Appendix C
Methodology for Confounder Control

C.1 Introduction

Two key analytic approaches are used in this report to investigate possible effects of the Campaign on various outcomes, namely marijuana use, the youth cognitive measures, and the parenting practices measures. One approach is to examine trends in outcomes over time (described in Chapter 4 for marijuana use and in Chapter 5 for the cognitive outcome measures). The other is to examine the dose–response relationship (described in chapters 5 and 6), where the dose is a unit of exposure to anti–drug advertising, and the response is the outcome measure. This type of analysis is called the dose–response approach because it is analogous to a drug study comparing the effects of, say, a 60mg dose, a 40 mg dose, a 20 mg dose, and a placebo. The purpose of this appendix is to describe the methods used in applying this second approach.

One of the problems with assessing trends over time is that the interpretation of change over time in outcomes relies on the assumption that other factors (everything other than the Campaign) affecting the outcomes held steady during the time period. However, it was beyond the scope of this evaluation to determine whether forces external to the Campaign did hold steady. These external forces might include such things as drug prices, drug availability, content of popular media, content of political speech and debate, celebrity actions, and seasonal variations. Consequently, the required assumption of constancy in all other societal forces is a strong one. Furthermore, data collection started after the start of the national phase (Phase III) of the Campaign. Thus, even if one were to accept the strong assumption about other forces holding steady, change in outcomes would reflect only the incremental effect of additional exposure beyond any effect that could have been initially achieved and/or the differential effect of a change in the Campaign. Given these caveats, it is clear a positive trend, while desirable, is insufficient for evaluating the effectiveness of the Campaign. Similarly, a negative trend does not negate the possibility that Campaign effects existed, but countervailing effects from other causes were stronger.

In view of the limitations to the interpretation of trends in the outcomes, considerable reliance is placed in this report on the dose-response approach (in combination with the trend approach). The dose-response approach involves examining trends in outcomes for individuals with different amounts (doses) of Campaign exposure. Suppose that individuals with higher levels of exposure to the Campaign are found to have better outcomes than those with lower levels of exposure. A possible explanation for this finding is that the better outcomes are the result of the higher levels of Campaign exposure. The attribution of the exposure-outcome association to the Campaign requires two key assumptions: (1) that no other variable or variables account for the observed association of exposure and outcome, and (2) that the association is not the result of the outcome causing the exposure rather than vice–versa. The technique used here to address concerns about the first of these assumptions is to control on potential confounding variables in the analysis. The technique used to address the second concern is to examine the associations of exposures at one survey round with outcomes at the next

round; with this delayed effects analysis, the outcome cannot be the cause of the prior exposure. These techniques are described later in this appendix.

Section C.2 discusses the strengths and weaknesses of the dose–response approach. Section C.3 provides more detailed information about the procedures used to implement it. Section C.4 provides detailed technical information on how effects were estimated. Section C.5 provides detailed technical information on how confidence intervals were formed on the effect estimates and how hypothesis testing was conducted.

C.2 Strengths and Weaknesses of the Dose–Response Approach

The dose-response approach is common in the epidemiology of chronic conditions brought on by environmental factors such as coal dust, primary smoking, second–hand smoke, indoor radon gas, and so on. The underlying theory in those disciplines is that if a substance is toxic, then a large dose of it should be at least as toxic as a small dose. If this expected relationship does not hold, the toxicity of the material has not been demonstrated. In the application of this theory to our evaluation of the Media Campaign, the underlying theory is that if advertising is effective, a large dose of consumed advertising should be at least as effective as a small dose. If this relationship does not hold, then the Evaluation generally cannot conclude that the effectiveness of the advertising has been demonstrated.

In dose–response analysis, one must assume that the variation in doses is random after controlling for known factors. In randomized experiments such as clinical trials, random assignment within groups of substantive interest is used to ensure that doses are randomly given. However, since Media Campaign doses are not randomly assigned, but are instead self–chosen by choices in media consumption and filtered through subject's recall, the Evaluation must instead assume that all sources of systematic (nonrandom) variation in doses have been measured.

This is a strong assumption, but as part of the questionnaire design and acquisition of geographic information, the Evaluation team considered a wide range of background variables that might affect dose reception. However, there is always the risk that the questionnaires might not have measured all the predisposing variables. The Evaluation team tried to include as many variables as seemed to be plausible predisposing variables, but limitations on the length of each interview meant information could not be recorded about every plausible predisposing variable.

Even among the set of data collected, some of the data items were not allowed into the "pool of admissible predisposing factors." This was necessary because some of the variables that were measured had an unclear temporal order with the outcomes. Some may be consequences of exposure to Campaign messages. Controlling on such "mediating" variables would be to underestimate Campaign effects. For example, if watching Campaign ads leads youth to change their beliefs about the consequences of marijuana use, and these belief changes lead, in turn, to changes in intentions to abstain from marijuana use (as would occur under the theoretical model described in Chapter 2), then it would be a serious mistake to allow marijuana beliefs into the pool of admissible predisposing factors, even though it is true that beliefs are predisposing factors in developing intentions about marijuana abstention.

Judgment was required to decide which variables were potential mediating variables and which were predisposing variables that were not subject to influence by exposure to the Campaign. There were some variables for which valid arguments were advanced both for classification as a mediator and for

classification only as a confounder. Resolving such conflicts was difficult and of the utmost importance, because each decision potentially affects the evaluation findings. The Evaluation team recognized that other researchers may disagree with these choices. A few of the decisions were extraordinarily difficult to make and are discussed in detail below.

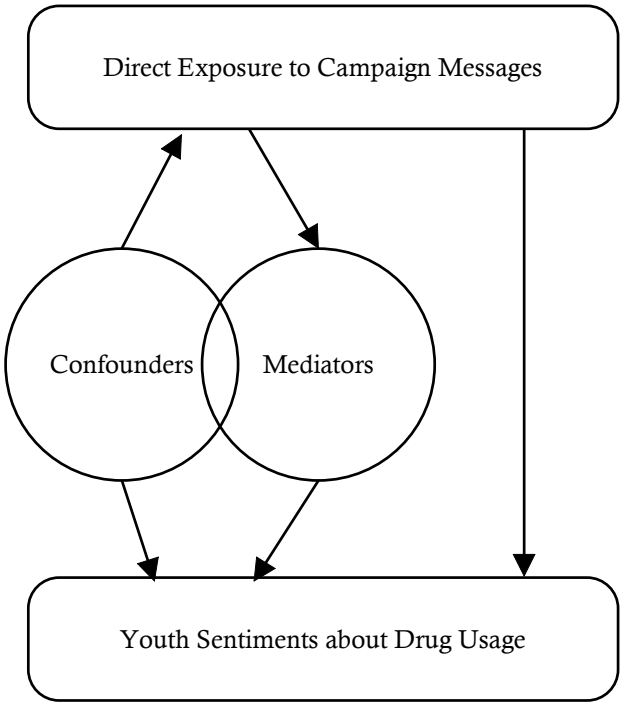
C.3 Admissible Confounder Selection

This section presents the set of variables that the evaluation team admitted into analysis as youth confounders, the set accepted as parent confounders, and concludes with a list of confounders considered as potential moderators. The presentation commences with a brief discussion of the concept of confounding and moderating variables and of the analytic difficulties that arise because some variables may play both roles.

