Causal Inference Lab

Methods/Results with Propensity Score Methods

The Methods Section

- Sample size in treatment and comparison groups
  - Initially, and after matching
- Propensity score model estimation procedure
- Matching method used
- Covariates included in the matching process
- Covariates were treated in a "special" way
- Estimand of interest
- Diagnostics
- Description of how effects were estimated following the matching

The Methods Section

- *Optional*
- Discussion of why a particular matching method was selected
- Mention of other matching methods used as sensitivity analyses
- Description of sensitivity analysis done to assess sensitivity to unobserved confounding

Sample size in treatment and comparison groups

- Each of the **279** young men who met DSM-IV criteria for a lifetime substance use disorder (American Psychiatric Association, 2000) was matched one-to-one with a young man who did not meet lifetime criteria but had similar propensity for these disorders. Data from the resulting matched sample (*n* = **558**) were used in the analyses. Results for these 558 individuals may not generalize to the entire PIRC sample of **780** males, as they represent a non-random minority selected for their high propensity for substance use disorders.

Propensity score model estimation procedure

- *Multivariable logistic regression (MLR), MLR with critically chosen interaction terms (39, 40), and generalized boosted modeling (GBM), a nonparametric regression tree technique (41), were used to estimate the propensity score. Each of these techniques models cannabis problem use as a function of the measured covariates.*

- *Two propensity score packages written for the R environment were used: MatchIt (46) and Twang (47). The two parametric propensity score estimation techniques used MatchIt, while the nonparametric estimation technique used Twang, which utilized the GBM package in R.*

Matching method used

We utilized the full matching approach, as described by Rosenbaum (1991) and Hansen (2004). This approach allows us to retain all adolescents in our data analysis sample, and has been shown to be particularly effective at reducing bias due to observed confounding variables (Stuart and Green, 2008). Unlike k:1 matching, full matching is a more flexible approach.

Covariates included in the matching process

Clinical significance guided the initial choice of covariates: age, sex, year, surgery, fifth of income, residence (urban v rural), hospital's characteristics (teaching status and volume of procedures), comorbid disease, preoperative medical consultation (general internist, cardiologist), preoperative anesthesia consultation, epidural anesthesia, and invasive monitoring. The comorbidities included in the model were coronary artery disease, congestive heart failure, atrial fibrillation, cardiac vascular disease, mechanical heart valve, cerebrovascular disease, peripheral vascular disease, hypertension, diabetes, pulmonary disease, renal disease, previous venous thromboembolism, liver disease, peptic ulcer disease, rheumatological disease, hemiplegia or paraplegia, malignancy, and dementia.

Estimand of interest

This article presents the estimated average causal effects of the "treatment on the treated." In the causal inference methodology literature, the exposure variable is referred to as a "treatment," but in this article we retain the epidemiologic terminology of exposure. The "treatment on the treated" is an estimate of the average causal effect that would be seen if everyone in the exposed group had been exposed versus no one in the exposed group being exposed. The other commonly reported average causal effect is referred to simply as the "average treatment effect" and is described elsewhere (38). In this article, we present the "treatment on the treated" estimate.

Diagnostics

Next, we assessed the adequacy of matching by performing a series of diagnostic checks as described in Stuart and Green (2008). The assessment included an *examination of the balance of each covariate, its square, and every two-way interaction as determined by standardized bias*. Standardized biases of less than .25, that is less than a quarter of a standard deviation difference in means between heavy users and light/non-users, were considered good matches (Ho et al., 2007). To improve the matching, *we included various squared terms and interactions in the matching equation and re-estimated the propensity scores*. The final model included a squared term of adolescent delinquency and of shyness and interactions between female-headed household and maternal substance use and between poverty status and maternal substance use. Once adequate sets were formed, each individual was then assigned a weight based on the ratio of heavy users to light/non-users within a set. In this analysis, we created 142 matched sets based on the propensity score. Propensity scores ranged from .01 (very low) to .95 (high). Each matched set contained an average of five individuals. Although sets varied in terms of the number of marijuana users and comparison individuals, each included at least one of the 185 heavy marijuana users (mean 1.30, median 1.00) and at least one of the 517 light/non-users (mean 3.64, median 1.00). *The average difference in propensity scores within a set ranged from 0 to .05, with a mean of .002, demonstrating similarity of propensity scores within sets*

Description of how effects were estimated following the matching

After matching, we first used weighted logistic regression in StataIC 10, to estimate the association of heavy adolescent marijuana use with each criminal outcome. All regression models include the matching variables as controls in order to further adjust for small differences remaining in the matched samples after matching (Ho et al., 2007). We also included the interaction of shyness and family history of substance use since the standardized bias of this interaction was greater than .25 in the final model.

The Results Section

- State the effects clearly, accounting for the estimand (ATT vs. ATE);
- Provide the effect estimates for each outcome and matching method selected.
- The discussion and tables should make it clear that effects estimated in the matched sample (vs. the original data)
- Discuss limitation of method to only adjust for observed confounders
- Optional: Discuss the analysis of sensitivity to an unobserved confounder

State the effects clearly, accounting for the estimand

Within the matched cohort, preoperative echocardiography was associated with a small and statistically significant increase in postoperative mortality, both at 30 days (relative risk 1.14, 95% confidence interval 1.02 to 1.27; P=0.02; number needed to harm (NNH) 423) and at one year (1.07, 1.01 to 1.12; P=0.02; NNH 222) (table 5). It was also associated with an increase in mean hospital stay (0.31 (95% confidence interval 0.17 to 0.44) days; P<0.001), but not surgical site infection (relative risk 1.03, 0.98 to 1.06; P=0.18).

Provide the effect estimates for each outcome and matching method selected

Unadjusted survival curves are shown in Figure 2, and the survival curves adjusted with the use of inverse probability weighting are shown in Figure 3 At 1 year, there was no significant difference in adjusted mortality between the groups (6.2% in the CABG group as compared with 6.6% in the PCI group; risk ratio, 0.95; 95% confidence interval [CI], 0.90 to 1.00). The adjusted 4-year mortality was 16.4% in the CABG group and 20.8% in the PCI group (risk ratio, 0.79; 95% CI, 0.76 to 0.82).

Limitation of method: only adjusting for observed confounders

Furthermore, we cannot rule out the possibility that unmeasured pre-exposure characteristics may have jointly influenced both exposure status at Assessment 2 and perpetration at Assessment 3. Yet, the omission of such variables from our analyses would constitute a violation of the assumption of strongly ignorable treatment assignment only to the extent that their influences were independent of estimated propensity scores. This would be most likely in the case of an omitted variable that was uncorrelated with the 153 covariates used in developing and testing our propensity model…

Analysis of sensitivity to an unobserved confounder

Figure 4 shows the method that can be used to determine whether an unmeasured binary risk factor could explain a hazard ratio of this magnitude. The x axis represents the hypothetical prevalence of the unmeasured confounder in the PCI population, and the y axis represents the hypothetical hazard ratio for mortality associated with this confounder. The curved lines indicate the hypothetical prevalence (5%, 10%, 20%, 30%, or 40%) of the potential confounder in the CABG group. For example, if an unmeasured risk factor was present in 10% of the patients in the CABG group (green curved line) and in 20%, 35%, or 50% of the patients in the PCI group, then the hazard ratio that would be required for an unmeasured confounder to account for the observed decreased risk with CABG (i.e., to shift the upper 95% confidence interval from 0.80 to 1.00) would be 4.25, 2.09, and 1.65, respectively.