Journal of Korean Society for Health Education and Promotion
Vol. 24, No. 5, 2007, pp.1-21

A meta-analysis of adolescent psychosocial smoking prevention programs in the United States: Identifying factors associated with program effectiveness

Myunghee Song Hwang
Dept. of Health Education and Management, Ewha Woman's University, Seoul, Korea

CONTENTS

Ⅰ. Introduction
Ⅱ. Methods
III. Results
IV. Discussion
References
Abstract

I. Introduction

Evaluation studies of adolescent psychosocial smoking prevention programs have been published since the late 1970s in the United States. Their success in reduction of adolescent smoking behaviors can be found throughout the literature, but individual studies have not provided consistently positive results in decreasing youth tobacco use. For instance, Botvin and Eng (1982) found a significant 58% reduction of new smoking behaviors among 7th grade students at 1-year follow-up while Burke et al. (1987) and Ary et al. (1990) failed to achieve significant results. The contrasting results among studies may come from the characteristics of psychosocial approaches that allow as many as multi-strategies and multi-channels when planned and implemented. Study programs employing different program delivery type, program leaders, instructional material, age of the program participants, use of media supplements, and level of parental involvement could not have consistent

Corresponding Author: Myunghee Song Hwang, PhD, CHES
Parkview 610-2002, Jeongja-dong, Bundang-gu, Seongnam-si, Gyeonggi-do, Korea
Dept. of Health Education and Management, Ewha Woman's University, Seoul, Korea
Tel: , E-mail: songhwang2000@yahoo.com

findings. For example, positive effects have been reported when psychosocial smoking prevention programs were delivered by project staff (Jason et al. 1982; Shaffer et al. 1983). Biglan (1987) and Colquhoun and Cullen (1981), however, found successful results in adolescent smoking behavior with teacher-led delivery. On the other hand, Clarke et al. (1986) and Lloyd et al. (1983) failed to achieve significant results in programs led by teachers. Mixed study results lead to confusion in planning, implementing, and evaluating adolescent tobacco education programs for health planners, educators, and decision-makers.

The second reason for conflicting results relates to poor quality research in the adolescent psychosocial smoking prevention literature. Flay (1985) found many internal validity issues based on the examination of 50 studies using social influences approach to smoking prevention. The methodological problems included mixed unit of analysis, different attrition rates from treatment and control condition, lack of biochemical validation of self-report measures, the number of random assignments to the conditions, the use of inappropriate comparison groups, and unequal baseline between treatment and control group (Battjes 1985; Best et al. 1988; Flay 1985; Glynn 1989; Snow et al. 1985). These studies with methodological limitations gave readers alternative interpretations of their reported effects and increased the gaps of program understanding.

Several meta-analysis studies measured small to medium overall program effects of adolescent psychosocial smoking prevention studies. For instance, Rundall and Bruvold (1988), Bruvold (1993), Rooney and Murray (1996), Tobler (1997), and Hwang, Lux-Yeagley, and Petosa (2004) estimated smoking behavior effect sizes .32, .19 - .31, .10 - .11, .16, and .18- .19, respectively. U.S. Dept. of Health and Human Services (1991) concluded social psychological smoking prevention models were only modestly effective across a variety of setting, times, and populations. At this point, rather than summarizing general program effects, the author would like to explore what variables associate with program effect enhancement.

This meta-analysis study is the secondary analysis after the previous study estimated mean effect sizes in smoking knowledge, attitudes, skills, and behaviors with treatment variables. To identify explanatory variables that are likely to increase program effects the author inputs demographic characteristics of program participants, research methodologies used in individual studies, and program variation factors related to ecology and implementation. The findings of the analysis recommend practical guidelines for smoking prevention efforts.

II. Methods

1. Research methods

The main purpose of this study was not to summarize the results of adolescent psychosocial smoking prevention programs (treatment variable) evaluated between 1978 and 1997, but to identify the program characteristics (explanatory variable) that influence smoking behavior outcomes and estimate how much they modify program effects.

The previous study examined 65 adolescent psychosocial smoking prevention programs (1978 to 1997) among students in grades 6 to 12 in the United States. An independent psychosocial smoking prevention program was the unit of analysis in the current meta-analysis. To summarize the procedure of previous meta-analysis, the study samples were identified through on-line and ancestry searches based on inclusion criteria defining a conceptual set of data and screening out studies with serious methodological flaws (See Hwang et al. 2004 MA for details).

The systematic coding of study characteristics permits an analytically precise examination of the relationships between study findings and such study features as respondent characteristics, nature of treatment, research design, and measurement procedures (Lipsy and Wilson 2001). One hundred forty-six variables per program were sorted by six categories regarding demographic characteristics, research methodology, modality, program variation, outcome assessment, and effect sizes (A code book and its descriptions can be requested from the author). Since not every study provided sufficient information to calculate effect sizes, primary researchers were contacted for the missing data on their published articles. Ambiguous coding interpretations were resolved in the regular meetings with a panel of four experts.

Each program effect was measured in terms of effect size, defining as the standardized mean difference between treatment and control group in experimental studies (Hedges 1985). Hedges' estimation of effect size was primarily used in this meta-analysis study that included unbiased estimator (d) and weighted mean effect size (d.). The effect sizes were estimated with the principle of using only one effect size in a given construct. Smoking behavior constructs were measured using three different time intervals: Short-term behavior effect ranged from 1 to 12 months after program completion, mid-term from 13 to 36 months, and long-term from 37 to 72 months.