C.3.1 Confounders and Mediators

A large number of cognitive and behavioral variables were obtained on each subject at a single point in time. It is impossible to say with any certainty the order in which these cognitions and behaviors manifested themselves in each subject. Nonetheless, in order to make causal inferences, it is necessary to make some assumptions about this ordering. Figure C–A defines different types of variables schematically.

Figure C–A. Types of variables



A confounder is a variable that leads to variation both in exposure and in outcomes but is itself not caused by exposure or outcomes. This is illustrated in Figure C–A by the directions of the line—confounders cause variation in exposure and cause variation in anti–drug sentiments. In order to avoid false claims of Campaign effects as well as false claims of counterproductive Campaign effects,

it is essential to remove the (confounding) effects of the confounder from the study of the dose–response relationship.

A mediating variable is one that is associated with both exposure and an outcome, as is the case with a confounding variable, but a mediating variable is a result of exposure rather than a cause of exposure. This is illustrated in Figure C–A by the direction of the arrow connecting Mediators and Exposure. In other words, the mediating variable is causally posterior to exposure rather than causally prior to exposure. In order to prevent errors of omission where we do not identify a Campaign effect, it is vitally important that nothing be done to remove the (mediating) effects of the mediator from the study of the dose–response relationship.

Unfortunately, some variables play both confounding and mediating roles. This is illustrated in Figure C–A by the overlap of the circles for confounders and mediators. For variables in this overlap area, we have conflicting imperatives. We must both remove and not remove their effects. As an example of a variable in that overlap, consider the role of cigarette smoking. Cigarette smoking makes it easier to try marijuana and could be related to choices of TV and radio programs and hours of viewing—so it is a confounding variable. At the same time, there may be youth who stopped smoking or were prevented from smoking because of generalized effects of exposure to the Campaign as discussed in Chapter 2.

Thus, it is also a mediating variable. We included items as confounders only when we could be confident that they were not mediators. In the case of cigarette smoking, the issue was resolved by including smoking initiation if it occurred more than 1 year before the date of the interview.

Decisions about which variables would be regarded as potential confounders and which as mediating were made after discussion by a committee of the evaluation team prior to any examination of the data. The committee did not use any of the data about the relationships among the potential confounders/mediators, exposure, and outcomes in making these decisions. Thus the decisions were made blinded to any possible effects on either finding or not finding any effects of the Media Campaign.

C.3.2 Admissible Pool of Youth Confounders

The following variables were judged by the committee to properly belong in the pool of admissible potential confounders for youth. The included variables can be divided into two broad groups. The first group, listed immediately below, includes confounders that directly measure the respondent youth's personal demographics, attitudes, family environment, and behaviors.

1. Age
2. Gender
3. Wave of interview
4. Race/ethnicity
5. Neighborhood characteristics from the Census
6. Urban, suburban, or rural nature of neighborhood
7. School enrollment status in the previous year
8. Whether school was in session in the last 30 days
9. Number of missed schooldays due to illness in the previous 30 days
10. Number of days the youth cut school in the previous 30 days

11. School grade level
12. Academic performance
13. Participation in extracurricular activities[1]
14. Respondent's primary postsecondary plan
15. Hours of TV consumption on weekdays
16. Hours of TV consumption on weekends
17. Hours of radio consumption on weekdays
18. Hours of radio consumption on weekends
19. Internet use
20. Magazine reading habits
21. Language of TV viewing
22. Language of radio programs heard
23. Availability of cable or satellite TV in the household
24. Consumption of specific cable channels targeted by the Media Campaign
25. Personal assessment of family fighting
26. Personal assessment of feelings of family togetherness
27. Degree of parental supervision
28. Respondent's perception of parental knowledge of his or her activities
29. Respondent's perception of parental knowledge of his or her plans
30. Degree of enjoyment of time spent with his or her family
31. Youth rating of the importance of religion in their lives
32. Attendance of religious services
33. Personal antisocial behavior
34. Association with antisocial peers
35. Youth close friends' drug use
36. Personal tobacco use of a long–standing nature
37. Personal alcohol use of a long–standing nature
38. Sensation seeking tendencies.
39. Last completed school year
40. Watched a music channel
41. Watched a sports channel
42. Watched an African American channel
43. Watched a Latino/Hispanic channel
44. Enjoys being with parents
45. Argues with parents
46. Risk score

All of the above reflect youth reports about themselves, their friends, and their families. Some of these variables might be possible outcomes of drug use, and it could be argued that, if the Campaign had reduced drug use, these were posterior to the Campaign not prior to it. However, a majority of the analyses reported in Chapter 5 focus on youth who had not used drugs, thus the concern is reduced. For example, the Campaign might potentially reduce drug use and that might decrease family tension

[1] It has been argued that some of the Campaign advertising in early 2001 may have encouraged youth to join extracurricular activities and thus, that this variable should be treated as a mediator rather than a confounder. This seemed of much less plausibility than a concern that such activities might both affect access to advertising as well as patterns of drug beliefs and use. The committee assumed that participation in extracurricular activities was largely a function of opportunity, physical fitness, other personal traits, accidents of friendship, and parental memories about extracurricular activities.

and increase a feeling of togetherness. Controlling for family togetherness might reduce that apparent dose response effect of the Campaign. However, given that for the majority of the analyses in Chapter 5 only nonusing youth are studied, family togetherness is appropriately seen as a confounder. Still some of these variables, contrary to the Evaluation team's judgment, might be causally posterior to either exposure or outcomes and thus not be true confounding variables.

Youth attendance at anti-drug programs (in or out of school) was excluded as a confounder. There is some risk that youth reports of attendance at such programs might reflect access to Campaign advertising or other outreach efforts, particularly since substantial advertising buys were made on Channel One, an in-school network.

The second broad category included as admissible potential confounders for the youth analysis covers information on parental characteristics and perceptions. These included:

1. Parental age
2. Wave of interview
3. Parental gender
4. Parental marital status
5. Parent has a child aged 9 to 11
6. Parent has a child aged 12 to 13
7. Parent has a child aged 14 to 18
8. Parental income
9. Parental educational attainment
10. Parental religiosity
11. Sharing of parental responsibilities
12. Parental use of the Internet
13. Parental consumption of newspapers
14. Parental consumption of magazines
15. Parental consumption of TV
16. Parental radio consumption
17. Parental consumption of specific cable channels targeted by the Media Campaign
18. The primary language in which the parent watches TV
19. Parental assessment of family togetherness
20. Parental enjoyment of time spent with children
21. Parent's perception of fights with children
22. Parent–child participation in fun indoor activities
23. Parent–child participation in fun outdoor activities
24. Parent's reports on the respondent youth's grade level
25. Parent's report on child's academic performance
26. Parent's report on the time their child spends with friends
27. Parental alcohol use
28. Parental tobacco use
29. Parental prior or current use of hard drugs
30. Parental prior or current use of marijuana
31. Parental prior or current use of inhalants
32. Watched African American or Hispanic TV
33. Watched Hispanic TV

34. Parent's reports of grades taught at school
35. Parent's report of child's grades

As with the youth variables, some of these variables have an ambiguous causal order with respect to outcomes and exposure. The fact that the majority of the associational analysis described in Chapter 5 concern youth who are nonusers of marijuana, strongly mitigates these concerns; but it is possible that youth viewing of advertising aimed at their parents may have influenced family functioning in some way such as decreasing youth resistance to parental monitoring activities. On balance, however, it was thought that it is far more likely that parental monitoring and family functioning would shape youth cognitions about marijuana use. Parent–child talk was not controlled for because of concerns that some of this talk may have been initiated by the youth after viewing Media Campaign ads and thus may be causally posterior to exposure.

