not be considered a potentially efficacious treatment (Kazdin & Weisz, 1998); thus, this finding could represent negative effects of aggregating adolescents for non-treatment purposes. Third, included among the “undesirable outcomes” used as the dependent variable were schizophrenia and bipolar disorder, which do not seem likely to be influenced by group treatment or deviancy training effects; this latter issue could be resolved through re-analysis.

A second finding that Dishion et al. interpret as evidence of an iatrogenic treatment effect is that boys in the treatment condition who received the most attention were most likely to

show negative effects, which suggests that increased therapeutic attention was iatrogenic. However, because boys in the study were not randomly assigned to the amount of attention they received, this finding could reflect the tendency for boys showing the least improvement to receive the most attention from the mentors, etc. Receiving more attention may have been a response to higher levels of difficulty and a failure of treatment, rather than vice-versa. Similarly, Dishion et al. (1999) cite as support for iatrogenic treatment effects the finding that negative outcomes occurred primarily among “cooperative” families. Again, it is possible that the most cooperative families were those with continuing child problems, and cooperativeness with treatment thus may have been a response to difficulties their child was having rather than vice versa.

Finally, Dishion et al. (1999) cite as support for the hypothesis that deviancy training underlaid these iatrogenic treatment effects the finding that treatment group adolescents assigned to summer camp by their counselors had worse outcomes than the matched control group participants. They reasoned that “participation in these camps could permit the type of audience and selective attention for misbehavior” (p. 760) that could result in deviancy training. For several reasons we think that this finding provides limited support for iatrogenic group treatment effects or deviancy training. First, as these authors note, the camps were “not dominated by participation of high-risk youth” (p. 760), which contradicts the hypothesis that deviancy training effects are most likely when antisocial youth are concentrated. Second, even if there was aggregation of delinquent youth that resulted in negative outcomes via deviancy training, this probably should not be considered an “iatrogenic” effect. “Iatrogenic” is variously defined as “induced inadvertently by a physician or surgeon or by medical treatment or diagnostic procedures” (Merriam-Webster Dictionary, n.d.) and thus, since these summer camps do not appear to have been designed for treatment or diagnostic purposes, they could not be iatrogenic.

Third, although children were randomly assigned within matched pairs to treatment or control, assignment to summer camp was not random but at the discretion of the counselor. Hence, the equating effects of random assignment within pairs were lost. Given the inevitable random variability in functioning of participants even within the carefully matched pairs (i.e., even though participants were well matched, there would still be random variability in their functioning, since the matching could not account for 100% of variability in negative outcomes), it is possible that counselors’ decision to recommend a particular child to go to camp was based on the child’s functioning poorly. Thus, the fact that children who went to camp did worse than their matched pairs who did not go to camp could simply reflect counselors having picked out the worst functioning children to send to camp. Dishion et al. did test whether children who went to camp differed from children who were not sent to camp in regards to delinquency prediction ratings, and found that they did not differ significantly. However, these “delinquency prediction scores” were based on each boy’s family history and home environments (McCord, 1992), so this comparison does not rule out the possibility that it was negative outcomes and the boys’ functioning during treatment that resulted in assignment to the summer camps.

Perhaps most critically, iatrogenic effects were not found at the end of treatment (which lasted five to six years), nor were they found at a follow-up evaluation five to ten years later

(McCord, 1990), but only after a thirty year delay. While it is possible that causal effects might be enduring, few if any treatment studies have shown delayed effects to be stronger than immediate effects (Weisz et al., 1987; 1995). One might hypothesize that the summer camps facilitated development of friendships that later led to deviancy training (Arnold & Hughes, 1999). However, the post-treatment follow-up occurred five to six years after treatment began, which would seem a reasonable length of time for deviancy training to have occurred.

Guided Group Interaction—Guided Group Interaction (GGI) is a group therapy approach that posits that delinquents will learn to conform to conventional social rules by receiving more social rewards through conformity than they do from nonconformity. Group sessions involve open discussion with an adult leader guiding the group by asking questions, summarizing, and encouraging participants to recognize problems with their behavior and attitudes. Gottfredson (1987) conducted a review of five evaluations of GGI. McCord et al. (2001) summarized his results by stating “Guided Group Interactions tended to increase misbehavior and delinquency” (p.135), and presented this review as evidence that in some circumstances peer group-based interventions may worsen young adolescents’ conduct problems.