2. Explanatory variables

According to Lipsey (1994), explanatory

variables should not be assumed to directly shape the results of the study. Unlike the interpretation of effect sizes by treatment variables, the effect sizes estimated from explanatory variables (associating factors) here suggest possible variables which may increase program effects for the reduction of adolescent smoking behaviors.

Based on previous research on smoking prevention and meta-analysis studies, the variables thought to be potentially related to study results were extracted and coded in considerable numbers. Explanatory variables measured in this study with four categories - demographic, research quality, implementation, and ecological factors.

3. Outcome variables

The outcome variable of smoking behaviors was estimated using three time intervals: short-term (1 to12 months), mid-term (13 to 36 months), and long-term (37 to 72 months). Several reasons for the specific definitions of time intervals can be given in this meta-analysis. A grand mean of effect sizes over time like other meta-analysis studies is not meaningful to comprehend the practical significance of the program effects. Considering the broad range of program outcome assessment periods (up to 5 years) among individual studies, the first-year follow-up studies represent higher program effect sizes than the effect sizes of five-year follow-up studies. Therefore, program effects should be analyzed by appropriate time intervals. The definition of short-term is given based on the fact that most studies reported their program outcomes within or at the point of time of 12 months after program completeness. The definition of mid-term is based on the findings of Tobler (1997) and Hwang (2000) that the initial program effects on drug/ or smoking behaviors were persistent over three years in their meta-analysis studies.

III. Results

1. Effect size variation by demographic characteristics

Weighted mean effect sizes were estimated to determine whether adolescent smoking behavior was influenced by the characteristics of program participants that included geographic location, social economic status, gender, predominant ethnicity, grade, and school type. Only one demographic characteristic, grade level, demonstrated an important difference in the magnitude of effect size. Figure 1 shows the changes in short and mid-term smoking behavior effect sizes by grade levels. The younger students (5^{th}-7^{th} grades) maintained higher effects with the psychosocial smoking prevention programs compared to

the older students (8^{th}-12^{th} grades). A sudden drop was observed for students in grade 8. Mean effect sizes of younger students in 5^{th} - 7^{th} grades were .19 for short-term smoking behavior and .18 for mid-term smoking behavior while the average effect sizes of older students in 8^{th} - 12^{th} grades were .11 for short-term smoking behavior and .05 for mid-term smoking behavior.

2. Effect size variation by research methodological factors

Methodologically rigorous studies are generally recommended for meta-analysis, but no agreement exists among meta-analysts about which factors influence the overall program effects. This meta-analysis combined study findings based on experimental research at least using treatment versus control group comparisons because only experimental studies could meet the methodological standards needed for meaningful results in meta-analysis. Table 1 demonstrates the mean effect size estimated by research methodological factors.

The reader should be careful to interpret the findings of this meta-analysis, especially the results of studies with methodological flaws. These results do not mean that methodologically poor studies are likely to increase program effects. High quality of studies is related to higher effect sizes. However, close examination shows this is not always the case. Since the studies with research methodological defects yield inaccurate average effect sizes, the effect sizes estimated from those studies were negatively interpreted here as having both inflated (increased) and deflated (decreased) meanings. For short-term smoking behavior, the small scale of studies was likely to have inflated effect sizes (ES= .22) when compared to the medium scale of studies. Similarly, the medium scale of studies was likely to have inflated effect sizes (ES= .20) compared to the large scale of studies (ES= .18). The effect size of mid-term smoking was .26 for the small scale of studies while the effect sizes were .19 and .17 for the medium and large scale of studies.

According to Flay (1985)'s typology of psychosocial smoking prevention studies, each generation is characterized by serious levels of research methodological limitations. For instance, the first generation of psychosocial smoking prevention studies has many methodological problems such as non-random assignment or only one school (or class) assignment to the condition. The fourth generation studies are the most developed smoking prevention studies including large-scale, efficacy trials with at least 5 or more units randomly assigned to each condition. As shown in Table 1, the

Table 1. The weighted mean effect sizes by research methodological factors in the adolescent psychosocial smoking prevention programs evaluated from 1978 to 1997 in the United States

Methodological Factors	Smoking Behavior		
	Short-term	**Mid-term**	**Long-term**
Research Designs:			
True-experimental design (n=5)	.53 (n=5)		
Quasi-experimental design (n=60)	.19 (n=49)	.18 (n=26)	.09 (n=7)
Quasi-experimental Design:			
Without equivalence (n=19)	.06 (n=12)	.11 (n=9)	.06 (n=3).
With equivalence (n=33)	.23 (n=30)	.22 (n=13)	.13 (n=3)
Number of Units:			
>=5 (n=22)	.18 (n=16)	.17 (n=8)	.06 (n=5)
3-4 (n=14)	.20 (n=11)	.19 (n=9)	
1-2 (n=14)	.22 (n=14)	.26 (n=3)	
Control Group:			
No treatment (n=13)	.18 (n=11)	.18 (n=3)	
Traditional tobacco education (n=14)	.04 (n=10)	.16 (n=6)	
General health curricula (n=10)	.22 (n=7)	.17 (n=7)	-.01 (n=3)
Attrition Rates:			
0 - 10% (n=9)	.30 (n=9)	.27 (n=3)	.13 (n=3)
11% or more (n=28)	.15 (n=24)	.16 (n=13)	.00 (n=2)
Differential Attrition Between Groups:			
1) Regarding whole group members			
Equal attrition (n=21)	.15 (n=14)	.17 (n=10)	.06 (n=3)
Unequal attrition (n=9)	.22 (n=8)	.21 (n=4)	.19 (n=2)
2) Regarding smokers			
Equal attrition (n=8)	.08 (n=4)	.15 (n=6)	.00 (n=2)
Unequal attrition (n=8)	.23 (n=7)	.21 (n=5)	.12 (n=3)
Research Generations:			
First generation (n=5)	.24 (n=5)	.10 (n=2)	
Second generation (n=16)	.20 (n=15)	.28 (n=5)	
Third generation (n=18)	.19 (n=16)	.21 (n=9)	
Fourth generation (n=25)	.18 (n=17)	.16 (n=10)	

weighed mean effect sizes were inflated as the generations were lowered.