Note that many of these parental attributes may be causally prior to *parental* exposure to Media Campaign advertising, but this is irrelevant for study of the association of youth cognitions with direct *youth* exposure.

This distinction between confounders and mediators is a concern for the cross-sectional models but not for the delayed effects models. In addition to the above list of measures, the following measures were included in the pool of admissible variables for the delayed effects youth models:

1. Enjoy being with parents
2. Friends marijuana use
3. Friends illicit drug use
4. Hard to talk to parents about drugs
5. Parent likely to know if using drugs
6. Parent likely to punish if found out using drugs
7. Visiting drug web sites
8. Offers of marijuana
9. Talk to parents about drugs
10. Talk to friends about drugs
11. Attending drug education classes/programs
12. Doing activities with parents
13. Any use of cigarettes
14. Any use of alcohol
15. Any use of inhalants
16. Baseline intentions to use marijuana
17. Own behavior influences child
18. Parent's reports of talking
19. Parent's reports of monitoring
20. Parent's report of likelihood of disciplining
21. Parent's report of writing letter
22. Parent's report of calling radio station
23. Parent's report of attending a meeting
24. Parent's report of joining a group
25. Parent's report of attending workshops on being an effective parent
26. Parent-child participation in fun indoor activities

27. Parent-child participation in fun outdoor activities
28. Parent's report on the time their child spends with friends

C.3.3 Admissible Pool of Parent Confounders

The committee judged that the following variables properly belong to the pool of admissible potential confounders:

1. Race/ ethnicity
2. Parent gender
3. Parent age
4. Parental income
5. Parental marital status
6. Parental religiosity
7. Parent has a child aged 9 to 11
8. Parent has a child aged 12 to 13
9. Parent has a child aged 14 to 18
10. Neighborhood characteristics
11. Urbanity
12. Parental use of the internet
13. Parental consumption of newspapers
14. Parental consumption of magazines
15. Parental consumption of TV
16. Parental radio consumption
17. Parental consumption of specific cable channels targeted by the Media Campaign
18. The primary language in which parents watch TV
19. Parental alcohol use
20. Parental tobacco use
21. Parental prior or current use of hard drugs
22. Parental prior or current use of marijuana
23. Parental prior or current use of inhalants
24. Availability of cable or satellite TV in the household
25. Sharing of parental responsibilities
26. Parental educational attainment
27. Watched African American or Hispanic TV
28. Lowest grade taught at child's school
29. Child's average grades
30. Parent's marijuana use
31. Highest grade taught at child's school

In addition to the above list of measures, the following measures were included in the pool of admissible variables for the delayed effects parent models:

1. Parental assessment of family togetherness
2. Parental enjoyment of time spent with children
3. Parent's perception of fights with children

4. Own behavior influences child
5. Baseline monitoring behavior
6. Baseline monitoring cognitions
7. Baseline talking behavior
8. Baseline talking cognitions
9. Baseline fun activities
10. Baseline youth marijuana use

C.3.4 Confounders as Moderators

A moderator is a characteristic or predisposition that makes respondents more or less susceptible to the Media Campaign. Moderators may cause the effects of the Media Campaign to be different in different subgroups of the population. In this case, there are interactions of Campaign effects with preexisting factors (the moderators). For youth, the moderators examined in this report are:

1. Age of youth
2. Gender of youth
3. Race of youth
4. Hispanic ethnicity of youth
5. Natural sensation–seeking tendencies of youth
6. Youth risk

For parents, the moderators examined in this report are:

1. Age of youth
2. Gender of youth
3. Race of youth
4. Hispanic ethnicity of youth
5. Gender of responding parent
6. Education of responding parent

C.4 Summary of Confounders

There were too many variables in the pool of admissible potential confounders to remove the effects of each individually. Instead, the information was summarized from the pool that tested as relevant. The summarization method is called propensity scoring. The method was introduced by Rosenbaum and Rubin (1983) and is widely used to analyze observational studies (D'Agostino, 1998). It can handle a large number of confounding variables. It is not necessary to develop complex models for all outcome variables, which is an advantage of this method over some of the alternative adjustment methods available. Exposure is conceptualized as a chance event. The probability distribution of exposure varies across people, (i.e., one person may have a high probability of achieving high exposure while others may have only moderate or low chance of doing the same). However, it is assumed that everyone has some chance of achieving every value of exposure. This rules out the existence of subgroups that are constrained to a subrange of the possible values of exposure.

The following discussion starts with a general overview of propensity scoring followed by an examination of the propensity scoring's "balance"—the extent to which the counterfactual projections of population means for the confounding variables vary across exposure levels. The remainder of Section C.4 looks first at the impact of the counterfactual projections on effective sample sizes. It then presents the four cross–sectional models that were fitted on the combined data from Waves 1 to 9—and the four delayed–effects affect models fitted for data for Waves 1 to 9.

C.4.1 Propensity Scoring

Within the group of individuals who have the same exposure propensity, associations between outcome and exposure are free of confounding. This is as if exposure had been randomly assigned to individuals as in a designed experiment. An individual's exposure propensity is estimated as his or her propensity score. Since there are two primary measures of exposure used in this report, two propensity scores were estimated, one for each measure of exposure. An individual's propensity is estimated in terms of confounding variables by complex statistical methods.

Propensity scoring frees the regression modeling process from its usual limitation of reliance on a small number of covariates and simplistic functional forms (e.g., linear main effects only). Rather, a complex model with interactions and higher–order terms can be fit at the propensity scoring stage without concern about overparameterization, since the goal is simply to obtain the best estimated probability of group assignment (in this case to exposure level) from the observed covariates. When subsequently included in the regression model, the propensity score carries all the information from the complex covariate model in a single variable, consuming only one degree of freedom. It also avoids the potentially adverse effects of multicolinearity on the stability of the estimates, regardless of the degree of correlation that exists among the covariates. Finally, propensity score technology can accommodate reasonable numbers of missing observations in the covariates, so fewer cases are lost in analytic procedures requiring complete cases for inclusion.

Despite these advances over traditional regression models, propensity scores have limitations. Like traditional methods for removing group nonequivalence, propensity score methods can adjust only for confounding covariates that are observed and measured. This is always a limitation of nonrandomized studies compared with randomized studies, where the randomization tends to balance the distribution of all covariates, observed and unobserved. However, tests can be devised to determine the robustness of the conclusions to potential influences of unobserved covariates. Such sensitivity analyses suppose that a relevant but unobserved covariate has been left out of the propensity score model. By explicating how this hypothetical unmeasured covariate is related to treatment assignment and outcome, one can estimate how the treatment effect that adjusts for it might change if such a covariate were available for adjustment. Moreover, propensity scores appear to be more robust to certain types of specification error than standard methods. In a simulation to investigate the relative influence of specification error in propensity scores versus regression models, Drake (1993) found that propensity scores are as vulnerable as standard methods to bias from omitted variables, but less vulnerable to bias from variables that are included but in the wrong functional form (e.g., linear rather than quadratic). A second limitation of propensity score methods—that they require reasonably large samples to support the subclassification—will not be a factor here because reasonably large samples are available.