We disagree that Gottfredson’s (1987) review provides evidence that peer group-based interventions are iatrogenic. Because of space limitations our review of Gottfredson (1987) is focuses on his conclusions and any significant limitations placed on his conclusions. The first study reviewed was the “Highfields Project” wherein adjudicated boys were assigned either to the Highfields residential facility or were incarcerated at the state reformatory. The experimental group showed lower rates of recidivism than the control group but because assignment to condition by judges was non-random, Gottfredson (1987) concluded that “the outcome of the Highfields study is best regarding as promising” (p. 675) rather than definitive.

A second GGI evaluation was the “Provo Experiment.” Adolescents who had been assigned to either probation or incarceration by a judge were then randomly assigned by the project to the option selected by the judge, or to the experimental GGI group. Certain design features made interpretation of results complex. For instance, outcome assessments occurred during and after treatment, and adjustments had to be made for the fact that the incarcerated youth had fewer opportunities to get arrested during their time in ‘treatment’ (i.e., while they were incarcerated). Ultimately, Gottfredson (1987) concluded “the practical implication of the results is that the experimental treatment resulted in fewer arrests for both the probation and incarceration populations” (p. 683), but because a “nonequivalent comparison” group had to be constructed for the incarceration group (because judges assigned few youth to incarceration), results for the incarceration group were less definitive than for the probation group.

The third GGI study reviewed was the Silverlake Experiment. Repeat male offenders were “usually” randomly assigned to (a) a community-based GGI program, or (b) a more traditional, institutional program. During the year after completing treatment, the experimental group had a mean number of arrests of .73 per adolescent vs. the control

group .80; this difference did not approach significance. However, the GGI-based treatment was more cost-efficient, costing 38% per adolescent of the cost of the traditional treatment.

Next reviewed was the Essexfields study. Assignment to condition was non-random, with judges assigning adolescents with less extensive delinquent histories to probation, those with more extensive histories to Essexfields' GGI program, and adolescents with the most extensive histories to the reformatory. There were substantial pre-treatment differences between the conditions (e.g., adolescents in the probation condition had a mean number of pre-treatment court appearances of .9 vs. 2.3 for Essexfields vs. 3.0 for the reformatory). At the three year follow-up, the GGI and the reformatory groups showed similar rates of recidivism that were higher than the probation group, but this result is difficult to interpret as adolescents with more extensive delinquent histories were assigned to Essexfields and to the reformatory, as opposed to probation.

The final GGI study reviewed was the Collegefields study. Adolescents were randomly assigned to a GGI condition or to one of two control groups. Most outcome comparisons favored the GGI condition, but only one was significant. In sum then, of the five GGI studies reviewed by Gottfredson (1987), only one produced results suggesting negative effects of GGI, and that study was limited by substantial pre-treatment differences in delinquency histories for the groups. If one were willing to ignore non-random assignment and pre-treatment differences, then one would need also to accept the positive results of the Highfields Project and the Provo Experiment. Thus, we disagree with McCord et al. (2001) that Gottfredson's (1987) review indicates that GGI tends to increase misbehavior and delinquency.

Meta-analytic Results

The studies reviewed above are ones that have been cited and discussed in the literature as indicating iatrogenic and / or deviancy training effects for peer group interventions, but a broader sampling of treatment studies would be useful to consider. Consequently, the following sections consider relevant meta-analytic findings based on larger samples of studies.

Lipsey (1992)—Lipsey's (1992) meta-analysis of delinquency treatment studies found that 29% of its studies had negative effect sizes, which Poulin et al. (2001) used as evidence of an iatrogenic treatment process. However, as Lipsey notes, because an effect size is a random variable, random variation will occur around the population effect size. If the population effect size is positive but relatively close to zero, observed effect sizes will include negative effect sizes due to sampling variability, despite the fact that the treatment actually produces positive effects. In Lipsey's data the population effect appears to be positive: 64% of observed effect sizes were positive, 29% negative, with a mean between +.10 and +.20. (As Lipsey notes, an effect size of .10 may sound trivial, but it represents a 10% decrease in recidivism from a baseline of 50% which is not trivial). Thus, unless there is a moderator of treatment effects, the negative effect sizes reported in Lipsey's meta-analysis likely represent normal variation around a positive effect size, due to random sampling effects rather than an iatrogenic process.

