RAND Drug Policy Research Center

A JOINT ENDEAVOR WITHIN RAND HEALTH AND RAND PUBLIC SAFETY AND JUSTICE

Estimating the Causal Effect of Time in Treatment Using Propensity Scores

Greg Ridgeway (RAND Corporation, Santa Monica, CA) • Daniel McCaffrey (RAND Corporation, Pittsburgh, PA) • Andrew R. Morral (RAND Corporation, Arlington, VA)

ABSTRACT

Length of stay (LOS) in treatment is often a good predictor of drug abuse treatment outcomes, but it is not clear whether this association results from a causal effect of treatment length, as some have concluded, or whether treatment length and outcomes are both functions of some third variable, like pretreatment motivation to overcome substance abuse problems. Identifying the causal effect of LOS is a high priority for planning cost-effective treatment services. We examined the causal effects of LOS on a measure of recovery from substance abuse using pretreatment features and 12-month outcomes collected on 486 youths between the ages of 13 and 18 in outpatient programs in the Cannabis Youth Treatment (CYT) study. A naïve comparison at 12-month follow-up would imply that the rate of recovery for those with a 30-day LOS is the same as those with a 90-day LOS, 33% and 30% respectively. To adjust for pretreatment characteristics we adopt a propensity score strategy to estimate what would happen to those youths who left a treatment program at 30 days if they had continued for a lengthier treatment.

We find that for subjects like those that stay in treatment for 30 days, additional days of treatment would increase the likelihood of recovery from substance abuse. However, few subjects with 90-day stays look like those with 30 days stays. As a result the standard error of the treatment effect estimates are large outside a small range around LOS = 30. We conclude that failing to adjust adequately for factors that affect both LOS and outcomes lead to conclusions that are biased and understate the uncertainty in the treatment effect estimates.

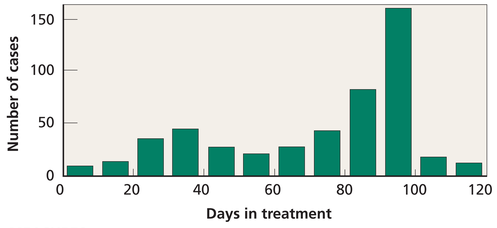
GOALS

- **Estimate** the causal effect of a subject's time-in-treatment on measures of recovery.
- **Demonstrate** a process for estimating the time-in-treatment effect that removes the effects of factors that influence both length of stay and outcome measures.

DATA

- **Source:** Youths in CSAT's Cannabis Youth Treatment (CYT).
  - Four-site experiment that randomized 600 youths to one of six manualized psychosocial interventions, all designed to last 90 days or less.
    - Actual LOSs varied, see Figure 1.
  - N = 564 with 12-month follow up data (94.5% follow-up rate) subsetted to N = 486 with LOS in the experimental intervention of less than 120 days.

Figure 1. Histogram of Actual Number of Days in Treatment.



MEASURES

- All youths interviewed at treatment admission and 12 months later using the Global Appraisal of Individual Needs (GAIN; Dennis, 1998).
- 56 pre-treatment characteristics include: treatment site and condition, basic demographics, criminal history, drug use severity, and indices of health, mental health, and treatment motivation.
- **Outcome of interest:** Recovery at 12 months post-admission, where recovery is a dichotomous outcome defined as no drug use or drug or alcohol problems in the past 90 days and not in a controlled environment (like juvenile detention or residential treatment) during this period.

METHODS

- Goal is to estimate the recovery rate for subjects that completed 30 days of treatment had they remained in treatment t days, E($Y(t)$ | LOS = 30).

Propensity score for continuous treatments.

- Propensity score analysis (Rosenbaum and Rubin, 1983) can remove the effect of *observed* subject features that confound the treatment effect.
- Commonly used to estimate treatment effects for binary treatments, the methods have been recently extended for use with continuous treatments.
- Imai and van Dyk (2004) show that the effect of confounding factors, $\mathbf{x}$, that influence both LOS and outcomes with the following process:
  1. **Fit** a propensity score model to predict the (continuous) LOS from baseline features, $p(\mathbf{x})$ = model predicting LOS from $\mathbf{x}$
  2. **Predict** the propensity scores for each subject, $p_i = p(\mathbf{x}_i)$
  3. **Stratify** the observations into J strata according to the propensity scores
  4. **Regress** the outcome Y on observed LOS within each stratum $g_j(t) = E(Y \mid \text{LOS} = t, p(\mathbf{x})$ in stratum $j)$
  5. **Estimate** the causal effect of time-in-treatment for subjects with 90-day stays as a weighted average of the with strata models $E(Y(t) \mid \text{LOS} = 30) = w_1 \times g_1(\text{LOS}) + \ldots + w_J \times g_J(\text{LOS})$ w_j is the fraction of subjects with 30-day stays in stratum j.