In summary, effect sizes were likely to be inaccurate (inflated or deflated compared to the average effect sizes) with poor quality smoking prevention studies characterized by one of following factors:

a. Quasi-experimental studies were likely to have deflated effect sizes compared to true-experimental studies.

b. Quasi-experimental studies with non-equivalence between groups were likely to have deflated effect sizes compared to quasi-experimental studies with equivalence

between groups.

c. Studies with smaller number of units assigned to the experimental conditions were likely to have inflated effect sizes compared to the studies with larger number of units assigned to the groups.

d. Control groups with traditional tobacco education (i.e. informative and/ or motivational programs) were likely to have deflated effect sizes compared to control groups with no-treatment/ or comprehensive health education in the psychosocial smoking prevention studies.

e. Higher rates of attrition (less than or equal to 10 percent) were likely to have deflated effect sizes compared to lower rates of attrition.

f. Differential attrition rates between treatment and control groups were likely to have inflated effect sizes compared to the equal attrition rates between groups for both whole members and smokers-only.

g. Lower generation studies were likely to have inflated effect sizes compared to the higher generation studies.

3. Effect size variability and influentialibility by common methodological problems

Methodological problems were further analyzed in this section based on literature reviews. First, the methodological issues that have been cited in a relatively large number of the smoking prevention literature were selected. They included non-random assignment, baseline non-equivalence, unit of analysis problem, failure to consider the impact of attrition on program outcome, absent or low instrument reliability, self-reported data only problem, and low or unknown implementation fidelity. Second, the problems that typically occurred in the research were examined to discover their prevalence among the studies in this meta-analysis. As shown in Table 2, the self-reported data only problem (20.0%) was the least prevalent while unit of analysis problems (76.9%) was the most prevalent among the studies. Many studies assigned the unit of school or class to experimental conditions, but they analyzed data at the level of students instead of using school or class unit. According to Zucker (1990), this method can result in Type I error rates and study results inflation. Implementation fidelity (refers to the degree to which the program is completely implemented as it is planned) was low or unknown among the studies included in this meta-analysis (70.8%) - a Type III error (Scanlon et al, 1977). Finally, effect sizes were estimated by these seven methodological problems to determine if the magnitude of effect sizes were affected by those factors (Table 2).

Generally speaking, the direction of effect size variability (inflated versus deflated) and

Table 2. The weighted effect sizes by research methodological problems in the adolescent psychosocial smoking prevention programs evaluated from 1978 to 1997 in the United States

Common Research Methodology Problems (Prevalence)		Smoking Behavior		
		Short-term	**Mid-term**	**Long-term**
Unit of analysis (76.9%)				
Yes	(n=50)	.18 (n=43)	.19 (n=21)	.12 (n=4)
No	(n=15)	.21 (n=10)	.17 (n=6)	.00 (n=2)
Implementation fidelity (70.8%)				
Yes	(n=46)	.14 (n=41)	.19 (n=18)	.06 (n=2)
No	(n=19)	.30 (n=12)	.17 (n=9)	.10 (n=4)
Instrument reliability (55.4%)				
Yes	(n=36)	.14 (n=31)	.17 (n=18)	.00 (n=2)
No	(n=29)	.28 (n=22)	.20 (n=9)	.12 (n=3)
Attrition (49.2%)				
Yes	(n=32)	.23 (n=29)	.20 (n=10)	.13 (n=3)
No	(n=33)	.16 (n=24)	.18 (n=17)	.06 (n=5)
Pretest equivalence (32.3%)				
Yes	(n=29)	.11 (n=21)	.14 (n=12)	.06 (n=3)
No	(n=36)	.23 (n=32)	.22 (n=15)	.13 (n=3)
Random assignment (29.2%)				
Yes	(n=19)	.22 (n=16)	.21 (n=13)	
No	(n=46)	.17 (n=37)	.16 (n=14)	.06 (n=5)
Self-report only (20.0%)				
Yes	(n=13)	.19 (n=12)	.27 (n=4)	
No	(n=52)	.19 (n=42)	.17 (n=23)	.06 (n=3)

the influentialibility (the differences in effect sizes between problem and no-problem subgroups) depended on the type of methodological problem. Deflated effect sizes related to problems with implementation fidelity, instrument reliability, and pretest equivalence. Inflated effect sizes related to methodological problems of attrition, random assignment, and self-report (only for mid-term smoking behavior). The problem of unit of analysis was not consistent with either direction of effect sizes. The differences in effect sizes between problem and no-problem subgroups indicated the degree to which the methodological problem affected the magnitude of effect sizes. Possible methodological problems affecting the magnitude of effect sizes for short-term smoking behavior were implementation fidelity, instrument reliability, pretest equivalence, attrition, and random assignment, in order from the largest difference between the subgroups. This order was also compatible with the prevalence of methodological problems except for the unit of analysis problem.

