Propensity Score Analysis with Hierarchical Data

Fan Li   Alan Zaslavsky   Mary Beth Landrum

Department of Health Care Policy
Harvard Medical School

May 19, 2008

Introduction

- Population-based observational studies are increasingly important sources for estimating treatment effects.
- Proper adjustment for differences between treatment groups is crucial to valid comparison and causal inference.
- Regression has long been the standard method.
- Propensity score (Rosenbaum and Rubin, 1983) is a robust alternative to regression for adjusting for observed differences.

Hierarchically structured data

- Propensity score has been developed and applied in cross-sectional settings with unstructured data.
- Data in medical care and health policy research are often hierarchically structured.
- Subjects are grouped in natural clusters, e.g., geographical area, hospitals, health service provider, etc.
- Significant within- and between-cluster variations are often the case.

Hierarchically structured data

- Ignoring cluster structure often leads to invalid inference.
  - Standard errors could be underestimated.
  - Cluster level effect could be confounded with individual level effect.
- Hierarchical regression models provide a unified framework to study clustered data.
- Propensity score methods for hierarchical data have been less explored.

Propensity score

- Propensity score: $e(x) = P(z = 1|x)$.
- Balancing on propensity score also balances the covariates of different treatment groups: $z \perp x|e(x)$.
- Two steps procedure.
  - Step 1: estimate the propensity score, e.g., by logistic regression.
  - Step 2: estimate the treatment effect by incorporating (e.g., weighting, stratification) the estimated propensity score.
- We will introduce and compare several possible estimators of treatment effect using propensity score in context of hierarchical data.
- We will investigate the large sample behavior of each estimator.

Notation

- h cluster; k individual.
- m no. of clusters; n_h no. of subjects in cluster h.
- z_{hk} binary treatment assignment - individual level.
- x_{hk} individual level covariates; v_h cluster level covariates.
- e_{hk} propensity score.
- y_{hk} outcome.
- Estimand: "treatment effect"
  $\Delta = E_x[E(Y|X, Z = 1)] - E_x[E(Y|X, Z = 0)]$.
- Note: Δ does not necessarily have a causal interpretation, it is the difference of the average of a outcome between two populations controlling for covariates.

Step 1: Marginal model

- Marginal analysis ignores clustering.
- Marginal propensity score model

$$\log\left(\frac{e_{hk}}{1-e_{hk}}\right) = \beta^e x_{hk} + \kappa^e v_h,$$

  where $e_{hk} = P(z_{hk} = 1 \mid x_{hk}, v_h)$.
- If treatment assignment mechanism (TAM) follows above

$$(x_{hk}, v_h) \perp z_{hk} \mid e_{hk}.$$

Step 1: Pooled within-cluster model

- Pooled within-cluster model for propensity score ($e_{hk} = P(z_{hk} = 1 \mid x_{hk}, h)$)

$$\log(\frac{e_{hk}}{1 - e_{hk}}) = \delta_h^e + \beta^e x_{hk},$$

  where δ_h^e is a cluster-level main effect, $\delta_h^e \sim N(0, \infty)$.
- General (weaker) assumption of TAM than marginal model:

$$(x_{hk}, h) \perp z_{hk} \mid e_{hk}.$$

- Assuming $\delta_h^e \sim N(0, \sigma_\delta^2)$ gives a similar random effects model.

Step 1: Surrogate indicator model

- Define $d_h = \frac{\sum_h z_{hk}}{n_h}$ = cluster-specific proportion of being treated.
- Propensity score model

$$\log(\frac{e_{hk}}{1-e_{hk}}) = \lambda \log(\frac{d_h}{1-d_h}) + \beta^e x_{hk} + \kappa^e v_h.$$

- logit(d_h) is "surrogate" for the cluster indicator with the coefficient being around 1.
- Analytic model is same as the marginal analysis with proportion treated d_h as additional variable.
- Greatly reduce computation, but based on strong linear assumption.

Step 2: Estimate "treatment effect"

Estimate treatment effect using propensity score.

- Weighting - weight as function of propensity score.
- Stratification.
- Matching.
- Regression using propensity score as a predictor.

Marginal weighted estimator - ignore cluster structure

- w_{h1}(w_{h0}: sum of w_{hk} with $z = 1(z = 0)$ in cluster h.
- $w_1 = \sum_h w_{h1}, w_0 = \sum_h w_{h0}, w = w_1 + w_0$.
- Marginal weighted estimator - difference of weighted mean

$$\hat{\Delta}_{.,marg} = \sum_{h,k}^{z_{hk}=1} \frac{w_{hk}}{w_1} y_{hk} - \sum_{h,k}^{z_{hk}=0} \frac{w_{hk}}{w_0} y_{hk}.$$

- Large sample variance under homoscedasticity of y_{hk}

$$\begin{aligned} s^2_{.,marg} &= var(\hat{\Delta}_{.,marg}) \\ &= \sigma^2 \left(\sum_{h,k}^{z_{hk}=1} \frac{w^2_{hk}}{w^2_1} + \sum_{h,k}^{z_{hk}=0} \frac{w^2_{hk}}{w^2_0} \right). \end{aligned}$$

- In practice, σ^2 estimated from sample variance of y_{hk}.