Standard propensity score methods assume that there are only two levels of exposure. However, in this study, exposure is a three– or four–level variable. For this more complex problem, the method

suggested by Joffe and Rosenbaum (1990) was used. With this method, an ordinal logit model is fit for each index. The structure of this model is

$$\ln\left(\frac{\sum_{j\le k} p_{ij}}{1-\sum_{j\le k} p_{ij}}\right)=a_k + X_i\beta\ .$$

Here p_{ij} is the propensity of the *i*–th subject for exposure level *j*, X_i denotes the vector of confounder scores for the same subject, α_k is a threshold parameter for the *k*–th exposure level, and β is a vector of slope parameters with one component for every confounder retained in the model. The point of the modeling exercise is to identify which of the admissible potential confounders are actually predictive of exposure and then to estimate the vector of slope parameters for those predictors.

Eight sets of models were developed for analysis. These included four for cross-sectional analyses and four for longitudinal (delayed effects) analyses. The four sets of models constructed for each type of analysis included a separate set of models for youth and dyads crossed by general and specific exposures.

To fit the initial models, a combination of a theoretically-driven variable specification and stepwise variable selection procedure in SAS was implemented on the set of potential confounders. (The sampling weights were ignored in fitting the model.)[2]

Once the initial models had been fit, the next step was to use the model to remove the effects of the confounding variables from the causal analysis. This was done by following a suggestion by Imbens (2000) with some innovations. The basic suggestion of Imbens was to use the estimated propensities to calculate the expected response across the entire sample, which would be expected in the counterfactual event that everyone in the sample had received the same exposure level. This could be achieved with the estimator

$$\hat{y}_{Ck} = \sum_i \frac{\delta_{ik} y_i}{\hat{p}_{ik}}\ ,$$

where δ_{ik} is an indicator variable for the *i*–th case having exposure level *k*, i.e.,

$$\delta_{ik} = \begin{cases} 1 & \text{if the } i\text{-th individual has observed exposure at level } k, \\ 0 & \text{otherwise} \end{cases}$$

and $\hat{p}_{ik}$ is the estimated propensity the *i*–th individual has for exposure level *k.* Note that for each *i*, $\sum_k \hat{p}_{ik} = 1$ for every i.

One innovation was to project the expected response to the entire eligible population by using the sampling weights. Taking account of the sampling weights is important because of the use of varying

[2] The modeling strategy addressed competing concerns about overfitting the model and not omitting key substantive variables that could confound the relationship between exposure and the outcomes.

probabilities of selection to select youth and parents for the study. As noted in Appendix A, youth 14 to 18 years of age had a higher probability of selection if they had siblings in the 12- to 13- or 9- to 11-year-old age brackets. Similarly, youth with siblings in the same age bracket had lower probabilities of selection than youth with no siblings in the same age bracket. The sampling weights also reflect the weighting adjustments that were made to compensate for differential rates of nonresponse and undercoverage. Using the sampling weights, the counterfactual estimator of response on variable *y* to exposure *k* would be

$$\hat{Y}_{Ck} = \sum_i \frac{\delta_{ik} y_i w_i}{\hat{p}_{ik}},$$

where w_i is the sampling weight for the *i*–th respondent, adjusted for nonresponse and poststratified to population controls. However, it was found that this estimator was unstable and the corresponding counterfactual projection (CFP) weights did not balance the covariates very well (see Section C.4.2). Much better results were obtained by smoothing and calibrating the individual propensities that were estimated by the ordinal logit regression model. The smoothing and calibration was done as follows.

First, the observations were ordered according to the value of $X_i\hat{\beta}$ obtained from the fitted ordinal logit model. The ordered observations were then divided into five mutually exclusive and exhaustive subgroups of approximately equal size denoted by $G_i, i = 1, 2, ..., 5.$ Within each group, smoothed propensities $\tilde{p}_{ik}$ were calculated according to the formula:

$$\tilde{p}_{ik} = \frac{\sum_{j \in G_i} \delta_{jk} w_j}{\sum_{j \in G_i} w_j},$$

where j denotes an observation in group G_i ($i = 1, 2, ..., 5$).

These propensities are smoothed in the sense that there are only five distinct values for each exposure level instead of different values for every study subject as would be the case if the individual propensities estimated by the ordinal logit model were used. The final step was to compute the CFP weight as follows:

$$w_{ik}^{CFP} = \frac{\delta_{ik} w_i}{\tilde{p}_{ik}} \left(\frac{\hat{N}_k}{\sum_j (\delta_{jk} w_j / \tilde{p}_{jk})} \right),$$

where

$$\hat{N}_k = \sum_j \delta_{jk} w_j.$$

The CFP weights are calibrated in the sense that the sum of the CFP weights equals the estimated number of persons in the population that received a given exposure level *k*. In other words, the

estimated exposure distribution based on the CFP weights will equal the corresponding distribution based on the original sampling weights. The calibration property can be expressed mathematically as:

$$\sum_i \delta_{ik} w_{ik}^{CFP} = \hat{N}_k \quad \forall\, k.$$

Using these smoothed and calibrated CFP weights, the counterfactual projection of a population characteristic *y* corresponding to exposure level *k* is given by

$$\tilde{Y}_{Ck} = \sum_i \delta_{ik} w_{ik}^{CFP} y_i.$$

C.4.2 Assessment of Balance

Because propensity scoring is designed to remove the effects of confounding variables from the association between outcomes and exposures, the counterfactual projections of population means for the confounding variables should not vary across the exposure levels. This property is referred to as balance. If a confounder has been successfully balanced, then it will have the same expected counterfactual projection across all exposure levels. Mathematically, this condition of balance is expressed as

$$E\left(\sum_i \delta_{ik} w_{ik}^{CFP} x_{ji} / \hat{N}_k\right) = E\left(\sum_i \delta_{im} w_{im}^{CFP} x_{ji} / \hat{N}_m\right), \text{ for } \forall\ j \text{ and } \forall k \neq m.$$

The same procedures that were implemented in the Fifth Semi-annual Report of Findings and the 2003 Report of Findings to test for balance were also implemented in Wave 9. For all variables in the final model and some variables that were not in the final model but were considered important, as well as for a few key subgroups,[3] WesVar was used to test linear trends and overall differences in the means of the variables across exposure levels for both general and specific exposure. After initial tests of balance, the models were rerun to incorporate variables that were considered to be out of balance using the test of linear trends in means. This process continued until less than 5 percent of the variables were out of balance for the overall sample and each of the subgroups.

C.4.3 Impact of Counterfactual Projections on Effective Sample Sizes

Table C–A contains illustrative information useful to understand how much the counterfactual projection reduces effective sample sizes. Table C–A summarizes the design effects due to the variation in propensities for youth general exposure across all nine waves. They were calculated using the standard Kish approximation (e.g., see Kish, 1965, page 403). The true effective sample sizes will be smaller because the variation in the sampling weights and clustering both will have the effect of increasing total design effects. The counterfactual projection weights did not increase variances appreciably for the groups with medium or high exposure. The increase in variance for the low–exposure group suggests that confounders were identified that successfully predicted who would have low exposure. The result for correcting for self–selection is a 21 percent reduction in the effective

[3] For the cross-sectional models, the subgroups included risk score and wave. For the lagged model the subgroups included marijuana use and wave.

sample size or a 26 percent increase in variances. This was judged to be a good exchange between variance and potential bias.