But might there in fact be a moderator, such that the mean effect size is not an appropriate summary of all of the effect sizes, and hence the variation around the mean effect size represents not just random sampling variability but also differences in true effect sizes as a function of this moderator? If there were, then a single mean effect size cannot represent the effect sizes, and the negative effect sizes might represent variation around a true negative effect size for a subset of the studies. The most obvious moderator of treatment effects would be group vs. individual treatment. However, as Table 4.7 in Lipsey et al. (1992) suggests, treatments that had null or negative effects do not appear to be those in which peers were aggregated, nor do treatments that aggregated peers appear to be those associated with negative effect sizes. Group counseling was associated with positive effect sizes, and treatment modalities that produced mean negative effect sizes were vocational training and “deterrence” interventions, which simply may be incapable of modifying the causal and perpetuating factors underlying delinquency (Lundman, 2001).

Ang and Hughes (2002)—Ang and Hughes (2002) meta-analysis reviewed 38 studies focusing on social skills training interventions for antisocial youth. The comparison most relevant here was between treatments that (a) involved homogeneous groups of antisocial youth ($ES=.55$, $n=29$); (b) mixed groups containing antisocial and non-antisocial youth ($ES=.60$, $n=4$) and (c) individual treatment ($ES=.78$, $n=5$). Because of the small samples, the mixed groups and individual treatments were combined into a “non-aggregated” comparison group ($ES=.70$, $n=9$), which differed significantly from the homogeneous groups. Ang and Hughes (2002) suggested that this finding provides support for the hypothesis that that group treatments that aggregate antisocial youth may be iatrogenic, and that peer reinforcement and other deviancy training effects may be responsible.

We believe that these data are supportive of neither conclusion. Because (a) the mixed groups’ effect ($ES=.60$) is actually closer to the homogeneous groups’ effect size ($ES=.55$) than it is to the individual treatments’ effect size ($ES=.78$), and because (b) the mixed group’s sample size is much smaller than the aggregated group’s sample size (4 versus 29, respectively), the significance of aggregated vs. “non-aggregated” treatments comparison is actually a function of the difference between group versus individual treatments. The finding that group treatments produce smaller effect sizes than individual treatments has been reported previously and does not appear to be specific to conduct problems. In our meta-analyses (Weisz et al., 1987; 1995), group treatments were associated with smaller effects sizes than individual treatments in general, not just for conduct problems. That group treatments produce smaller effect sizes does not mean that they necessarily contain iatrogenic components; clearly, there are reasons (e.g., less time spent on any one individual) other than “deviancy training” or other iatrogenic effects why group treatments might be less effective than individual treatments.

The most relevant comparison, we believe, would be between groups that aggregated antisocial youth vs. groups that contained a mixture of antisocial and non-antisocial youth. As has been suggested (e.g., Ang & Hughes, 2002; Dishion et al., 1999), the presence of nonantisocial youth in mixed groups should control or attenuate deviancy training effects, through decreased reinforcement of deviant actions, and increased reinforcement of normative behavior. Thus, the homogeneous versus mixed groups comparison would be a

test of deviancy training effects during group treatment. In addition, in contrast to the group versus individual contrast, comparison of two group conditions would control for factors associated with group interventions, such as therapists' need to attend to multiple youth simultaneously. The difference in Ang and Hughes data between the effect sizes for the aggregated versus mixed groups was quite small (.05). Because the number of mixed groups in this comparison was small ($n=4$), it would be inappropriate to firmly conclude that aggregation of antisocial youth does not result in deviancy training effects; however, it would be even more inappropriate to conclude that these data suggest that aggregation of antisocial youth does result in deviancy training effects.

New Analyses in our Meta-analytic Datasets

Next, we used our own youth psychotherapy outcome meta-analytic dataset to address some of these questions. Because our datasets were not designed to evaluate potential iatrogenic group treatment or deviancy training effects, we coded several new variables, and we also updated the dataset by coding studies published since our most recent meta-analysis (Weisz et al., 1995), following guidelines used in the original meta-analyses (see Weisz et al., 1987; 1995). To be included, the focus of the treatment had to be on externalizing conduct problems, excluding treatments that focused exclusively on impulsivity / attentional problems without addressing aggression, etc. Studies with less than 20 participants were dropped, as their results were unlikely to be reliable.

Codings—New codings included: (a) whether treatment included a peer group treatment component (inter-rater $\kappa = .98$); (b) if the treatment did, whether the group was homogeneous in regards to conduct problems ($\kappa = .86$), or included non-antisocial youth; (c) number of children and adults in the treatment group (both $\kappa = 1.0$); and (d) intervention duration ($\kappa = .96$). In our analyses here, we also used previously coded variables (e.g., type of treatment).