Clustered weighted estimator

- Cluster-specific weighted estimator

$$\hat{\Delta}_h = \sum_{k \in h}^{z_{hk}=1} \frac{w_{hk}}{w_{h1}} y_{hk} - \sum_{k \in h}^{z_{hk}=0} \frac{w_{hk}}{w_{h0}} y_{hk}.$$

- The overall clustered estimator

$$\hat{\Delta}_{.,clu} = \sum_h \frac{w_h}{w} \hat{\Delta}_h.$$

Clustered weighted estimator

- Variance of $\hat{\Delta}_h$ under within-cluster homoscedasticity

$$s_h^2 = var(\hat{\Delta}_h) = \sigma_h^2 \Big(\sum_{k \in h}^{z_{hk}=1} \frac{w_{hk}^2}{w_{h1}^2} + \sum_{k \in h}^{z_{hk}=0} \frac{w_{hk}^2}{w_{h0}^2} \Big).$$

- Overall variance

$$s_{.,clu}^2 = var(\hat{\Delta}_{.,clu}) = \sum_h \frac{w_h^2}{w^2} s_h^2.$$

- Standard error can also be obtained using bootstrap.

Doubly-robust estimators (Scharfstein et al., 1999)

- Weighted mean can be viewed as a weighted regression without covariates.
- In step 2, replace the weighted mean by a weighted regression.
- Estimator is consistent if either or both of step 1 and 2 models are correctly specified.
- Numerous combination of regression models in two steps.

Choice of weight

- Horvitz-Thompson (inverse probability) weight

$$w_{hk} = \begin{cases} \frac{1}{e_{hk}}, \text{for } z_{hk} = 1 \\ \frac{1}{1-e_{hk}}, \text{for } z_{hk} = 0. \end{cases}$$

- Balance covariates distribution between two groups:

$$E\left[\frac{XZ}{e(X)}\right] = E\left[\frac{X(1-Z)}{1-e(X)}\right].$$

- H-T estimator compares the counterfactual scenario:
  all subjects placed in trt=0 vs. all subjects placed in trt=1.

$$E\left[\frac{YZ}{e(X)} - \frac{Y(1-Z)}{1-e(X)}\right] = E[(Y|Z=1) - (Y|Z=0)].$$

- H-T has large variance if e(X) approaches 0 or 1.

Choice of weight

- Population-overlap weight

$$w_{hk} = \begin{cases} 1 - e_{hk}, \text{for } z_{hk} = 1 \\ e_{hk}, \text{for } z_{hk} = 0. \end{cases}$$

- Each subject is weighted by the probability of being assigned to the other trt group.
- Balance covariates distribution between two groups:

$$E\{XZ[1 - e(X)]\} = E[X(1 - Z)e(X)].$$

- Small variance, different estimand.

$$\begin{aligned} & E\{YZ[1 - e(X)] - Y(1 - Z)e(X)\} \\ = \; & E\{[(Y|Z = 1) - (Y|Z = 0)]e(X)[1 - e(X)]\}. \end{aligned}$$

Bias of Estimators

- Focus on the simplest case with two-level hierarchical structure and no covariates.
- $n_{h1}(n_{h0})$: no. of subjects with $z = 1(z = 0)$ in cluster h.
- $n_1 = \sum_h n_{h1}, n_0 = \sum_h n_{h0}, n = n_1 + n_0$.
- Assume *outcome generating mechanism* is:

$$y_{hk} = \delta_h + \gamma_h z_{hk} + \alpha d_h + \epsilon_{hk}, \tag{1}$$

where $\delta_h \sim N(0, \sigma_\delta^2), \epsilon_{hk} \sim N(0, \sigma_\epsilon^2)$, and the true treatment effect: $\gamma_h \sim N(\gamma_0, \sigma_\gamma^2)$.