Table C–A. Design effects and sample sizes for youth across nine waves of data collection by exposure level

General exposure level	Nominal sample size*	Design effect	Effective sample size**
1	4,309	1.26	3,420
2	4.664	1.02	4,573
3	9,802	1.08	9,076

*Youth 12½- to 18-years age included in analysis. Excludes youth with missing general exposure.

**Does not reflect the effects of clustering or standard (non-CPF) weighting.

C.4.4 Detailed Models of Exposure

In this section, the final cross-sectional and delayed-effects models are presented. Four cross–sectional models were fitted, one for each type of parent exposure index and one for each type of youth exposure index. The estimated parameters for the final models are too numerous to present, but a list of the first-order terms in each model is given below along with a brief summary of each of the models.

C.4.4.1 Cross–Sectional Model for the Youth General Exposure Index

The final cross–sectional model for the youth general exposure index included a total of 150 terms. The final model included 65 main effects, 84 interaction effects, and a constant term. After the estimation of the initial model, an additional 11 main effects and 47 interaction terms were included as a result of the tests of balance. The statistically significant first-order variables are presented below in Table C–B.

C.4.4.2 Cross–Sectional Model for the Youth Recall–Aided Exposure Index

The statistically significant first-order variables for the youth recall–aided exposure index are presented in Table C–C. The final cross–sectional model for the youth recall–aided exposure index included a total of 171 terms. The final model included 80 main effects, 90 interaction effects, and a constant term. After the estimation of the initial model, an additional 19 main effects and 60 interaction terms were included as a result of the tests of balance.

C.4.4.3 Cross–Sectional Model for the Parent General Exposure Index

The final cross–sectional model for the parent general exposure index included a total of 128 terms. The final model included 69 main effects, 58 interaction effects, and a constant term. After the estimation of the initial model, an additional 19 main effects and 44 interaction terms were included as a result of the tests of balance. The statistically significant main effects are presented below in Table C–D.

Table C–B. Main effects for cross–sectional model for youth general exposure index among youth aged 12½ to 18

Wave of survey response	Parent watched African American TV**
Youth age**	Parent watched Hispanic TV
Youth's gender	Percent Mexican American ***
School grade level	Percent Cuban American***
Youth does not have an associated parent	Percent of persons under age 18***
Last completed school year***	Percent of households that are linguistically isolated Spanish-speaking households
School enrollment status in the previous year**	Percent of persons age 16+ employed in mining**
Whether school was in session in the last 30 days**	Percent of persons with public assistance income**
Youth's average grades**	Percent of persons below poverty in 1989***
Participation in extracurricular activities**	Percent of persons per vehicle***
Highest grade taught at child's school***	Percent of housing without complete plumbing facilities***
Lowest grade taught at child's school***	Magazine reading habits**
Youth's post-secondary plan is to graduate from a 4-year college**	Language of radio programs heard**
Youth's post-secondary plan is none of the above**	Hours of radio consumption on weekdays
Parental age	Hours of radio consumption on weekends**
Parental gender**	Internet use**
Parental education***	Availability of cable or satellite TV in the household**
Parent has a child aged 14-18**	Hours of TV consumption on weekdays**
Parent's reports on the respondent youth's grade level	Hours of TV consumption on weekends**
Parental marital status is widowed***	Language of TV viewing
Parental race/ethnicity is White	Watched an African American channel**
Parental race/ethnicity is Black	Watched a sports channel**
Parental race/ethnicity is Hispanic	Watched a Hispanic channel
Parental use of the internet	Urban persons living inside urbanized areas **
Parental radio consumption	Metropolitan status is rural
Parental consumption of magazines	Metropolitan status is town
Parental consumption of newspapers	Metropolitan status is city**
Parental consumption of TV	Metropolitan status is suburban**
The primary language in which parents watch TV	Risk score**
Watched a music channel**	Sensation seeking tendencies

**Statistically significant predictors

***Included after tests of balance

Table C–C. Main effects for cross–sectional model for youth specific index among youth aged 12½ to 18

Wave of survey response**	Magazine reading habits**
Youth age	Hours of radio consumption on weekdays
Youth's gender**	Hours of radio consumption on weekends**
Youth does not have an associated parent	Hours of TV consumption on weekdays**
Last completed school year	Hours of TV consumption on weekends**
Whether school was in session in the last 30 days	Watched a music channel**
Participation in extracurricular activities	Watched a sports channel**
Highest grade taught at child's school***	Watched a Hispanic channel**
Lowest grade taught at child's school***	Watched an African American channel**
High school dropouts	Availability of cable or satellite TV in the household
Youth's post-secondary plan is to attend graduate or professional school**	Percent of persons 16+ employed in mining**
	Percent of persons below poverty in 1989***
Youth's post-secondary plan is to graduate from a 4-year college**	Percent of persons with public assistance
	Percent of persons in noninstitutional group quarters **
Parental age	Percent of persons living on farms **
Parental gender	Percent of persons under age 18 below poverty in 1989 ***
Parental education***	Percent of persons with BA plus **
Parent has a child aged 12-13**	Percent of persons working in manufacturing
Parent has a child aged 14-18**	Percent of persons who are Cuban American***
Parent's reports on the respondent youth's grade level***	Percent of persons who are American Indian, Eskimo and Aleut***
Parental race/ethnicity is Black	Percent of persons who are Asian and Pacific Islander**
Parental race/ethnicity is Hispanic	Percent of persons who are White***
Parental race/ethnicity is White	Percent of households that are linguistically isolated Asian and Pacific Islander speaking households***
Parental religiosity***	Foreign born naturalized citizens
Parental marital status is married**	Urban persons living outside urbanized areas**
Income**	Metropolitan status is city
Parental prior or current use of hard drugs***	Metropolitan status is suburban
Parental prior or current use of marijuana***	Metropolitan status is rural
Parental use of the internet	Metropolitan status is town
Parental radio consumption	Military population age 16-64 male and female**
Parental consumption of magazines	Household Units(HU) occupied by renters with no cash rent***
Parental consumption of newspapers	HUs without complete plumbing facilities**
Parental consumption of TV**	Vacant HUs for seasonal, recreational, or occasional use***
Parent watched African American TV	Risk score
Parent watched Hispanic TV	Sensation seeking tendencies**
The primary language in which parents watch TV	
Language of radio programs heard	
Language of TV viewing	
Internet use**	

**Statistically significant predictors

***Included after tests of balance

Table C–D. Main effects for cross–sectional model for parent general exposure index among all parents of youth aged 9 to 18