Results—Because “iatrogenic” can refer to (a) a reduction in effectiveness or (b) negative effects, we conducted two sets of analyses. The first used a continuously distributed effect size to assess possible reduction in effect size, and the second dichotomized these effect sizes to either negative (reflecting an iatrogenic effect) or positive, analyzing these dichotomized effect sizes using a logit model (Agresti, 1990). We used weighted least squares, weighting by the inverse of the variance of the effect size (Hedges & Olkin, 1985). We report mean effect sizes and percentages in their original unweighted metric, as these will be more readily interpretable, but parenthetically note the statistics as they were analyzed (weighted). We considered any test with $p < .05$ significant, but also report marginally significant tests of $p < .10$.

Our two databases (Weisz et al., 1987; 1995) combined with the new studies produced a pool of 66 studies containing 115 separate treatment groups, 53% involving a peer group. The number children in the groups ranged from 3 to 20 (mean=8.4), the number of adults ranged from 1 to 10 (mean=1.8; the one study with 10 adults involved a treatment program with combined parent / child / therapist groups), and the adult / child ratio ranged from .1 to 1.3 (mean=.3). Ten percent ($n=6$) of groups were mixed in regards to children with conduct

problems, and most group treatments (87%) used behavioral techniques (generally cognitive-behavioral or social skills training). In the analyses that follow, we defined potential iatrogenic as an effect wherein treatments involving peer groups were associated with significantly smaller effects or a significantly greater likelihood of a negative effect size. Interaction tests evaluated whether “iatrogenic” effects might be stronger across levels of other variables (e.g., age).

We first assessed whether overall, treatments containing a peer group component were associated with smaller effect sizes or a higher probability of having a negative effect size. Treatments that did not have a peer group component had a mean effect size of .68 (.48) and a .12 (.19) likelihood of having a negative mean effect size, whereas treatments that included a peer component had a mean effect size of .79 (.34) and a .08 (.12) likelihood of a negative effect size. The test of the continuous effect sizes was non-significant whereas the test of the dichotomized effect sizes was significant ($p < .005$), indicating that the presence of a peer group component was associated with significantly *less* likelihood of having a negative effect size.

Since it has been suggested that iatrogenic effects may be more likely to occur among early adolescents, we tested the interaction between Peer Group and the curvilinear effect of age. The interaction was non-significant for the continuous effect sizes and marginally significant for the dichotomous effect sizes. For treatments with a peer group, the log odds of having a negative effect size peaked at age 11 whereas for treatments without a peer group, it peaked at age 8.6.

To determine if iatrogenic effects might vary as a function of type of treatment (reflecting degree of structure), we tested the interaction between type of treatment (a pre-existing coding, cognitive / behavioral vs. non-behavioral treatments) and Peer Group. Both tests were nonsignificant. We tested whether the sub-type of externalizing problems (e.g., aggression; delinquency) being treated influenced the effect of peer group on treatment outcome. Again, this interaction was non-significant for continuous and dichotomous outcomes. The next analysis focused on whether there was an interaction between the total duration of treatment and the effect of Peer Group, to determine whether peer group treatments that were longer might show stronger iatrogenic effects, as one might hypothesize based on deviancy training effects. Both interaction tests were non-significant. Finally, analyses of gender indicated that the effect (or lack thereof) of a peer group component did not differ as a function of the proportion of males in the sample.

We next conducted several analyses separately in the subset of studies that included a peer group treatment component ($N=40$) to assess the effects of certain group characteristics that the deviancy training hypothesis would suggest may be related to outcome. We first tested whether the mean effect size, .79 (.34), differed significantly from zero; it did. We tested whether the probability of finding a positive effect size, .93 (.88), differed significantly from .50 (the chance level); it did. Next, we tested whether homogeneity of the groups for antisocial behavior was related to outcome. Both effects were non-significant, effect sizes for peer group interventions where all subjects were antisocial were not significantly different from effect sizes for groups with mixed youth. Only six of the treatments in this

comparison were non-homogeneous, which means that this test had relatively limited power; however, the mean effect size for the homogeneous groups was .79 (.34), which argues against the homogeneous groups being iatrogenic.