Bias of Marginal Estimator

- For marginal model in step 1, $\hat{e}_{hk} = \frac{n_1}{n}, \forall h, k$.
- The marginal estimator is

$$
\begin{aligned}
& \hat{\Delta}_{marg,marg} \\
= \quad & \sum_{h,k}^{z_{hk}=1} \frac{y_{hk}}{n_1} - \sum_{h,k}^{z_{hk}=0} \frac{y_{hk}}{n_0} \\
= \quad & \sum_h \frac{n_{h1}}{n_1}\gamma_h + \sum_h \left(\frac{n_{h1}}{n_1} - \frac{n_{h0}}{n_0}\right)\delta_h + \left(\sum_{h,k}^{z_{hk}=1} \frac{\epsilon_{hk}}{n_1} - \sum_{h,k}^{z_{hk}=0} \frac{\epsilon_{hk}}{n_0}\right) \\
& + \alpha \frac{\frac{n}{n_1 n_0} - \sum_h n_h d_h (1 - d_h)}{\frac{n}{n_1 n_0}}
\end{aligned}
$$

Bias of Marginal Estimator

- By WLLN of weighted sum of i.i.d. random variables (assuming $\sum_h^\infty \frac{n_{h1}^2}{n_1^2} < \infty$):

$$\sum_h \frac{n_{h1}}{n_1} \gamma_h \overset{n_h, m \to \infty}{\to} \gamma_0.$$

- Similarly the middle two parts go to 0 as $n_h, m \to \infty$.
- $\frac{n}{n_1 n_0} = var(n_1)$: variance of total no. of treated, if all clusters follow the same TAM, $z \sim Bernoulli(\frac{n_1}{n})$.
- $\sum_h n_h d_h (1 - d_h) = \sum_h var(n_{h1})$: sum of variance of no. of treated within each cluster, if each cluster follows a separate TAM: $z_{k \in h} \sim Bernoulli(\frac{n_{h1}}{n_h})$.

Bias of Marginal Estimator

- Exact form of bias

$$Bias(\hat{\Delta}_{marg,marg}) = \alpha \left(\frac{var(n_1) - \sum_h var(n_{h1})}{var(n_1)} \right). \tag{2}$$

- Controlled by two factors: (1) variance ratio - treatment assignment mechanism; (2) $|\alpha|$ - outcome generating mechanism.
- Both are ignored by the marginal estimator $\hat{\Delta}_{marg,marg}$.

Bias of Clustered Estimator

- For pooled within-cluster model in step 1, $\hat{e}_{hk} = \frac{n_{h1}}{n_h}, k \in h$.
- The clustered estimator with p.s. estimated from pooled within-cluster model $\hat{\Delta}_{pool,clu}$ is consistent

$$
\begin{aligned}
& \hat{\Delta}_{pool,clu} \\
= \quad & \frac{\sum_h (\sum_{k \in h}^{z_{hk}=1} \frac{y_{hk}}{n_{h1}})}{m} - \frac{\sum_h (\sum_{k \in h}^{z_{hk}=0} \frac{y_{hk}}{n_{h0}})}{m} \\
= \quad & \frac{\sum_h \gamma_h}{m} + \frac{\sum_h (\sum_{k \in h}^{z_{hk}=1} \frac{\epsilon_{hk}}{n_{h1}})}{m} - \frac{\sum_h (\sum_{k \in h}^{z_{hk}=0} \frac{\epsilon_{hk}}{n_{h0}})}{m} \\
\overset{n_h, m \to \infty}{\longrightarrow} \quad & \gamma_0
\end{aligned}
\tag{3}
$$

- This result is free of type of weights.

Bias of Clustered Estimator

- Clustered estimator with p.s. estimated from marginal model, $\hat{\Delta}_{marg,clu}$, exactly follows (3), thus consistent.
- Marginal estimator with p.s. estimated from pooled within-cluster model, $\hat{\Delta}_{pool,marg}$, also consistent.
- But different small sample behavior between H-T and population-overlap weights.

Extensions

- Without covariates, surrogate indicator model gives the estimated p.s. as pooled within-cluster model.
- Above results regarding pooled within-cluster model automatically hold for surrogate indicator model.
- Proofs are analogous for data with higher order of hierarchical level.

Double-robustness

- For the simplest case without covariates, we show “double-robustness” of the p.s. estimators:
  - When both of the true underlying treatment assignment mechanism and outcome generating mechanism are hierarchically structured:
  - Estimators using a balancing weight are consistent as if hierarchical structure is taken into account in at least one of the two steps in the p.s. procedure.
- A special case of Scharfstein et al. (1999), but free of form of weights.

Cases with covariates

- No closed-form solution to p.s. models, thus no closed-form of the bias of those estimators.
- Can be explored by (1) large-scale simulations; or (2) adopting a probit (instead of logistic) link for estimating p.s.
- Intuitively, “double-robustness” property still holds.
- Bias of $\hat{\Delta}_{marg,marg}$ is affected by:
  - α and $\frac{var(n_1)-\sum_h var(n_{h1})}{var(n_1)}$ in (2);
  - Size of true trt effect γ (negative correlated);
  - Ratio of between-cluster and within-cluster variance, $g = \frac{\sigma_\delta^2}{\sigma_\epsilon^2}$ (positively correlated).