Wave of survey response**	Percent of persons under age 18**
Highest grade taught at child's school***	Percent of persons under age 18 below poverty in 1989***
Lowest grade taught at child's school**	Percent of persons per room**
Youth's average grades	Percent of persons per vehicle***
Parental age	Percent of households that are linguistically isolated Spanish-speaking households***
Parental gender**	Percent of households with children under age 18 that are headed by female with no husband present***
Parental education	Percent of housing that is vacant***
Parental income**	Percent of households with income above $75,000 per year***
Parental marital status is divorced	Percent of occupied housing that is renter-occupied***
Parental marital status is living as married**	Household Units (HU) occupied by renters with no cash rent **
Parental marital status is married**	Percent of housing that are in large structures with 50 or more HUs**
Parental marital status is separated	HUs without complete plumbing facilities
Parental marital status is widowed	Vacant HUs for seasonal, recreational or occasional use***
Parental race/ethnicity is Black	Percent of persons age 16+ in the labor force who are unemployed
Parental race/ethnicity is Hispanic	Percent of persons age 16+ with farming, forestry and fishing occupations**
Parental race/ethnicity is White	Percent of persons in non-institutional group quarters
Parental religiosity**	Military population age 16-64 male and female**
Parent has a child aged 12-13	Metropolitan status is city
Parent has a child aged 14-18	Metropolitan status is rural
Parent's reports on the respondent youth's grade level***	Metropolitan status is suburban
Parental prior or current use of hard drugs***	Metropolitan status is town
Parental prior or current use of marijuana	Youth's risk score
Parental alcohol use***	
Parental tobacco use**	
Parental use of the internet**	
Parental radio consumption**	
Parental consumption of magazines**	
Parental consumption of newspapers**	
Parental consumption of TV**	
Parent watched African American TV**	
Parent watched Hispanic TV**	
The primary language in which parents watch TV**	
Availability of cable or satellite TV in the household**	
Percent of persons who are American Indian, Eskimo and Aleut**	
Percent of persons who are Black**	
Percent of persons who are Cuban American***	
Percent of persons who are Puerto Rican***	
Percent of persons who are White***	
Percent of persons with BA degree+**	
Percent of persons with public assistance income	

**Statistically significant predictors

***Included after tests of balance

C.4.4.4 Cross–Sectional Model for the Parent Recall–Aided Exposure Index

The statistically significant main effects for the parent recall–aided exposure index are presented in Table C–E. The final cross–sectional model for the parent recall–aided exposure index included a total of 129 terms. The model included 60 main effects, 68 interaction effects, and a constant term. After the estimation of the initial model, an additional 9 main effects and 53 interaction terms were included as a result of the tests of balance.

Table C–E. Main effects for cross–sectional model for parent specific exposure index among all parents of youth aged 9 to 18

Wave of survey response**	Percent of persons who are foreign born naturalized citizens***
Highest grade taught at child's school**	Percent of persons who are foreign born non-citizens***
Lowest grade taught at child's school***	Percent of households that are linguistically isolated Asian and Pacific Islander speaking households**
Parental age	Percent of households that are linguistically isolated Spanish-speaking households
Parental gender**	Percent of households that are other linguistically isolated households**
Parental education**	Percent of households where English language is spoken primarily***
Parental income**	Percent of persons who are American Indians, Eskimos and Aleuts***
Parental marital status is divorced**	Vacant HUs for seasonal, recreational or occasional use**
Parental marital status is living as married	Percent of housing that is vacant***
Parental marital status is married**	Percent of persons per room**
Parental marital status is separated**	Percent of persons 16+ who are employed (military and civilian)***
Parental marital status is widowed	Percent of persons 16+ working in manufacturing**
Parental race/ethnicity is Black**	Percent of persons below poverty in 1989**
Parental race/ethnicity is Hispanic**	Percent of households with income above $75,000 per year**
Parental race/ethnicity is White**	Percent of persons who are institutionalized**
Parental religiosity**	Percent of persons with BA degree+**
Parent has a child aged 12-13**	Percent of persons who are high school dropouts**
Parent has a child aged 14-18	Metropolitan status is city**
Parental alcohol use***	Metropolitan status is rural
Parental tobacco use**	Metropolitan status is suburban
Whether parent has ever used marijuana	Metropolitan status is town
Parental use of the internet**	Youth's risk score
Parental radio consumption**	
Parental consumption of magazines	
Parental consumption of newspapers	
Parental consumption of TV**	
Parent watched African American TV**	
Parent watched Hispanic TV	
The primary language in which parents watch TV**	
Sharing of parental responsibilities**	
Availability of cable or satellite TV in the household**	
Percent of households with children under age 18 that are headed by female household with no husband present**	

**Statistically significant predictors

***Included after tests of balance

C.4.4.5 Delayed–Effects Model for the Youth General Exposure Index

The first-order variables for the delayed–effects model for the youth general exposure index are tabulated in Table C–F. The model included a total of 167 terms. The final model included 101 main effects, 65 interaction effects, and a constant term. After the estimation of the initial model, an additional 28 main effects and 51 interaction terms were included as a result of the tests of balance.

Table C–F. Main effects for delayed-effects model for youth general exposure index among youth aged 12½ to 18

Wave of survey response	Language of TV viewing
Youth age**	Hours of TV consumption on weekdays**
Youth's average grades	Hours of TV consumption on weekends**
School grade level	Watched a Hispanic channel
Youth's race/ethnicity is Black	Watched a music channel**
Youth's race/ethnicity is Hispanic	Watched a sports channel**
Youth's race/ethnicity is White	Watched an African American channel**
Gender	Percent of persons age 9-18***
Last completed school year**	Percent of persons age 16-64
Participation in extracurricular activities	Percent of persons who have BA plus***
Number of days the youth cut school in the previous 30 days**	Percent of households with income above $75,000 per year***
Whether school was in session in the last 30 days**	Percent of housing occupied by renters with no cash rent***
Parental age	Percent of housing that are detached single-family structures***
Parental educational	Percent of housing that are in large structures with 50 or more HUs***
Parental gender	Percent of occupied housing that is renter-occupied***
Parent has a child aged 14-18**	Percent of persons 16+ employed in mining**
Parent's reports on the respondent youth's grade level**	Percent of persons 16-64 who are in the military**
Parental alcohol use	Percent of persons in same house as in 1985**
Parental tobacco use***	Percent of persons per room***
Parental consumption of magazines	Percent of persons per vehicle***
Parental consumption of newspapers	Percent of persons who are high school dropouts***
Parental radio consumption	Percent of persons who are institutionalized
Parental consumption of TV	Percent of persons who are urban and live inside urbanized areas**
The primary language in which parents watch TV	Percent of persons who are urban but live outside urbanized areas
Parent watched African American TV	Percent of persons who live in noninstitutional group quarters***
Parent watched Hispanic TV	Percent of persons who live on farms***
Parental use of the internet	Percent of persons with public assistance income**
Parents report of attending a meeting***	Percent of households that are linguistically isolated Asian and Pacific Islander speaking households***
Parents report of attending workshops on being an effective parent***	Percent of households that are linguistically isolated Spanish-speaking households***
Parents report of joining a group***	
Parents report of likelihood of disciplining	
Parents report of writing letter**	
Parent likely to punish if found out using drugs**	
Degree of parental supervision	
Availability of cable or satellite TV in the household**	

Table C–F. Main effects for delayed-effects model for youth general exposure index among youth aged 12½ to 18 (continued)

Respondent is home schooled or school does not give grades	Percent of households where English language is spoken primarily**
Respondent's perception of parental knowledge of his or her plans***	Percent of persons who are American Indian, Eskimo and Aleut***
Personal antisocial behavior**	Percent of persons who are Cuban American**
Personal assessment of family fighting***	Percent of persons who are Mexican American **
Talk to friends about drugs**	Percent of persons who are Puerto Rican***
Talk to parents about drugs**	Percent of persons who are foreign born naturalized citizens***
Use of marijuana	Metropolitan status is city**
Baseline intentions to use marijuana	Metropolitan status is rural
Time with friends no adults around**	Metropolitan status is suburban
Hard to talk to parents about drugs**	Metropolitan status is town
Argues with parent	Youth's risk score
Association with antisocial peers**	Sensation seeking tendencies
Attending drug education classes/program	
Offers of marijuana **	
Visiting drug websites	
Hours of radio consumption on weekdays	
Hours of radio consumption on weekends**	
Language of radio programs heard**	
Magazine reading habits**	
Internet use**	

**Statistically significant predictors

***Included after tests of balance

C.4.4.6 Delayed–Effects Model for the Youth Recall–Aided Exposure Index

The first-order variables for the delayed–effects model for the youth specific exposure index are tabulated in Table C–G. The model included a total of 151 terms. The final model included 80 main effects, 70 interaction effects, and a constant term. After the estimation of the initial model, an additional 9 main effects and 52 interaction terms were included as a result of the tests of balance.