We tested whether the number of children in the peer treatment group or the ratio of the number of children to the number of adults, both of which would provide a measure of the difficulty that the adults might have controlling the group, was related to outcome. The effect of adult/child ratio on the dichotomous outcomes was significant, with *higher* ratios associated with a higher probability of a negative effect size. However, because there was only a single negative effect size in this analysis, it should be approached with caution. As a last step in evaluating the iatrogenic and deviancy training hypotheses, we conducted an analysis designed to maximize the likelihood of finding iatrogenic effects, by restricting the sample to (a) treatments involving a group component; (b) youth being treated specifically for delinquency and / or aggression; (c) youth between the ages of 10 to 14; and (d) groups homogeneous for antisocial youth. In this highly restricted sample, the mean effect size was .52 (.48) and in the studies not included in the restrictive subsample, .79 (.44); the difference was not significant. Nine percent (1 of 11) of the restrictive subsample showed a negative effect size whereas 5% (3 of 53) of the remainder showed a negative effect size; this difference was non-significant.

Funnel Plot—All literature reviews are subject to what Rosenthal (1979) has called the “file drawer problem.” That is, studies that fail to produce significant results may tend to end up in the “file drawer” rather than being published. One technique that can be used to identify this effect is the “funnel plot,” wherein the sample size is plotted on the y-axis, against the effect size on the x-axis, across the different studies in the meta-analysis (Greenhouse & Iyengar, 1994). Because the variance of the effect size is an inverse function of the sample size, one should see an upside-down funnel as studies with large samples (and smaller variance) cluster more closely with less variability around the population effect size and studies with smaller samples (and larger variance) spread out around the population effect size. Bias against negative findings will produce a funnel that is truncated to the left, close to where the effect size equals zero, as studies with null or significant negative effect sizes are more likely to end up in the file drawer.

In order for the funnel plot to be appropriate, the sample data must be able to be represented by a single population mean (i.e., there must be no moderators of treatment effects). Thus, when we conducted our funnel plot test, we restricted our sample to studies that involved behavioral treatments, as previous meta-analyses indicate that the behavioral vs. non-behavioral treatment distinction probably is the most consistent and strongest moderator of treatment effects (Weisz et al., 1987; 1995). We separately plotted treatments that included and did not include a peer group component. Inspection of both funnel plots indicated clear censoring, although there were a number of negative effect sizes for both groups; median effect sizes were .41 and .42, respectively. We also plotted a funnel plot for treatment studies (from Weisz et al., 1987; 1995) targeting internalizing problems; this plot also indicated censoring.

Because the population effect sizes appear (based on their medians) to be positive, the likelihood of a study being censored is inversely correlated with its sample size; i.e., studies with larger samples will produce effect sizes closer to the (positive) population effect size, and hence will be less likely to be censored due to a spurious null or negative effect. Since WLS weights by the inverse of the variance (which is itself an inverse function of the sample size), to some extent WLS will mitigate the effects of censoring as it underweights studies with small sample sizes, which will be those that tend to be furthest from the population effect size and most likely to be censored. As a sensitivity analysis, to further reduce potential censoring effects we reconducted our analyses, restricting the sample to studies containing at least 50 subjects. By increasing the minimum sample size, we removed the bottom of the funnel, which is the part of the sample size distribution most likely to be censored. These reanalyses differed from the main analyses in only one respect: The interaction between Peer Group and the curvilinear effect of age went from marginal to non-significant.

Discussion

One possible response to concerns that group treatments may be iatrogenic would be to simply discontinue their use and focus on alternative forms of treatment, avoiding any risk of iatrogenic effects. This would ignore, however, the fact that group treatments can be an economical and convenient approach to providing interventions (Ang & Hughes, 2000). To discontinue their use unnecessarily would ignore the economic and cost-containment realities that have made the reduced clinician / client ratios of group therapy desirable (Steenbarger & Budman, 1996; Spitz, 2002). Thus, we believe that, as with any treatment, consideration of potential negative aspects must be balanced with potential positive aspects, including economic factors.

In this paper, we considered two inter-related hypotheses: (a) that adolescent group treatments are iatrogenic; and (b) that deviancy training underlies these iatrogenic effects. In our review, we found little strong evidence for either hypothesis. Conceptually, questions arise regarding the iatrogenic impact of intervention sessions relative to the numerous other influences on adolescent antisocial behavior. This conclusion is, however, limited by a key gap in the literature: There has been little research regarding the extent to which and how the group therapy context might magnify the effects of deviancy training or other iatrogenic processes.

