Appendix

Analysis Methods for Testing the Interactive or Additive Effects of the Combined Interactive R01 Projects Addressing School Students and their Parents

Design of Integrated Projects

The Parent Project is a 3 (treatment groups) X 4 (occasions) design with repeated measures over the second factor. The School Project is also a 3 (treatment groups) X 4 (occasions) design with repeated measures over the second factor. These two projects can be combined to form a 3 (parent intervention groups) X 3 (adolescent intervention groups) X 4 (occasions) design with repeated measures over the third factor. Figure 1 illustrates this design. The treatment will occur during the first two years of the study and the third year will be a no treatment follow-up in both studies. This design will integrate the individual projects by investigating both the individual and combined effects of the two intervention channels. There will be two separate analyses, one focusing on the outcome data for the parent channel and the other focusing on the outcome data for the school channel. This design will allow us to investigate the combined effects of the two channels. There are three potential outcomes possible, all of which are of scientific and practical interest. This same interpretation can be made whether analyzing the parent or adolescent data. First, the two interventions (parent and student) could produce no interaction effects; i.e., both exhibit independent effects and the impact of the two interventions is *additive*. For the parent data, this would mean that the intervention on the students has no impact on the intervention on the parents. Second, the two series of interventions could produce an interaction that is *synergistic*; i.e., the combined interventions are more effective than the additive effects of the individual interventions. For the parent data, this would mean that the intervention on the students enhances the impact of the interventions on the parents. Third, the two series of interventions could produce an interaction that is conflicting; i.e., the two interventions could interfere with each other. For the parent data, this might mean that the intervention on the students decreases the impact of the intervention on the parents. These analyses will be conducted contrasting each of the intervention levels of the Parent project with each of the intervention levels of the School project for each of the six cancer risk behaviors.

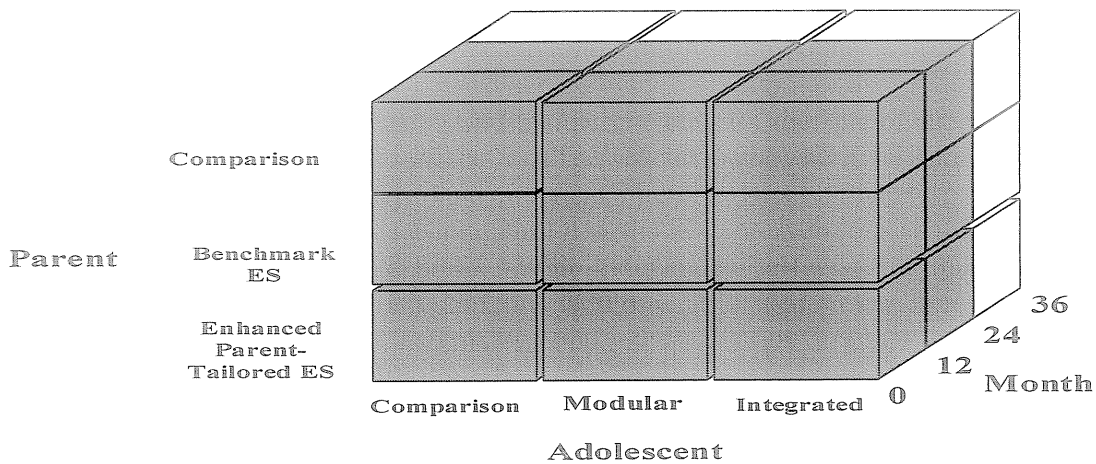

Figure 1. The overall design for the combination of the Adolescent Project and the Parent Project.

Statistical Analyses for the Integrated Project: Outcome

The overall integrated project design is a 3 (parent treatment group) X 3 (adolescent treatment group) X 4 (occasions) design with two sets of dependent repeated measures over the third factor (the parent data and the student data). The data management task will be to match adolescents and parents. Since multiple outcome and/or intermediate variables are available, the type of statistical analysis to be performed will be determined by the nature (categorical or continuous) of each variable and the number of dependent measures selected. The classification of the analyses by primary and secondary outcomes reflects the wide acceptance and history of use of the primary dependent measures. Secondary outcomes are characterized by greater sensitivity, but are without a long record of extensive use. All of these outcome variables are described in greater detail in the individual project proposals.