C.4.4.7 Delayed–Effects Model for the Parent General Exposure Index

The main effects for the delayed–effects model for the parents general exposure index are tabulated in Table C–H. The final delayed–effects model for the parent general exposure index included a total of 72 terms. The final model included 50 main effects, 21 interaction effects, and a constant term. After the estimation of the initial model, an additional 9 main effects and 21 interaction terms were included as a result of the tests of balance.

Table C–G. Main effects for delayed-effects model for youth specific index among youth aged 12½ to 18

Wave of survey response**	Attendance of religious services**
Youth age**	Availability of cable or satellite TV in the household**
Gender**	Hours of radio consumption on weekdays
Youth's race/ethnicity is Black	Hours of radio consumption on weekends
Youth's race/ethnicity is Hispanic	Hours of TV consumption on weekdays**
Youth's race/ethnicity is White	Hours of TV consumption on weekends**
Highest grade taught at child's school	Internet use**
Lowest grade taught at child's school	Language of radio programs heard
Last completed school year***	Language of TV viewing
Whether school was in session in the last 30 days	The primary language in which parents watch TV
Youth's average grades***	Watched a Hispanic channel
Youth's post-secondary plan is to graduate from a 4-year college**	Watched a music channel**
Parental age	Watched a sports channel
Parental gender	Watched an African American channel***
Parent has a child aged 9-11**	Attending drug education classes/program
Parental marital status is married	Participation in extracurricular activities
Parental consumption of magazines	Visiting drug websites**
Parental consumption of newspapers	Use of marijuana
Parental consumption of TV**	Percent of households with income above $75,000 per year**
Parent watched African American TV	Percent of housing that are detached single-family structures**
Parental radio consumption	Percent of housing that is vacant**
Parental use of the internet	Percent of housing without complete plumbing facilities**
Magazine reading habits**	Percent of persons age 16+ employed in mining**
Parents report of attending workshops on being an effective parent**	Percent of persons age 16+ with farming, forestry and fishing occupations**
Parents report of joining a group***	Percent of persons age 16-64 who are in the military**
Parents reports of monitoring***	Percent of persons age 16-64**
Parent likely to punish if found out using drugs**	Percent of persons below poverty in 1989**
Hard to talk to parents about drugs**	Percent of persons in same house as in 1985**
Degree of enjoyment of time spent with his or her family***	Percent of persons who are rural but do not live on farms**
Sharing of parental responsibilities***	Percent of persons who are urban and live inside urbanized areas**
Talk to parents about drugs	Percent of persons with public assistance income**
Talk to friends about drugs**	Percent of persons who are Asian and Pacific Islander**
Time with friends no adults around**	Metropolitan status is city
Respondent's perception of parental knowledge of his or her plans	Metropolitan status is rural**
Respondent is home schooled or school does not give grades***	Metropolitan status is suburban
Personal assessment of family fighting**	Metropolitan status is town**
Association with antisocial peers	Youth's risk score**
	Sensation seeking tendencies

**Statistically significant predictors

***Included after tests of balance

Table C–H. Main effects for delayed–effects model for parent general exposure index among all parents of youth aged 12½ to 18 at followup wave

Wave of survey response	Parental consumption of TV**
Youth age	Availability of cable or satellite TV in the household**
Youth's race/ethnicity is Black	The primary language in which parents watch TV**
Youth's race/ethnicity is Hispanic	Parent watched African American TV**
Youth's race/ethnicity is White	Parent watched Hispanic TV**
Lowest grade taught at child's school***	Parent's perception of fights with children**
Parental age	Parents reports of talking**
Parental gender	Baseline fun activities**
Parental education**	Baseline talking cognitions**
Parental income**	Percent of households with children under age 18 that are headed by female household with no husband present***
Parental marital status is divorced	Percent of occupied housing that is renter-occupied***
Parental marital status is living as married**	Percent of persons in same house as in 1985**
Parental marital status is married**	Percent of persons with public assistance income***
Parental marital status is separated	Percent of persons below poverty in 1989***
Parental marital status is widowed	Percent of persons 16+ employed in mining**
Parent has a child aged 12-13	Percent of persons 16-64 who are in the military**
Parent has a child aged 14-18	Percent of persons who have BA degree+**
Parent has a child aged 9-11	Percent of persons who are American Indian, Eskimo and Aleut***
Parental alcohol use**	Percent of persons per vehicle**
Parental prior or current use of marijuana**	Use of marijuana
Parental tobacco use**	Metropolitan status is city
Parental use of the internet**	Metropolitan status is rural
Parental consumption of magazines**	Metropolitan status is suburban
Parental consumption of newspapers**	Metropolitan status is town
Parental consumption of radio**	Youth's risk score

**Statistically significant predictors

***Included after tests of balance

C.4.4.8 Delayed–Effects Model for the Parent Recall–Aided Exposure Index

The final delayed–effects model for the parent specific exposure index included a total of 82 terms. The final model included 57 main effects, 24 interaction effects, and a constant term. After the estimation of the initial models, an additional 11 main effects and 24 interaction terms were included as a result of the tests of balance. The first-order variables for the delayed–effects model for the parents specific exposure index are tabulated in Table C–I.

Table C–I. Main effects for delayed–effects model for parent–specific exposure index among all parents of youth aged 12½ to 18 at followup wave

Wave of survey response	Parent watched African American TV**
Youth age	Parent watched Hispanic TV**
Youth's race/ethnicity is Black	Parent's perception of fights with children**
Youth's race/ethnicity is Hispanic	Parents reports of talking**
Youth's race/ethnicity is White	Baseline fun activities**
Lowest grade taught at child's school***	Baseline talking cognitions**
Parental age	Percent of households with children under age 18 that are headed by female household with no husband present***
Parental gender	Percent of occupied housing that is renter-occupied***
Parental education**	Percent of persons in same house as in 1985**
Parental income**	Percent of persons with public assistance income***
Parental marital status is divorced	Percent of persons below poverty in 1989***
Parental marital status is living as married**	Percent of persons 16+ employed in mining**
Parental marital status is married**	Percent of persons 16-64 who are in the military**
Parental marital status is separated	Percent of persons who have BA degree+**
Parental marital status is widowed	Percent of persons who are American Indian, Eskimo and Aleut***
Parent has a child aged 12-13	Percent of persons per vehicle**
Parent has a child aged 14-18	Use of marijuana
Parent has a child aged 9-11	Metropolitan status is city
Parental alcohol use**	Metropolitan status is rural
Parental prior or current use of marijuana**	Metropolitan status is suburban
Parental tobacco use**	Metropolitan status is town
Parental use of the internet**	Youth's risk score
Parental consumption of magazines**	
Parental consumption of newspapers**	
Parental consumption of radio**	
Parental consumption of TV**	
Availability of cable or satellite TV in the household**	
The primary language in which parents watch TV**	

**Statistically significant predictors

***Included after tests of balance

C.5 Testing for Significance of Counterfactual Effects

Both visual and technical approaches were employed to assess the significance of estimated effects. The actual mean on each outcome for the weighted sample and all of the counterfactual means for each exposure group were displayed with their confidence intervals and were available for visual inspection. A more technical approach was to adapt a test of the Gamma statistic of significance for monotonic relationships. The monotonic dose–response test assessed the overall association between exposure and outcome. In calculating the gamma statistic, the variance introduced by complex sample design, nonresponse adjustment, and counterfactual projection were reflected as fully as possible.