4. Effect size variation by program implementation factors

Implementation factors varied by type of program leader, type of leadership training, drug of treatment, and treatment strength. Effect size variability in smoking behaviors for implementation factors are shown in Table 3.

Table 3. The weighted effect sizes by implementation factors in the adolescent psychosocial smoking prevention programs evaluated from 1978 to 1997 in the United States

Implementation Factors	Smoking behavior		
	Short-term	**Mid-term**	**Long-term**
Program Leaders:			
Peer-led delivery (n=17)	.20 (n=17)	.25 (n=6)	.13 (n=3)
Peer-facilitated delivery (n=14)	.14 (n=9)	.19 (n=10)	
Teacher-led delivery (n=23)	.19 (n=17)	.16 (n=9)	.08 (n=2)
Staff-led delivery (n=19)	.19 (n=17)	.20(n=9)	
Type of Leadership Training:			
Training with written or AV (n=11)	.16 (n=8)	.15 (n=6)	
Training with both written and practice (n=28)	.17 (n=23)	.20 (n=12)	
Training with practice (n=8)	.34 (n=7)	.33 (n=2)	.06 (n=2)
Trained hours <=4 (n=12)	.09 (n=11)	.15 (n=3)	
Trained hours >=5 (n=25)	.23 (n=21)	.21 (n=13)	.06 (n=2)
Drug of Treatment:			
Cigarette only (n=33)	.24 (n=28)	.19 (n=13)	.06 (n=2)
Tobacco, alcohol, and other drugs (n=19)	.17 (n=15)	.16 (n=10)	.06 (n=2)
Treatment Strength:			
1) Treatment session			
Less than 10 (n=35)	.12 (n=30)	.16 (n=14)	.07 (n=5)
10 or more (n=30)	.23 (n=24)	.19 (n=13)	
2) Treatment period			
1-40 days (n=25)	.11 (n=19)	.16 (n=9)	.00 (n=2)
41-200 days (n=24)	.24 (n=20)	.18 (n=10)	.09 (n=2)
3) Treatment Tense			
Daily (n=10)	.17 (n=7)	.19 (n=2)	
Semi-weekly (n=12)	.18 (n=11)	.28(n=4)	
Weekly (n=23)	.22 (n=17)	.15 (n=10)	.06 (n=2)
Monthly (n=4)	.16 (n=4)		
4) Booster programs			
Not given (n=47)	.17 (n=39)	.19 (n=17)	.06 (n=2)
Given (n=18):	.23 (n=14)	.18 (n=10)	.10 (n=4)
1-4 sessions (n=8)	.17 (n=6)	.14 (n=4)	.00(n=2)
5-15 sessions (n=10)	.26 (n=8)	.20 (n=6)	.15 (n=2)

The program leaders investigated in this meta-analysis included peer leaders, peer assistants, teachers, and program staff. Even though the effect sizes were not noticeably different among program leaders, the use of peer leaders was the most effective for overall smoking behavior. If peer leader were defined as students who were recruited and trained from junior high or high schools, then the mean effect size (n=11) was .23 for short-term smoking behavior and .28 for mid-term smoking behavior. The peer -facilitated method was the least effective method using peers assisting adult leaders.

The mean effect sizes of leader's training with practice were higher for smoking behavior (ES= .34 at short-term; and ES= .33 at mid-term) than other types. Leadership training hours ranged from 2 to 120 hours with a median of 7. The training hours of 4 or less were less effective for smoking behavior (ES= .09 at short-term; and ES= .15 at mid-term) than the training hours of more than 4 hours (ES= .23 at short-term; and ES= .21 at mid-term).

The programs only targeted to cigarette smoking had higher effect sizes for all outcome variables than the programs comprehensively targeted to tobacco, alcohol, and other illegal drugs. The effect sizes of cigarette-only programs demonstrated .24 at short-term and .17 at mid-term for smoking behavior while the effect sizes of comprehensive drug programs were .17 at short-term and .16 at mid-term.

Treatment strength was separately examined by the following factors: number of treatment sessions, whole period of treatment given (unit of days), treatment tense (unit of frequencies), and booster programs given. For the number of treatment sessions, this meta-analysis investigated effect sizes by two groups of treatment sessions: less than 10 sessions and 10 or more sessions based on Desenbury (1997)'s recommendation. For short-term smoking behavior, the effect sizes were.23 for 10 or more treatment sessions (10>= sessions) and .12 for less than 10 sessions (<10 sessions). Treatment period was separately calculated by counting only weekdays (i.e. 5 days per week, 20 days per month), thus converting both the number of treatment sessions and the intensity level to one figure. Treatment periods of 40 days or shorter (<=40 days) had a low effect size of .11 for short-term smoking behavior while treatment periods of longer than 40 days (40> days) had an effect size of .24. Treatment tense refers to the frequency levels when the treatment programs were given. The levels of tense included daily, semi-weekly, weekly, and monthly programs. Short-term smoking behavior was the only variable available to measure and compare all four tense levels for their effect sizes. Weekly-based

Table 4. The weighted effect sizes by ecological factors in the adolescent psychosocial smoking prevention programs evaluated from 1978 to1997 in the United States

Ecological Factors	Smoking Behavior		
	Short-term	Mid-term	Long-term
Research Centers:			
University of Southern California (n=12)	.22 (n=6)	.19 (n=9)	
Cornell University (n=9)	.44 (n=8)		
University of Washington (n=7)	.29 (n=6)	.28 (n=4)	
University of Minnesota (n=5)	.22 (n=5)	.28 (n=3)	.13 (n=3)
Stanford University (n=5)	.19 (n=5)	.28 (n=2)	
Study Result Significance:			
Significance (n=33)	.22 (n=30)	.21 (n=16)	
Insignificance (n=29)	.14 (n=23)	.10 (n=11)	.06 (n=5)

treatment programs had the highest effect sizes (ES= .22) and the monthly-based treatment programs had the lowest effect sizes (ES= .16). The majority of studies (72%) were not given booster programs among studies included in this meta-analysis. The programs given booster sessions (n=18) ranged from 1 to 15 sessions. The smoking behavior outcome variables were likely to have higher effect sizes for programs given booster sessions rather than the programs not given booster sessions.