Multivariate analysis of variance (MANOVA) with repeated measures will be used for the initial examination of outcome differences for variables that are continuous in nature. Such analyses will treat the separate projects (adolescent and parent) as independent variables and the separate project treatments (Comparison, Modular, & Integrated in the School Project; Comparison, Benchmark Expert System, and Enhanced Parent-Tailored Expert System in the Parent Project) as levels of the independent variables. In addition, specialized design applications of Generalized Estimating Equation (GEE) and random effects regression analyses, which can model separate, combined and interactive treatment effects over time, will be used for analyzing outcome differences that are either dichotomous or continuous in nature. More specialized between projects analyses will also be conducted using both Latent Transitions Analysis and Structural Equation Modeling. These methods are not described in detail here.

Generalized Estimating Equation (GEE) Analysis and Random Effects Models

Initial analyses of each outcome variable will assess the efficacy of the intervention using repeated measures regression analyses under the Generalized Estimating Equation (GEE) method (Zeger & Liang, 1986). GEE is a powerful and versatile procedure for analyzing discrete and continuous longitudinal data under minimal assumptions about time dependence. It provides estimates of population averaged effects, and is especially advantageous when the objective is to make inferences about group differences. It enables use of linear, logistic and Poisson regression methods with repeated measures, and provides consistent estimates of regression coefficients and robust variance estimates, even in the presence of unbalanced group data. Even though the assumptions regarding the working correlation matrix in any regression model may be incorrect, use of the robust variance estimator proposed by Zeger and Liang (1986) will ensure that we have an accurate test of the intervention effect.

Repeated measures analyses will be conducted with logistic regression for dichotomous outcomes (logit link function) and with linear regression with normally distributed continuous outcomes (identity link function). These analytic models will include dummy variables representing the intervention groups, time point, and intervention group X time point interaction terms. The efficacy of the intervention will be determined by the test of statistical significance of a time by treatment interaction effect, indicating, for example, that subjects in the enhanced intervention group show greater behavior change over the course of the study than in the control group. In the GEE procedure, the continuous or dichotomous dependent variable with the proper link function will be regressed first against treatment group, and other individual-level covariates such as age, gender, race, and education. Subsequent analytic models will examine gender specific treatment effects.

Both GEE and random effects models are extensions of models for independent observations to time-dependent data, and both can be used to analyze binary or continuous outcomes longitudinally. GEE models are desirable when the research focus is on differences in population averaged response over time (i.e., treatment vs. control or comparison group differences at follow-up), and random effects models are appropriate when the emphasis is on changes in individuals' behavior across time (Hu et al., 1998; Laird & Ware, 1982; Park, 1993). Random effects models have a key advantage for handling missing data because subjects are not assumed to be measured at the same number of time points, thus subjects with missing data on the dependent variable are not excluded from the analysis. In addition, random effects models are also less restrictive with respect to missing data assumptions than GEE and, realistically, allow missingness to depend upon an individual's previously observed values of the dependent variable (Hedeker & Gibbons, 1997; Little, 1995). Since both types of analyses are of interest, both types of logistic models will be applied. Analyses of primary treatment vs. comparison group differences and tests of the interaction between projects will utilize GEE models. Joint distribution analyses of rates of behavior change over time accounting for the missing data mechanism will use random effects models.

For example, primary outcomes which are measured categorically, such as being in maintenance versus other stages for dietary fat, will be evaluated using GEE with a logit link function or the random effects logistic model. Separate logistic models will be constructed for each dependent variable (Y) measured at the final follow-up, and alternative coding schemes will be used to address specific comparisons of the intervention components of the two Interactive R01 projects. A general example of a model for the Parent Project is as follows:

$$\text{logit pr}(Y=1 \mid x_{1,}\ x_{2,\ldots,}\ x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \ldots + \beta_k x_k$$

where: Y = 1 if subject is in maintenance stage for dietary fat at t_{36}
0 if subject is not in maintenance stage at t_{36}

x_1 = 1 if the subject received the Parent Benchmark Expert System (Parent Treatment group 1)
0 if the subject was in the comparison group

x_2 = 1 if the subject's child received the School based Integrated intervention (School Treatment group 1)
0 if the subject's child was in the comparison group