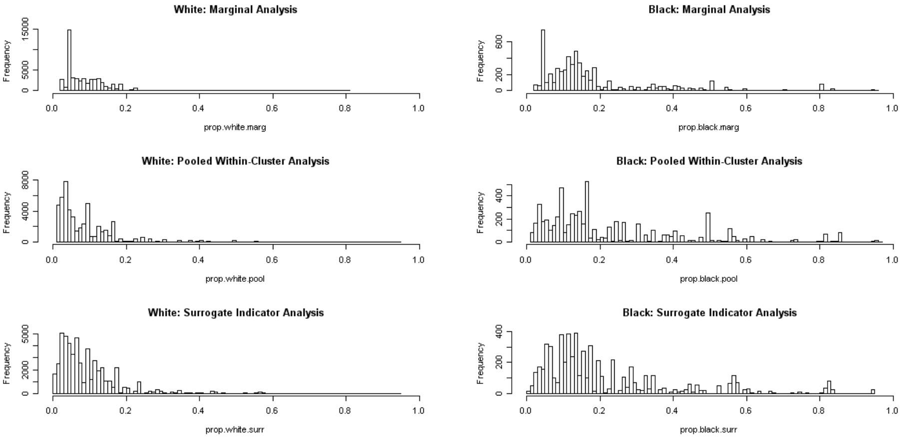
Racial disparity data

- Disparity: racial differences in care attributed to operations of health care system.
- Breast cancer screening data are collected from health insurance plans.
- Focus on the plans with at least 25 whites and 25 blacks: 64 plans with a total sample size of 75012.
- Subsample 3000 subjects from large (>3000) clusters to restrict impact of extremely large clusters, resulting sample size 56480.

Racial disparity data

- Cluster level covariates v_h: geographical code, non/for-profit status, practice model.
- Individual level covariates x_{hk}: age category, eligibility for medicaid, poor neighborhood.
- "Treatment" variable z_{hk}: black race (1=black, 0=white).
- Not strictly causal. Compare groups with balanced covariates.
- Outcome y_{hk}: receive screening for breast cancer or not.
- Research aim: investigate racial disparity in breast cancer screening.

Estimated propensity score



Estimated propensity score

- Different propensity score models give quite different estimates.
- Each method leads to good overall covariates balance between groups in this data.
- Marginal analysis does not lead to balance in covariates in each cluster, surrogate indicator analysis does better, pooled- within the best.

Analysis results: racial disparity estimated from Horvitz-Thompson weight

	weighted		doubly-robust		regression
	pooled	clustered	marginal	pooled-within	
marginal	-0.050	-0.020	-0.042	-0.021	-0.044
	(0.008)	(0.008)	(0.004)	(0.004)	(0.007)
pooled-within	-0.024	-0.021	-0.018	-0.022	-0.032
	(0.009)	(0.008)	(0.004)	(0.004)	(0.007)
surrogate indicator	-0.017	-0.015	-0.012	-0.015	-0.014
	(0.009)	(0.008)	(0.004)	(0.004)	(0.007)

Analysis results: racial disparity estimated from population-overlap weight

	weighted		doubly-robust		regression
	pooled	clustered	marginal	pooled-within	
marginal	-0.043	-0.030	-0.043	-0.032	-0.044
	(0.007)	(0.008)	(0.004)	(0.004)	(0.007)
pooled-within	-0.030	-0.031	-0.031	-0.031	-0.032
	(0.007)	(0.008)	(0.004)	(0.004)	(0.007)
surrogate indicator	-0.035	-0.030	-0.031	-0.030	-0.014
	(0.007)	(0.008)	(0.004)	(0.004)	(0.007)

Diagnostics

- Check the balance of weighted covariates between treatment groups.
  Each method leads to good balance in this data.
- Quantiles table.

Remarks on results

- Ignoring cluster structure in both steps gives results greatly defer from others.
- Results from surrogate indicator analysis are different from others, suggesting Portion treated is correlated with covariates.
- Taking into account cluster structure in at least one of the two steps leads to similar results - “doubly-robustness”.
- Doubly-robust estimates have smaller s.e., extra variation is explained by covariates in step 2.
- Incorporating cluster structure in step 2 is preferable to step 1.
- Between-cluster variation is large in breast cancer data.
- Standard errors obtained from bootstrap are much larger than those from analytic formula.

Summary

- We introduce and compare several possible propensity score analyses for hierarchical data.
- We show “double-robustness” property of propensity score weighted estimators: cluster structure must be taken into account in at least one of the two steps.
- We obtain the analytic form of bias of the marginal estimator.
- Case by case. In practice, total number of clusters, size of each cluster, within- and between- cluster variations can greatly affect the conclusion.