5. Effect size variation by program ecological factors

According to Hedges and Becker (1989), subtle kinds of interdependence among effect sizes cannot always be removed. When a series of related studies is published by an individual or by a set of colleagues, they may be more alike than studies by other unassociated persons or groups. As shown in Table 4, research centers/ or projects tended to publish identical (or similar) program intervention studies (but using other subjects and methods) and to yield higher effect sizes.

The research center that published the greatest number of psychosocial smoking prevention studies was the University of Southern California (n=12). Unlike other research centers, the university conducted a variety of program interventions under the program/ or project names of MPP, HASP, TAPP, TVSFT, SMART, and TNT that represented different modalities and setting levels. The trials of various programs may lead to produce relatively lower effect sizes compared to other research centers like Cornell University and the University of Washington. The overall average effect sizes were .22 for short-term smoking behavior, .19 for mid-term smoking behavior, and .31

for short-term knowledge outcome. The TNT (Toward No Tobacco) represented higher effect sizes for smoking behavior among school-setting programs while MPP (Midwestern Prevention Project) and TVSFP (Television, School, and Family Project) among school-community setting. The Midwestern Prevention Project (MPP) represented the highest effect sizes (ES= .31 for short-term smoking behavior; and ES= .21 for mid-term smoking behavior) among the programs from the University of Southern California (USC).

The psychosocial smoking prevention studies from Cornell University demonstrated the highest effect sizes of all other research centers. Unlike the USC, the nine studies represented homogeneous program characteristics – Life Skills Training (LST) in the school setting only. The programs were more likely to be implemented with white, middle class, suburban students in New York State (n=7). In addition to the homogeneous intervention trials, the quality of the program structure and the quantity of the effect sizes distinguished LST from the other existing smoking prevention programs. The weighted mean effect sizes were .44 for smoking behavior, .16 for skills, .95 for knowledge, and .34 for attitudes in the short-term measure.

In overall program/ or project names, the Life Skills Training (LST) program revealed the highest effect size (.44) and the programs representing medium effects (effect size is .20 to 49) included LST, TAPP, MPP, RASP, MHHP, CLASP, NTE, SHOUT, and PMH for short-term smoking behavior.

The studies with significant results tend to estimate higher effect sizes than the studies with insignificant results. For smoking behavior, the effect sizes of statistically significant studies were .22 and .21 at short-term and mid-term while the effect sizes of insignificant studies were .14 and .10 at short-term and mid-term.

IV. Discussion

The objectives and methods of this meta-analysis are distinctive from other meta-analysis studies that summarized quantitative findings of a body of empirical research by averaging program effect sizes. This study attempted to determine the source of differences in adolescent psychosocial smoking prevention study findings that represent modest levels of program effects or mixed results in reducing adolescent smoking behaviors in the US. The variation in effect sizes across studies was examined by specific study characteristics, including demographic, research quality, implementation, and ecological factors.

Meta-analysis does not require large number of studies and, in some circumstances, can be usefully applied to as few as two or three study findings (Lipsey and Wilson 2001; Rosenberg et al. 1997). Robust and better meta-analysis statistics with small number of studies, however, should have large-sample approximations (Hedges 1994). According to Rosenberg (1997), Hedges' effect size of 'd' estimated in this meta-analysis works well with large-sample theory when sample sizes are at least 10. For this meta-analysis, mid- or long-term effects with small number of studies do not violate any condition indicating poor analysis, but have limitations to provide consistent interpretations about study results. Considerable differences of study numbers were often observed between short-term and long-term effects. Study results were not consistent over time intervals with some variables. The inconsistency can be explained by neutralizing effects over time or a large difference in the number of studies analyzed over time. No literature discusses or provides theoretical and/ or statistical explanation of this phenomenon to date.

Some critics said that research synthesis should be based only on findings from high quality studies. Studies with very strict methodological criteria for inclusion are likely to exclude much of the available evidence and can be unrepresentative in a given research area. On the other hand, studies with more relaxed methodological standards may reach potentially misleading conclusions. For example, Tobler's first meta-analysis (1986) of substance abuse prevention programs was criticized by being overly tolerant of methodologically poor studies (n=146). Bangert-Drown's meta-analysis (1988) included only 33 studies out of 126 studies on alcohol or drug abuse education with three fourths of the studies excluded based on strict standards for inclusion (i.e. studies were excluded if a researcher provided evidence of significant pretreatment differences between experimental and control group). The behavior effect sizes were .27 for the former study and .12 (not significantly different from zero) for the latter one.

There is little agreement among researchers on what constitutes good methodological studies to provide meaningful meta-analysis results. It is not also known which method issues would be most influential in research findings. This meta-analysis study examined a wide range of methodological characteristics in relationship to effect size variations. An important conclusion drawn from this comprehensive analysis of research methodology used among studies is that the factors related to poor quality of studies do

not always represent lower effect sizes than the factors related to high quality studies. Depending on methodological problems, the effect sizes were likely overestimated (inflated) or underestimated (deflated) compared to the studies with methodological rigor.