Our empirical review of the literature found little support for either hypothesis. Specific studies as well as meta-analyses previously cited as suggesting iatrogenic effects contain what we believe to be little evidence for deviancy training or iatrogenic group treatment effects. Consequently, we conducted several analyses in our own meta-analytic dataset to evaluate some of these questions. We found that overall, the likelihood of a study producing a negative effect size was actually significantly *smaller* for studies that involved a peer group component. We also tested a number of interaction effects, to determine if there were specific subgroups of subjects or specific characteristics of treatment that might be associated with stronger iatrogenic effects. Of eighteen tests, only one produced a result supportive of potential iatrogenic effects. Although this effect was statistically marginal, it is

interesting to consider its implications. In this test, we found that the likelihood of producing a negative effect size among treatments involving a peer group peaked at age 11, at the beginning of early adolescence as suggested by the deviancy training hypothesis. The fact that the likelihood peaked at age 8.6 for treatments not involving a peer group rather than in early adolescence suggests that this effect is not simply due to developmental factors influencing treatment outcome in general. However, this effect was marginal, and the test using the continuous effect sizes was non-significant, which suggests that this result probably should be viewed as suggestive rather than conclusive.

Although the analyses conducted in our review were not supportive of iatrogenic or deviancy training effects, it is certainly possible that deviancy training might occur in some group treatment contexts, under conditions not tested here. For instance, results of our review (e.g., the fact that the potential amount of deviancy training occurring during therapy sessions is relatively limited compared to what could occur outside therapy) suggest that, if iatrogenic effects do occur, they would most likely occur (a) with youth who have experienced rejection by mainstream peers but are not yet associated with deviant peers, or similarly (b) with youth who live in the same neighborhoods as delinquent youth with whom they are not yet acquainted. For these youth, therapy could be iatrogenic insofar as it provides a route to new friendships with delinquent peers that are maintained outside of treatment. However, if youth are at risk of delinquency, then it appears that they also are at risk of associating with deviant peers, regardless of their involvement with group treatment. One might argue that parental monitoring could be preventing these youth from associating with deviant peers outside of treatment. Yet, if parents monitor their children sufficiently closely to prevent negative peer associations independent of therapy, then it seems unlikely that parents would allow negative peer relationships initiated during group sessions to develop.

In considering how our findings may have been impacted by a possible publication bias, we note three things: (a) the fact that there were a number of negative effect sizes indicates that any bias was not absolute; (b) we used WLS which means that studies with smaller samples influenced to our results less. Because the population effect size appears to be positive, WLS to some extent will have mitigated effects of any publication bias, since it de-emphasizes the bottom of the funnel where the censoring occurs; (c) when we restricted our sample to larger sample size studies (essentially giving the small sample studies a weight of zero) our results were essentially unchanged.

This censoring undoubtedly led to an overly-positive overall mean effect size. However, because in this review we were not interested in the overall effects of treatment but rather relative effects, this would not necessarily have impacted our results. In order for it to have done so, there would have to be differential bias, such that the mean effect size for the treatments with peer groups was more inflated than the mean effect size for treatments without peer groups. It is even less clear how censoring would have influenced other tests, such as our evaluation of variables such as length of treatment, since they were not based on mean effect sizes. It perhaps could have restricted the range of the effect sizes and attenuated the correlation, but our data did include a number of negative effect sizes, which argues against substantial restriction of range.

Although concerns regarding publication bias have been raised before – unfortunately with little apparent effect (cf., Cooper, DeNeve & Charlton, 1997; Rosenthal, 1978) – it nonetheless is important to again urge the publication of results based not on their statistical significance but rather on the importance of the question being addressed, and the quality of the design and implementation. This will require both researchers as well as journal editors and reviewers to change their perspectives (Cooper et al., 1997) but perhaps recent discussions regarding potential iatrogenic effects will highlight the importance of null or negative findings.

Based on our review, we have a number of research recommendations. First, if we are to understand when iatrogenic group treatment effects might occur, it will be important to more fully understand the processes occurring during group sessions that may foster such iatrogenic effects. A key aspect of this will be determining what, if anything, is special about group therapy, what aspects of group therapy might exacerbate the effects of deviancy training or other iatrogenic processes, such that their impact is disproportionate to their time. For instance, is deviant talk especially reinforcing when conducted during an adult-sanctioned and monitored group, perhaps because adult reactions intended to control inappropriate behavior actually may be reinforcing, as these reactions validate to his or her peers the adolescent's disregard and lack of fear for adult authority?

It also will be critical to track participants' friendships, both inside and outside the group, prior to the treatment beginning, during the groups, and subsequent to the end of treatment. It will be important to determine not only who group participants' friends are, but how they spend their time together outside of therapy, the extent to which they engage in deviancy training outside therapy as well as during sessions. Further, participants' reactions to the deviancy training, both in regards to their attitudinal change and self-reports of its reinforcement value as well as overt behavior should be assessed.