Estimating the propensity score with gbm.

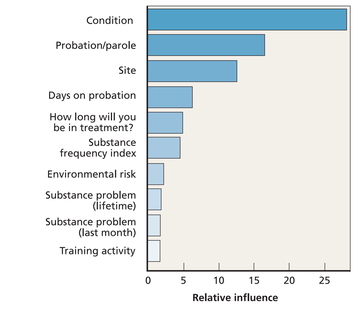
- We estimate the propensity scores with generalized boosted models (GBM), a multivariate nonparametric regression technique (Friedman, 2001; Ridgeway; 1999). GBM is a general, automated, data adaptive modeling algorithm that can estimate the nonlinear relationship between a variable of interest and a large number of covariates.
  - GBM initially predicts the mean LOS for all subjects.
  - GBM then iteratively searches for piecewise constant functions of the baseline features that are predictive of the residuals of the current model fit.
  - The final model is a sum of many piecewise constant functions that can approximate a large class of feature/outcome relationships.
- GBM's advantages include:
  - Captures non-linear effects and interaction terms
  - Selects variables automatically
  - Handles continuous, nominal, ordinal, and missing features
  - Is invariant to one-to-one transformations of the features
  - Can handle large numbers of correlated features.
- The free, open-source gbm package (Ridgeway, 2004) is available as an R-project library (www.r-project.org).

RESULTS

Step 1. Fit a propensity score model.

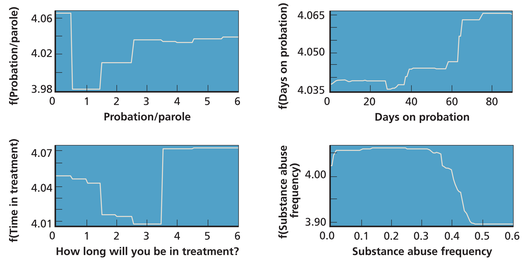
- GBM models log(LOS) well.
  - R^2 = 0.37.
- Figure 2 shows which baseline features most influence the length of stay.
  - Condition and site are among the most important.
  - Probation and parole measures are also predictive of LOS.
- Differences in outcomes by LOS may be a result of differences in these factors rather than exposure to the treatment.

Figure 2. Primary Predictors of Length of Stay.



- GBM captures the non-linear relationship between the baseline factors and length of stay.
  - Figure 3 shows the marginal relationships for four of the six primary predictors of LOS.
- In a propensity score analysis involving a binary treatment, McCaffrey et al., (2004) found that GBM outperformed standard linear models by four measures of model performance.

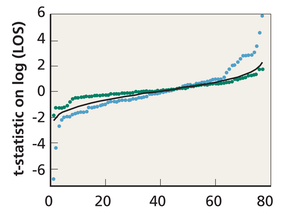
Figure 3. Marginal Relationship Between Baseline Features and Length of Stay.



Step 2 and 3. Predict the propensity score and stratify.

- We used J = 6 strata, enough to reduce bias while preserving a sufficient sample size within each stratum.
- The propensity strata remove the confounding of the baseline features and LOS.
- For each baseline feature, x_i:
  - fit a linear model predicting x_i from log(LOS). Store the t-statistic for the log(LOS) coefficient (blue)
  - fit a linear model predicting x_i from log(LOS) and propensity strata indicators. Store the t-statistic for the log(LOS) coefficient (green)
  - If we have properly adjusted, the latter should be roughly N(0,1).
- Figure 4 shows that after adjusting for the propensity strata, the relationships between LOS and the baseline predictors resemble random noise.

Figure 4. Adjusting for the Propensity Score Removes the Relationship Between Baseline Features and LOS.

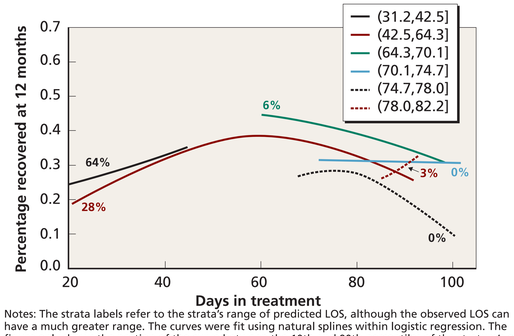


Notes: Blue points refer to unadjusted model and green dots refer to propensity score adjusted models. The black curve is what would be expected from a random sample from a standard normal distribution.

Step 4. Regress the outcome Y on observed LOS within each stratum.