Previous studies have shown that variation in study effect sizes is often associated with methodological variation among studies (Hedges, Shymansky & Woodworth 1989; Lipsey 1994). Several meta-analysts have tried to compare the results from randomized experiments to those from quasi-experiments. In medical and surgical areas, the research suggests that randomized trials of medical innovations yield smaller estimates of effectiveness of the innovation (Colditz et al. 1988; Gilbert et al. 1978). In psychotherapy studies, the findings suggest that random assignment may make little difference to outcome (Smith et al 1980). For educational research, Becker (1990) found that experimental studies yield larger effect sizes than quasi-experimental studies on Scholastic Aptitude Test (SAT) coaching methods. Using the previously mentioned research areas, Heinsman (1993) found that the weighted average effect size of randomized experiments (d+ = 0.42*) was significantly higher than the effect size for quasi-experiments (d+ = 0.03). He also showed an effect size difference in the drug use prevention studies where the weighted least spuares (WLS) average effect size was over three times significantly larger for randomized (d+ = 0.51*, n=13) compared to quasi-experiments (d+ = 0.15*, n=17). The findings of these previous studies were consistent with the results from the present meta-analysis of psychosocial smoking prevention studies even though the current number of studies was small (n=5) for true-experimental studies compared to the number of studies (n=60) for quasi-experimental studies. The weighed mean effect size of smoking behavior was almost three times larger (ES= .53) for true-experimental studies than quasi-experimental studies (ES= .19).

This meta-analysis used the number of units assigned to the experimental conditions to examine effect size variations while the other meta-analysis studies used the number of students. Since most studies used the unit of school (65%) or class (19%) for the assignment of experimental conditions rather than individuals (8%), using the number of assignment units was more reasonable than using the number of program participants. Tobler (1997), on the other hand, analyzed using the number of students that were defined small sample as 20 to 400 youths and large sample as 401 to 4000 youths. Her study results were consistent with current

study findings in the direction of effect variations (the smaller number of units; the larger effect sizes), but her effect sizes were more inflated than the present study. The magnitude of effect size was .41 for small sample sizes and .13 for large sample sizes compared to .22 - .26 for 1-2 units, .19 - .20 for 3-4 units, and .17 - .18 for 5 or more units.

Of the methodological issues in the smoking prevention research area, the most significant may be the unit of analysis problem (Rooney and Murray 1996). Because the problem was the most common (77% of the present meta-analysis study) in the literature and can result in an inflated Type I error rate as well as an inflated estimate of the significance of a study's finding (Zucker 1990). The current study, however, did not find evidence that the research problem would affect the magnitude of effect size estimations. The results of this study are consistent with Rooney and Murray (1996)'s findings of no difference in average effect sizes before and after correcting the unit of analysis problem. The unadjusted effect size was .1145 at posttest and .1094 at follow-up for smoking behavior. After the correction of the wrong unit of analysis, the adjusted effect size changed little - .1130 at posttest and .1006 at follow-up. The results from these two meta-analyses, however, do not suggest granting a license to analyze at the level of the individual regardless of the unit of assignment (Rooney and Murray 1996). Rather than the unit of analysis problem, the current study found that the methodological issues of implementation fidelity, instrument reliability, attrition, pretest equivalence, and random assignment are more influential in research findings among psychosocial smoking prevention studies with experimental designs.

Adolescent psychosocial smoking prevention programs feature many variations in implementation strategies such as delivery leaders. Based on the fact that kids talk to kids, trained adolescents make sense as smoking intervention leaders to help their friends correct myths and misconceptions of tobacco use and to disseminate assertive communication techniques and decision-making skills. The relationship between the use of peer leaders and higher effect sizes was consistently found in meta-analysis studies. For instance, Bangert-Drowns (1988) found that the use of peer leaders was related to a significantly higher average effect size on attitudinal change in substance abuse education. Programs administered by peers had an average effect size of .64 while programs administered primarily by adults averaged .26. Rooney (1996)'s meta-analysis study also found a positive relationship between

same-age peer leaders and smoking behavior effect sizes. Unlike the results of these previous meta-analyses, Tobler (1997) found that mental health specialists (ES= .31) were more effective for drug behavior than peer leaders (ES= .15). This meta-analysis study examined further the effects of peer use in delivering smoking prevention programs by dividing peer-led and peer-facilitated methods. Peer-led delivery method was more effective than peer-facilitated method, but was not dominantly effective compared to research staff-led delivery. The smoking behavior effect sizes were re-estimated by categorizing same-aged peer leaders, older-aged peer leaders using high school students in middle school settings, and older-aged peer leaders using college students in high school settings. A conclusion was made that the successful use of peers in the delivery of smoking prevention programs should be primarily led by peers who are same-aged peers or older-aged high school students.