C.5.1 Estimating Variances on Counterfactual Projections

Replicate weights had been prepared for variance estimation of ordinary survey statistics as explained in Appendix A. There are 100 of these replicate weights for every subject. The process of adjusting the standard survey weights for counterfactual projection was partially repeated on each set of replicate weights. As explained in Section C.4.1 of this appendix, there were four major steps in this process. The first was to model exposure. The second was to create a partition of the data set based on the values of $X_i\hat{\beta}$. The third was to estimate the exposure propensity within each cell of the partition for each of the different exposure levels. The fourth was to apply the inverse of these estimated propensities to the sampling weights. To estimate the variances of the counterfactual projections, only the third and fourth steps were replicated. Ideally, all the steps would have been replicated, but technical issues made this infeasible. As a result, the variance estimates are likely to be a little too small and the confidence intervals a little tighter than they should be.

The reason for this is that confidence intervals do not reflect the uncertainty due to selecting the most important predictors of exposure. Different samples would no doubt have resulted in different choices of which variables to include in the ordinal logit model. However, the extra uncertainty introduced by model selection among the variables considered is probably small. Note that the confidence intervals are also conditioned on the assumptions made about exposure. If there were important covariates that were omitted from the modeling process because they were never asked in the questionnaire, the confidence intervals will not provide the 95 percent coverage promised.

Let w_{itr} be the *r*–th replicated counterfactual weight for the *t*–th exposure level for the *i*–th observation. Let w_{it0} be the full sample counterfactual weight. Note that these weights are equal to zero for the *i*–th observation unless the *i*–th observation actually experienced the *t*–th exposure level. Let δ_{it} be an indicator flag for the *t*–th exposure level for the *i*–th observation. A unified set of counterfactual weights was then created by stacking these weights according to

$$w'_{ir} = \sum_k \delta_{ik} w_{ikr} \text{ and } w'_{i0} = \sum_k \delta_{ik} w_{ik0} .$$

The counterfactual mean for some outcome *y* on some class *c* indicated by ε_{ci} and exposure level *t* is then

$$\hat{y}_{ct} = \frac{\sum_i w'_{i0}\delta_{it}\varepsilon_{ci}y_i}{\sum_i w'_{i0}\delta_{it}\varepsilon_{ci}} \text{ with variance estimate } \operatorname{var}\hat{y}_{ct} = \sum_r b_r \left(\frac{\sum_i w'_{ir}\delta_{it}\varepsilon_{ci}y_i}{\sum_i w'_{ir}\delta_{it}\varepsilon_{ci}} - \frac{\sum_i w'_{i0}\delta_{it}\varepsilon_{ci}y_i}{\sum_i w'_{i0}\delta_{it}\varepsilon_{ci}} \right)^2 ,$$

where the b_r are factors chosen to correspond to the replication method.

C.5.2 Testing for a Monotonic Dose–Response Relationship

A standard nonparametric estimate for a linear relationship is the Gamma statistic. It is appropriate for testing whether two ordinal variables have a monotonic relationship to each other. It does not require that the response (outcome) variable have a normal distribution, as is the case in standard analysis of variance procedures. This is important in this report because the outcomes of interest are generally not normally distributed. In this application, a monotonic relationship is a relationship such that as the level of exposure increases, the level of the outcome variable moves in one direction only.

There is no requirement that the outcome rise linearly or steadily. It can rise in jerks and pauses, but there can be no reversals. In terms of the cognitive processes, it is assumed that extra exposure to advertising will either have an effect or not have an effect, but that the direction of the effect will never reverse. Although it might be possible to imagine a situation where light exposure is beneficial while heavy exposure actually has the opposite of the desired effect, this does not seem plausible in general.

In the Wave 4 report, two statistics were used in the cross–sectional and delayed–effects association tables: the Jonckheere–Terpsta (JT) test of monotonicity for significance testing and the Spearman rank correlation coefficient (*rho*) to represent strength of association or effect size. It was apparent from the reviews that this engendered some confusion, with some readers thinking that the significance level applied to the *rho* value, which it did not. Beginning in Wave 5, the significance testing and effect size statistics were unified by using *gamma* for both purposes. Both the gamma and the JT are similar in many respects—they are both nonparametric tests that do not make strong distributional assumptions (see Nadimpalli, Judkins and Zador, 2003). The choice to report the gamma statistic instead of the JT test is driven primarily by convenience: the gamma measure is more easily interpretable than the JT statistic. Gamma is a symmetric measure whose values range, like *rho*, from –1 to 1. Under statistical independence, the gamma statistic will be 0.

SAS has an option to use a weight in calculating the Gamma test. This feature was used. If a subject has a weight of W, using the weight has the same effect on the calculations as if W copies of the subject were included in the database. Since the weights were in the tens of thousands, SAS perceives the sample size as being much larger than it really is and returns inappropriate significance levels. This was corrected by replicating the Gamma.

Let Γ_0 be the value of the Gamma produced by SAS using the full sample counterfactual weights w'_{i0} and Γ_r be the value of the Gamma produced by SAS using the r–th replicated counterfactual weights w'_{ir}. The variance on the Gamma statistic was calculated as

$$v = \sum_{r=1}^{100} b_r \left(\Gamma_r - \Gamma_o\right)^2 .$$

The corrected test for significance of Gamma is then given as

$$\Gamma_C = \frac{\Gamma_0}{\sqrt{v}} .$$

Under the null hypothesis that there is no relationship between exposure and the outcome, the statistic Γ_C has an approximate t–distribution with 100 degrees of freedom. So the alternative hypothesis of a monotonic relationship between exposure and outcome is accepted if Γ_C >1.98.

References

D'Agostino, R.B., Jr. (1998). Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non–randomized control group. *Statistics in Medicine,* 17, 2265–2281.

Drake, C. (1993). Effects of misspecification of the propensity score on estimators of treatment effect. Vol. 49, 1231–36. Imbens, G. W. (2000). The role of the propensity score in estimating dose–response functions. *Biometrika*, 87, 706–710.

Imbens, G.W., (2000). The role of the propensity score in estimating dose–response functions. *Biometrika*. 87. 706–710.

Joffe, M. M., and Rosenbaum, P. R. (1990). Propensity scores. *American Journal of Epidemiology*, 150, 327–333.

Kish, L. (1965). *Survey Sampling*. New York: J. Wiley & Sons.

Nadimpalli, V., Judkins, D., and Zador, P. (2003). Tests of Monotone Dose–Response in Complex Surveys . *Proceedings of the American Statistical Association.*

Rosenbaum, P.R., and Rubin, D.B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association,* 79, 516–524.

Rosenbaum, P.R., and Rubin, D.B. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika,* 70, 41–55.