APPENDIX I

Analytic Approach for Regression Analysis

As a first step in building multivariate linear or logistic regression models, bivariate models are run, including only one independent variable in each model. These models estimate the crude, or unadjusted, effect of each independent variable on the outcome variable. Those independent variables that have a statistically significant bivariate relationship with the outcome variable ($p < 0.10$) are considered for inclusion in the multivariate model. Independent variables that do not have statistically significant bivariate relationships with the outcome variables are still often tested for inclusion in the multivariate model based on theory alone.

Program Outcomes: The goal of these analyses was to identify characteristics of youth who completed the program and had good program outcomes. Program outcomes were represented by filings while enrolled in the program, program recidivism, education status at the time of discharge, change in the CCAR substance use score from admit to, and change in the average CCAR overall problem severity score from admit to discharge from the program. For each site, statistical models were built in order to determine groups of characteristics of the youth measured before they were admitted to the program, or at the time they were admitted, which predicted these outcomes. Linear regression was used to predict change in the CCAR substance use score and change in the average CCAR overall problem severity score from admit to discharge from the program because each was measured as a continuous variable with a normal distribution. Logistic regression was used to predict program completion, having at least one filing while enrolled in the program, education status at the time of discharge, and program recidivism. Only the 99 youth (52 in Denver, 47 in Sterling) who have been discharged from the program were included in these models.

12-Month Outcomes: The goal of these analyses was to identify characteristics of youth who: (1) had high costs during the 12 months following discharge from the program, (2) produced a cost savings, with the costs during the 12 month follow-up period being compared to the costs accumulated before being admitted to the program, (3) accumulated more days in residential treatment during the 12 months following discharge from the program, and (4) accumulated more commitment days during the 12 months following discharge from the program. For each site, statistical models were built in order to determine groups of characteristics of the youth measured before they were admitted to the program that predicted these four outcomes. Linear regression was used to predict cost, days in residential treatment, and number of commitment days during the 12 months following discharge because each was measured as a continuous variable with a normal distribution [1]. Logistic regression was used to predict cost savings versus no cost savings.

[1] Models of cost often use a transformation of the cost as the outcome variable because the distribution is often non-normal (i.e., skewed). However, a linear regression model was found to be appropriate for these cost data because the distribution did not differ significantly from normal.