Yet another area where we need more information is in regards to whether there might be particular domains of antisocial behavior that are more susceptible to iatrogenic group effects and deviancy training. For instance, anecdotal reports and Dishion et al.'s (1999) findings regarding smoking suggest that illegal substance use may be one area particularly vulnerable to deviancy training or other iatrogenic effects.

We have two final recommendations. First, at present there has been relatively little discussion regarding the timeframe during which negative effects might occur. Although we believe that the specific studies we reviewed do not provide much evidence for deviancy training or other iatrogenic effects, they do highlight the fact that we have not clearly specified the timeframe through which one might expect iatrogenic effects to occur. On the one hand, deviancy training can clearly have an immediate impact on behavior (Dishion et al., 1999) but a case also could be made for delayed effects, as friendships develop and lead to more involvement with even more deviant peer groups.

Finally, because it is the hypothesized process that has received the most attention, our review focused on deviancy training as the primary cause of potential group treatment iatrogenic effects. However, there are other potentially iatrogenic processes, such as stigma

associated with being a member of a group of deviant youth, that will be important to study and monitor.

In conclusion, results of our review concur with those of Handwerk et al. (2000), who concluded that the risk of iatrogenic effects may at present be overstated. This is not to say that one should not be cautious when considering a group treatment for children or adolescents: When one considers any treatment, one should be cautious and consider how the situation of the individual fits and does not fit with various forms of treatment. We are not advocating for the use of group treatments but simply that the evidence be evaluated carefully. We agree with Dishion et al. (1999) and others that more research is warranted, but at present we find little support in the literature for iatrogenic effects, deviancy training based or otherwise.

Acknowledgments

This research was supported in part by: NIMH grants R01-MH-54237, and R01-MH-58275 to the first author, the American Psychological Foundation's Elizabeth Munsterberg Koppitz Fellowship to the second author, and NIMH grant R01-MH068806 and a Research Network on Youth Mental Health grant from the MacArthur Foundation to the six author, and by the Vanderbilt Institute for Public Policy Studies. The authors wish to thank Tom Dishion for providing additional information regarding published studies.

References

- Agresti, A. *Categorical data analysis*. New York City: Wiley; 1990.
- Ang RP, Hughes JN. Differential benefits of skills training with antisocial youth based on group composition: A meta-analytic investigation. *School Psychology Review*. 2002; 31:164–185.
- Arnold ME, Hughes JN. First do no harm: Adverse effects of grouping deviant youth for skills training. *Journal of School Psychology*. 1999; 37:99–115.
- Berger, RJ.; Crowley, JE.; Gold, M.; Gray, J.; Arnold, MS. *Experiment in juvenile court: A study of a program of volunteers working with juvenile probationers*. Ann Arbor, MI: Institute for Social Research; 1975.
- Brendgen M, Vitaro F, Bukowski WM. Stability and variability of adolescents' affiliation with delinquent friends: Predictors and consequences. *Social Development*. 2000; 9:205–225.
- Cooper H, DeNeve K, Charlton K. Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods*. 1997; 2:447–452.
- De-Haan LG, MacDermid SM. Identity development as a mediating factor between urban poverty and behavioral outcomes for junior high school students. *Journal of Family and Economic Issues*. 1999; 20:123–148.
- Dishion TJ, Andrews DW. Preventing escalation in problem behaviors with high-risk young adolescents: Immediate and 1-year outcomes. *Journal of Consulting and Clinical Psychology*. 1995; 63:538–548. [PubMed: 7673531]
- Dishion TJ, Bullock BM, Granic I. Pragmatism in modeling peer influence: Dynamics, outcomes and change processes. *Development and Psychopathology*. 2002; 14:969–981. [PubMed: 12549712]
- Dishion TJ, Capaldi DM, Spracklen KM, Li F. Peer ecology of male adolescent drug use. *Development and Psychopathology*. 1995; 7:803–824.
- Dishion TJ, McCord J, Poulin F. When interventions harm: Peer groups and problem behavior. *American Psychologist*. 1999; 54:755–764. [PubMed: 10510665]
- Dodge KA. Cost effectiveness of psychotherapy for child aggression: First, is there cost effectiveness? *Group Dynamics: Theory, Research and Practice*. 1999; 3:275–278.
- Elliott, DS.; Menard, S. *Delinquent friends and delinquent behavior: Temporal and developmental patterns*. In: Hawkins, D., editor. *Delinquency and crime: Current theories*. Cambridge, England: Cambridge University Press; 1996. p. 28–67.