- Within each stratum, we fit a smooth relation between observed LOS and recovery.
  - Logistic regression models using natural splines with 2 degrees of freedom.
- Figure 5 shows the relationships for each of the strata. The range of each curve indicates the middle 80% of the stratum's observed LOS. Unfortunately there is very little overlap in the domain of these curves. For example, only the (42.5,64.3] stratum has a substantial number of subjects that stay 30 days and 90 days. As a result, many of the subjects in the CYT sample offer little information on the effect of time in treatment for the 30-day stayers.

Figure 5. Within Strata Relationships Between Observed LOS and Recovery.

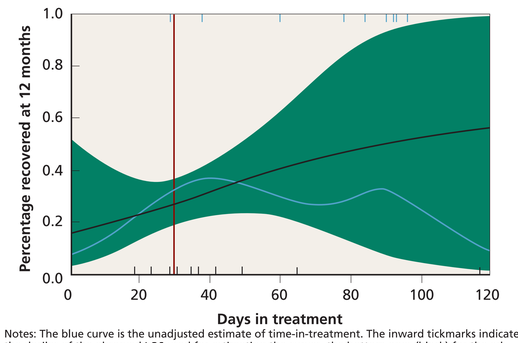


Notes: The strata labels refer to the strata's range of predicted LOS, although the observed LOS can have a much greater range. The curves were fit using natural splines within logistic regression. The figure only shows the portion of the curve between the 10th and 90th percentiles of the stratum's observed LOS. The percentages indicate the fraction of the study's subjects with stays between 26 and 34 days that are in each of the strata.

Step 5. Estimate the causal effect of time-in-treatment.

- To estimate the causal effect of time-in-treatment for subjects with 30-day stays, we compute pointwise weighted averages of the curves in Figure 5. Each curve receives a weight, shown as percentages in Figure 5, equal to the fraction of subjects with 30-day stays in that stratum. Since few subjects had exactly 30 day stays we considered all subjects with stays in 26-34 days.
- LOS tends to be associated with increased rates of recovery on average.
- Outside of 20 to 60 days, we have little information on what would have happened to subjects that actually stayed 30 days.

Figure 6. Causal estimate of the effect of time in treatment for subjects staying for 30 days (black).



Notes: The blue curve is the unadjusted estimate of time-in-treatment. The inward tickmarks indicate the deciles of the observed LOS used for estimating the curves, the bottom ones (black) for the adjusted curve and the top ones (blue) for the unadjusted curve. The green shading marks ±2 standard errors.

DISCUSSION

- The findings demonstrate that cases with LOS around 30 days are highly distinguishable at baseline from those with LOS greater than 60 days. This indicates that unadjusted, or naïve, comparisons of the outcomes of those with shorter and longer LOS are more likely to illustrate differences in the groups than the effects of treatment duration.
  - For example, subjects that stayed for 30 days were on probation for an average of 24 days in the three months prior to admission. 90-day stayers were on probation for 39 days. Therefore, differences in outcomes between the 30- and 90-day stayers may be due to differences in probation status rather than time-in-treatment.
- GBM is a tool that automates the process of estimating the propensity score and is particularly useful in the presence of a large number of baseline features.
- We showed that the propensity score adjustment can eliminate the confounding effects of observed pretreatment factors.
  - If important predictors of both length of stay and recovery were not included in the propensity score model, the approach described above may yield biased estimates.
- Failing to adjust for confounding factors can lead to bias in or understate the variance of the estimates of the effect of time in treatment.

ACKNOWLEDGMENTS

This report was developed with support from the National Institute on Drug Abuse 1R01 DA015697 and R01 DA017507-01 and the Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration (grant KD1-TI11433, and contract number 270-97-7011).

REFERENCES

Dennis, M. L. (1998). *Global Appraisal of Individual Needs (GAIN) manual: Administration, scoring and interpretation.* Bloomington, IL: Lighthouse Publications. Retrieved April 28, 2004, from: http://www.chestnut.org/LI/gain/index.html.

Friedman, J.H. (2001). "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics* 29(5):1189-1232.

Imai, K. and D.A. van Dyk. (2004, to appear). "Causal Inference With General Treatment Regimes: Generalizing the Propensity Score," *Journal of the American Statistical Association.*

McCaffrey, D.M., G. Ridgeway, A.R. Morral (2004, to appear). "Propensity Score Estimation with Boosted Regression for Evaluating Adolescent Substance Abuse Treatment," *Psychological Methods.*

Ridgeway, G. (1999). "The state of boosting," *Computing Science and Statistics* 31:172-181.

Ridgeway, G. (2004). *GBM 1.1-2 package manual.* Retrieved April 28, 2004, from http://cran.r-project.org/doc/packages/gbm.pdf.

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