Treatment intensity levels varied across psychosocial smoking prevention programs. The present meta-analysis found that focused (including only tobacco topics) and

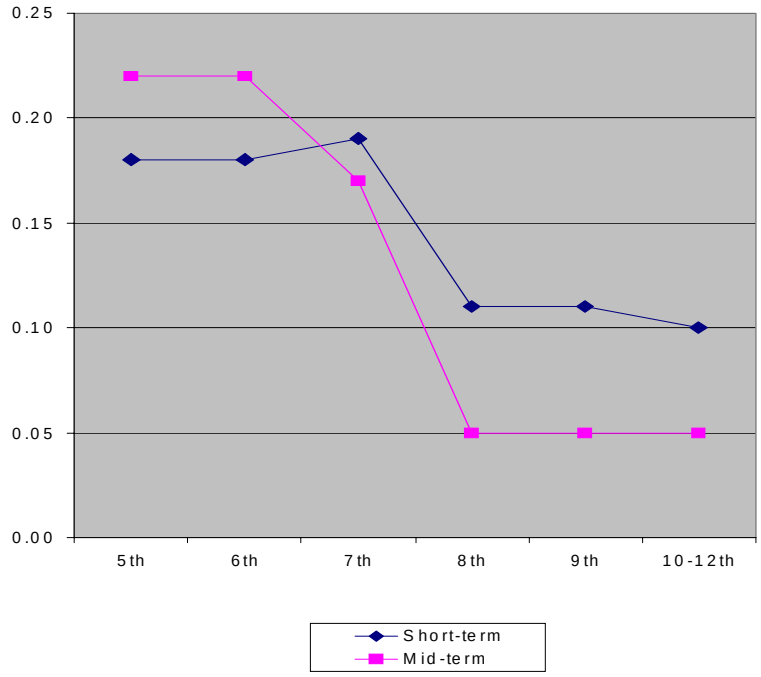

Figure 1. Short and mid-term smoking behavior effects by grade level in the adoleschent psychosocial smoking prevention programs evaluated from 1978 to 1997 in the United States

intense smoking prevention programs produced higher effects in deterring smoking behaviors. The study results are consistent with Tobler's (1997) finding of higher effects with cigarette-only programs, and Glynn (1989) and Dusenbury's (1997) guidelines of at least 10 or more treatment sessions for school-based smoking (drug abuse) prevention strategies.

Regarding program participants' characteristics, the crucial time to implement the program has been discussed in the tobacco education literature. The CDC guidelines for school health programs to prevent tobacco use (1994) recommend prevention education should be especially intensive in junior high or middle school and should be reinforced in high school. Earlier meta-analysis studies, however, did not find any significant difference of program effects between lower grades and higher grades of students. Tobler (1992) found no difference of drug behaviors between 6^{th} - 8^{th} middle school and 9^{th} - 12^{th} high school students. The present meta-analysis examined the effects on smoking behaviors by each grade level of students when the intervention begins (Figure 1). More successful results were shown when smoking prevention programs were implemented to students in grades 5 to 7 than the students in grades 8 to 12. Also, a sudden drop of program effects in grade 8 observed in this meta-analysis may lead to misleading results in earlier meta-analyses. The present study results are consistent with the CDC's guideline (1994) and Dusenbury's (1997) recommendations of 'begin the program during the transition year from elementary to middle school or junior high when new students are exposed to older students who use tobacco at higher rates'. Additionally, this study's results show specific evidence that the program should be reinforced in the students starting with 8^{th} grades.

References

Ary, D.V. et al. 1990. The efficacy of social influence prevention programs versus "standard care": Are new initiatives needed? *Journal of Behavioral Medicine* 13:281-296.

Battjes, R.J. 1985. Prevention of adolescent drug abuse. *International Journal of Addiction 20:*1113-1134.

Bangert-Drown, R.L. 1988. The effects of school-based substance abuse education - A meta-analysis. *Journal of Drug Education* 18(3):243-264.

Best, J.A. et al. 1988. Preventing cigarette smoking among school children. *Annual Review of Public Health* 9:161-211.

Botvin, G.J., and Eng, A. 1982. The efficacy of a multi-component approach to the

prevention of cigarette smoking. *Preventive Medicine* 11:199-211.

Bruvold W.H. 1993. A meta-analysis of adolescent smoking prevention programs. *American Journal of Public Health* 83:872-880.

Burke, J.A. et al. 1987. The short-term effects of competition and rewards in an adolescent smoking prevention program. *Health Education Quarterly* 14(2):141-152.

Centers for Disease Control and Prevention. 1994. Content and teaching strategies in 10 selected drug abuse prevention curricula. *Morbidity and Mortality Weekly Report* 43(RR-2):1-18.

Clarke, J.H. et al. 1986. Reducing adolescent smoking: A comparison of peer-led, teacher-led, and expert interventions. *Journal of School Health* 56:102-106.

Colquhoun, J., and Cullen, K. 1981. Improved smoking habits in 12-to-14-year-old Busselton children after antismoking programs. *Medical Journal of Australia* 1:586-587.

Dusenbury, L. and Falco, M. 1997. School-based drug abuse prevention strategies: From research to policy and practice. Pp.47-75 in *Enhancing Children's Wellness* edited by R. Weissberg, T. et al. Thousand Oaks: SAGE Publication.

Flay B.R. 1985. Psychosocial approaches to smoking prevention: a review of findings. *Health Psychology* 4:449-488.

Glynn T.J. 1989. Essential elements of school-based smoking prevention programs. *Journal of School Health* 59:181-188.

Hedges, L.V., and Olkin, I. 1985. *Statistical methods for meta-analysis*. Orlando: Academic Press.

Hedges, L.V., and Becker, B.J. 1989. Statistical methods in the meta-analysis of research on gender differences. Pp. 14-15 in *The psychology of gender: Advances through meta-analysis* edited by J.S. Hyde and M.C. Linn. Baltimore: The Johns Hopkins University Press.

Hwang, M.S., Yeagley, K.L., and Petosa, R. 2004. A meta-analysis of adolescent psychosocial smoking prevention programs published between 1978 and 1997 in the United States. *Health Education & Behavior* 31(6):702-719.