- Fergusson DM, Horwood LJ. Prospective predictors of deviant peer affiliations in adolescence. *Journal of Child Psychology and Psychiatry*. 1999; 40:581–592. [PubMed: 10357164]
- Fo WSO, O'Donnell CR. The buddy system: Effect of community intervention on delinquent offenses. *Behavior Therapy*. 1975; 6:522–524.
- Gottfredson GD. Peer group interventions to reduce the risk of delinquent behavior: A selective review and a new evaluation. *Criminology*. 1987; 25:671–714.
- Greenhouse, JB.; Iyengar, S. Sensitivity analysis and diagnostics. In: Cooper, H.; Hedges, LV., editors. *The handbook of research synthesis*. New York: Russell Sage Foundation; 1994. p. 383-398.
- Handwerk ML, Field CE, Friman PC. The iatrogenic effects of group intervention for antisocial youth: Premature extrapolations? *Journal of Behavioral Education*. 2000; 10:223–238.
- Hedges, LV.; Olkin, I. *Statistical methods for meta-analysis*. Orlando, FL: Academic Press; 1985.
- Kazdin AE, Weisz JR. Identifying and developing empirically supported child and adolescent treatments. *Journal of Consulting and Clinical Psychology*. 1998; 66:19–36. [PubMed: 9489260]
- Larson RW. How U.S. children and adolescents spend time: What it does (and doesn't) tell us about their development. *Current Directions in Psychological Science*. 2001; 10:160–164.
- Lipsey, MW. Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In: Cook, TD.; Cooper, H.; Cordray, DS.; Hartmann, H.; Hedges, LV.; Light, RJ.; Louis, TA.; Hosteller, F., editors. *Meta-analysis for explanation: A casebook*. New York: Russell Sage Foundation; 1992. p. 83-127.
- Lundman, RJ. *Prevention and control of juvenile delinquency*. 3rd ed.. London: Oxford Press; 2001.
- McCord J. A thirty-year follow-up of treatment effects. *American Psychologist*. 1978; 33:284–289. [PubMed: 655479]
- McCord J. Crime in moral and social contexts: The American Society of Criminology, 1989 Presidential Address. *Criminology*. 1990; 28:1–26.
- McCord, J. The Cambridge-Somerville Study: A pioneering longitudinal-experimental study of delinquency prevention. In: McCord, J.; Tremblay, RE., editors. *Preventing antisocial behavior: Interventions from birth through adolescence*. New York: Guilford Press; 1992. p. 196-206.
- McCord, J.; Widom, CS.; Crowell, NA. *Juvenile crime, Juvenile justice*. Washington, DC: National Academy Press; 2001.
- Merriam-Webster Dictionary. (n.d.). from <http://www.m-w.com>.
- Poulin F, Dishion TJ, Burraston B. Three-year iatrogenic effects associated with aggregating high-risk adolescents in cognitive-behavioral preventive interventions. *Applied Developmental Science*. 2001; 5:214–224.
- Rosenthal R. The file drawer problem and tolerance for null results. *Psychological Bulletin*. 1979; 86:638–641.
- Spitz, HI. The impact of managed care on the group therapy of substance abuse. In: Spitz, HI.; Brook, DW., editors. *The group therapy of substance abuse*. New York: Haworth Press; 2002. p. 3-18.
- Steenbarger BN, Budman SH. Group psychotherapy and managed behavioral health care: Current trends and future challenges. *International Journal of Group Psychotherapy*. 1996; 46:297–309. [PubMed: 8753149]
- Tarter RE. Etiology of adolescent substance abuse: A developmental perspective. *The American Journal on Addictions*. 2002; 11:171–191. [PubMed: 12202010]
- Weisz JR, Weiss B, Han SS, Granger DA, Morton T. Effects of psychotherapy with children and adolescents: A meta-analysis of treatment outcome studies. *Psychological Bulletin*. 1995; 117:450–468. [PubMed: 7777649]
- Weisz JR, Weiss B, Alicke MD, Klotz ML. Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology*. 1987; 55:542–549. [PubMed: 3624610]
- Wills TA, Cleary SD. Validity of self-reports of smoking: Analyses by race / ethnicity in a school sample of urban adolescents. *American Journal of Public Health*. 1997; 87:56–61. [PubMed: 9065227]