x_3 = 1 if the subject received the parent Enhanced tailored System intervention (Parent Treatment group 2)
0 if the subject was in the comparison group

x_4 = if the subject's child received the school based Modular intervention (School Treatment group 2)
0 if the subject's child was in the comparison group

x_{5-8} = a series of dummy variables representing the four interaction terms for each level of the Interactive R01 project interventions: $x_5 = x_1 * x_2$; $x_6 = x_1 * x_4$; $x_7 = x_3 * x_{2;}$; $x_8 = x_3 * x_4$

x_{9-11} = a series of dummy variables representing stage of change for dietary fat at baseline

x_{12-14} = individual level covariates (e.g., age, educational level, sex)

x_{15-17} = a series of dummy variables representing linear, quadratic, and/or spline effects of time

x_{18-k} = a series of dummy variables representing the interaction of time X treatment group

Evaluation of the effectiveness of the interventions upon stage of dietary fat and other dichotomous dependent variables will be done for each population studied, and for the above example β_1 and β_3 will be the parameters of primary interest for the Parent project, whereas β_2 and β_4 will be the parameters of primary interest for the School project. The estimate of these parameters will be exponentiated to provide the multivariate adjusted odds of reaching maintenance for dietary fat for individuals who receive each intervention after control for baseline stage of change and individual covariates, and confidence intervals will be calculated.

The efficacy of the intervention at follow-up is determined by the statistical significance of a positive time X intervention interaction effect in the above model, β_{18-k}. That is, participants in the intervention arm would show greater behavior change over the course of this study than in the comparison arm. The test statistic is the corresponding interaction parameter estimate divided by its standard error from the robust variance estimate from GEE and compared to a normal distribution. For the outcomes measured on 4 occasions, we will examine whether this time trend is linear, or whether behavior change rates increase or level off. There is some evidence from unpublished data from the current parent-school project that behavior change among early stage individuals at baseline may be delayed and thereby occur in a nonlinear fashion. An understanding of this trend is critical for future prevention trial design and evaluation, and nonlinear curve fitting methods (e.g., use of fractional polynomial or spline terms for time) will be appropriate (Greenland, 1995a, 1995b).

The effectiveness of the combined adolescent and parent interventions can be assessed by considering the interaction terms of the treatment variables from the two projects (e.g. β_7= Parent Enhanced treatment X Adolescent Integrated treatment) as the parameter of primary interest. This coefficient will allow determination of whether the effects of each project specific intervention are modified by the other intervention. In the dichotomous case, multivariate adjusted odds ratios will be calculated for individuals receiving both interventions as compared to one intervention, and to no intervention, and confidence intervals will be calculated.

Rothman (1986) provides an alternative coding scheme for testing the interaction of multiple treatments which can be more informative than the traditional dummy coded treatment terms illustrated above. Used in a logistic regression model, this alternative coding yields several statistical tests of effect of the treatment interactions, including tests for additivity, synergy and attributable proportion of effect due to interaction (Hosmer & Lemeshow, 1992; Rothman 1986). These analyses provide a test for the primary hypotheses and, in addition, provide an estimate and confidence interval of the intervention effects for the two Interactive R01 projects.

Statistical Power to Test R01 Interactive Effects

After attrition, the Parent project is estimated to have approximately 1,856 respondents at t_{36}. Since the number of adolescents enrolled in the School project will be greater than the number of parents enrolled in the Parent project, and since these analyses require a matched parent-student dataset, the size of the sample for testing interactive effects is limited by the size of the Parent project sample. Given this, each of the nine treatment cells of the Parent Treatment X School Treatment interaction will contain approximately N=200 parent-student pairs. This sample is sufficient to enable detection of even fairly small standardized effect sizes (Cohen 1988; Rossi, 1990) for many pair-wise cross project comparisons of continuous outcome measures. For example, a simple 3 X 3 analysis of variance with $N = 1856$ and $\alpha = .05$ has power > .90 for a small effect size (Cohen's $f = .10$; $\eta^2 = .01$). For dichotomous outcomes, somewhat larger standardized effect sizes may be needed to achieve comparable levels of power. Although the two projects may not have sufficient sample size to enable the detection of all project specific comparison groups, there is sufficient statistical power to provide some important information on many exploratory analyses of the efficacy of the joint contributions of the Parent and Student interventions.