Jason, L.A. et al. 1982. Evaluating an early secondary smoking prevention intervention. *Preventive Medicine* 11:96-102.

Lipsey, M.W. 1992. Juvenile delinquency treatment: A meta-anlaytic inquiry into the variability of effects. Pp. 83-128 in *Meta-analysis for explanation: A casebook* edited by T. Cook, H. et al. New York: Russell Sage Foundation.

Lipsey, M.W. 1994. Identifying potentially interesting variables and analysis

opportunities. Pp. 111-124 in *The handbook of research synthesis* edited by H. Cooper and L.V. Hedges. New York: Russell Sage.

Lipsey, M.W., and Wilson, D.B. 2001. *Practical meta-analysis*. Thousand Oaks, CA: SAGE Publications.

Lloyd, D.M. et al. 1983. Cigarette smoking and drug use in school children: III-Evaluation of a smoking prevention education program. *International Journal of Epidemiology* 12(1):51-58.

Rooney B.L., and Murray D.M. 1996. A meta-analysis of smoking prevention programs after adjustment for errors in the unit of analysis. *Health Education Quarterly* 23:48-64.

Rosenberg, M.S., Adams, D.C., and Gurevitch, J. 1997. *MetaWin: Statistical software for meta-analysis with resampling tests*. Sunderland: Sinauer Associates Publishing.

Rundall T.G., and Bruvold W.H. 1988. A meta-analysis of school-based smoking and alcohol use prevention programs. *Health Education Quarterly* 15:317-334.

Scanlon J.W. et al. 1977. Evaluability assessment: Avoiding type III and IV errors. In *Evaluation Management: A Source Book of Readings* edited by G.R. Gilbert and P.J. Conklin. Charlottesville, VA: US Civil Service Commission.

Shaffer, H. et al. 1983. The primary prevention of smoking onset: An inoculation approach. *Journal of Psychoactive Drugs* 15(3):177-184.

Snow, W.H., Gilchrist, L.D., and Schinke, S.P. 1985. A critique of progress in adolescent smoking prevention. *Children Youth Service Review* 7:1-19.

Tobler, N.S. 1986. Meta-analysis of 143 adolescent drug prevention programs: Quantitative outcome results of program participants compared to a control for comparison group. *The Journal of Drug Issues* 16(4):537-567.

Tobler, N.S. 1992. Drug prevention programs can work: Research findings. *Journal of Addictive Disease* 11(3):1-28.

Tobler N.S. 1997. Meta-analysis of adolescent drug prevention programs: Results of the 1993 meta-analysis. *NIDA Research Monograph*. 170:5-68.

U.S. Department of Health and Human Services. 1991. Strategies to control tobacco use in the United States: A blueprint for public health action in the 1990's. Washington, DC: U.S. Government Printing Office (NIH Publication No. 92-3316).

Zucker D.M. 1990. An analysis of variance pitfall: The fixed effects analysis in a nested design. *Psychological and Educational Measurement and Evaluation* 50: 731-738.

ABSTRACT

Adolescent psychosocial smoking prevention programs have been successful, but limited in the magnitude of program effects. The present study is the secondary analysis after the previous study estimated mean effect sizes in smoking knowledge, attitudes, skills, and behaviors with treatment variables. Regardless of overall program effect estimations that other meta analysis studies have done, this study is conducted to identify explanatory variables that are likely to increase program effects.

A decrease of adolescent smoking behaviors is associated with the following factors:

a. Younger students (5^{th}-7^{th}) than older students (8^{th}-12^{th}).
b. Research methodology using true experimental design, quasi experimental design with equivalence between groups, use of random assignment, 10% or less attrition rate, use of a no treatment control group, high implementation fidelity, and/or acceptable instrumentation reliability
c. Programs using trained peer leaders, targeting cigarette smoking only, implementing 10 or more treatment sessions and/ or providing booster sessions.

Key Words: Meta-analysis, Smoking prevention, Psychosocial, Adolescent, Explanatory factors, Program effectiveness

〈국문초록〉

사회 심리 이론에 근거한 학교 흡연 예방 프로그램의 메타분석: 미국 사례와 Explanatory Variables

청소년을 위한 학교 흡연예방 프로그램은 사회심리 이론에 근거한 프로그램이 대체로 성공적이었다고 알려져 있으나, 각 프로그램의 효과 정도에는 많은 차이가 있다. 이 연구는 다른 메타 분석처럼 전체적인 프로그램 효과도를 측정하여 일반적인 결론을 유도한 것이 아니라, 프로그램의 효과와 관계가 깊은 요인 (Explanatory Variables)을 자세히 파악하여 보건교육 담당자, 연구원, 또는 정책 결정자들에게 구체적인 가이드라인을 제공하는 데에 목적을 두고 있다. 주요한 연구결과는 다음과 같다.

1. 8-12학년 학생들보다는 초등학교에서 중등학교로 바뀌는 5-7학년 학생들에게 흡연예방 프로그램은 더 효과가 있었다.
2. 연구 방법론에 있어서는 experimental design, random assignment, 순수 비교그룹을 사용하였을 경우, implementation fidelity와 instrument reliability가 높은 경우, 또는 10% 미만의 attrition rates일 때 프로그램 효과도 (effect size)가 더 높게 나타났다.
3. 프로그램 실행 시 또래 리더를 사용하였을 경우, 알코올 등 다른 약물을 배제한 담배만을 중점적으로 다루었을 경우, 적어도 10회 이상 연속적으로 이루어지거나 프로그램 종료 후 일년 뒤에 추가 프로그램이 주어진 경우가 더욱 효과적이었다.