Matching Estimators for Causal Effects of Multiple Treatments

Anthony D. Scotina

(joint work with Roee Gutman)

Anthony_Scotina@brown.edu

July 30, 2017

BROWN
School of Public Health

Setting

- **Objective**: Compare the effects of treatment regimens on the occurrence of cardiovascular events in patients with type 2 diabetes.
- Metformin: common first-line treatment for type 2 diabetes
  - low-cost, high efficacy, associated with weight loss, low risk of hypoglycemia
- Of interest: second-line treatments when used in combination with Metformin

Estimand of Interest

- We are interested in simultaneously estimating treatment effects of all second-line treatment regimens.
  - Three drug classes: Sulfonylureas, DPP-4 inhibitors, Thiazolidinediones
- Why simultaneous estimation?
  - Patient cohort in "clinical equipoise"
  - Avoid the need to perform many separate pairwise studies
- Need to carefully consider treatments of interest

Background

- Matching methods have been proposed, for both binary and multiple treatment settings, to balance units on pre-intervention characteristics to replicate a randomized experiment.
- We refer to point estimates of treatment effects derived from matching as *matching estimators*.
- The theory and use of matching estimators is well-developed for the binary treatment setting (Abadie & Imbens, 2006, 2008, 2011).
- But...many studies actually contain multiple levels of treatment!
  - Example: {Tylenol, Advil, Generic}, {Smoke, Alcohol, Neither}

Framework

- For Z possible treatments:
  - W_i is the treatment group identification for unit i, $i = 1, \ldots, n$.
  - n_w is the size of treatment group w
  - $T_{iw} = 1$ if $W_i = w$, and 0 otherwise
  - Covariates $\mathbf{X}_i = (X_{i1}, \ldots, X_{iP})$

- Assuming SUTVA, $Y_i^{obs} = T_{i1}\, Y_i(1) + \cdots + T_{iZ}\, Y_i(Z)$.

- Estimand: Average treatment effect on the treated (ATT),

$$\tau_{jk}^{t} \equiv E[Y(j) - Y(k) \mid W = t],\ (j,k) \in \mathcal{W}^2,\ j \neq k$$

Vector Matching (VM)

- Lopez and Gutman (2017) – Match within strata of the generalized propensity score vector (GPS)
  - $R(\mathbf{X}_i) = \{P(W_i = 1 \mid \mathbf{X}_i), \ldots, P(W_i = Z \mid \mathbf{X}_i)\}$
- Choice of estimand is important: generalizes to plausible candidates for each of the treatments under study.
- Matching within strata ensures that units matched to one another are nearly perfect matches within one GPS and roughly similar on other treatment assignment probabilities.
- VM has been shown to produce matched sets with low covariate bias for $Z = 3$ treatments, and yield a high proportion of matched units.

Matched Data Example

Table: Matched Dataset ($Z = 3$)

ID	Treatment (W)	$Y_i(1)$	$Y_i(2)$	$Y_i(3)$	$\widehat{r}(1\|X_i)$	$\widehat{r}(2\|X_i)$	$\widehat{r}(3\|X_i)$	$\widehat{Y}_i(1)$	$\widehat{Y}_i(2)$	$\widehat{Y}_i(3)$
1	1	136	?	?	0.52	0.37	0.11	136	112	122
2	1	107	?	?	0.16	0.68	0.16	107	111	113
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1028	1	173	?	?	0.32	0.51	0.16	173	131	119
1029	2	?	112	?	0.52	0.38	0.10	–	–	–
1030	2	?	111	?	0.16	0.69	0.15	–	–	–
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1614	2	?	131	?	0.32	0.50	0.18	–	–	–
1615	3	?	?	122	0.52	0.29	0.19	–	–	–
1616	3	?	?	113	0.16	0.69	0.15	–	–	–
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1963	3	?	?	110	0.26	0.59	0.15	–	–	–

Matched Data Example

Table: Matched Dataset ($Z = 3$)

ID	Treatment (W)	$Y_i(1)$	$Y_i(2)$	$Y_i(3)$	$\widehat{r}(1\|X_i)$	$\widehat{r}(2\|X_i)$	$\widehat{r}(3\|X_i)$	$\widehat{Y}_i(1)$	$\widehat{Y}_i(2)$	$\widehat{Y}_i(3)$
1	1	136	?	?	0.52	0.37	0.11	136	112	122
2	1	107	?	?	0.16	0.68	0.16	107	111	113
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1028	1	173	?	?	0.32	0.51	0.16	173	131	119
1029	2	?	112	?	0.52	0.38	0.10	–	–	–
1030	2	?	111	?	0.16	0.69	0.15	–	–	–
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1614	2	?	131	?	0.32	0.50	0.18	–	–	–
1615	3	?	?	122	0.52	0.29	0.19	–	–	–
1616	3	?	?	113	0.16	0.69	0.15	–	–	–
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1963	3	?	?	110	0.26	0.59	0.15	–	–	–

Matched Data Example

Table: Matched Dataset ($Z = 3$)

ID	Treatment (W)	$Y_i(1)$	$Y_i(2)$	$Y_i(3)$	$\widehat{r}(1\|X_i)$	$\widehat{r}(2\|X_i)$	$\widehat{r}(3\|X_i)$	$\widehat{Y}_i(1)$	$\widehat{Y}_i(2)$	$\widehat{Y}_i(3)$
1	1	136	?	?	0.52	0.37	0.11	136	112	122
2	1	107	?	?	0.16	0.68	0.16	107	111	113
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1028	1	173	?	?	0.32	0.51	0.16	173	131	119
1029	2	?	112	?	0.52	0.38	0.10	–	–	–
1030	2	?	111	?	0.16	0.69	0.15	–	–	–
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1614	2	?	131	?	0.32	0.50	0.18	–	–	–
1615	3	?	?	122	0.52	0.29	0.19	–	–	–
1616	3	?	?	113	0.16	0.69	0.15	–	–	–
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1963	3	?	?	110	0.26	0.59	0.15	–	–	–

Matched Data Example

Table: Matched Dataset ($Z = 3$)

ID	Treatment (W)	$Y_i(1)$	$Y_i(2)$	$Y_i(3)$	$\widehat{r}(1\|X_i)$	$\widehat{r}(2\|X_i)$	$\widehat{r}(3\|X_i)$	$\widehat{Y}_i(1)$	$\widehat{Y}_i(2)$	$\widehat{Y}_i(3)$
1	1	136	?	?	0.52	0.37	0.11	136	112	122
2	1	107	?	?	0.16	0.68	0.16	107	111	113
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1028	1	173	?	?	0.32	0.51	0.16	173	131	119
1029	2	?	112	?	0.52	0.38	0.10	–	–	–
1030	2	?	111	?	0.16	0.69	0.15	–	–	–
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1614	2	?	131	?	0.32	0.50	0.18	–	–	–
1615	3	?	?	122	0.52	0.29	0.19	–	–	–
1616	3	?	?	113	0.16	0.69	0.15	–	–	–
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1963	3	?	?	110	0.26	0.59	0.15	–	–	–

Matched Data Example

Table: Matched Dataset ($Z = 3$)

ID	Treatment (W)	$Y_i(1)$	$Y_i(2)$	$Y_i(3)$	$\widehat{r}(1\|X_i)$	$\widehat{r}(2\|X_i)$	$\widehat{r}(3\|X_i)$	$\widehat{Y}_i(1)$	$\widehat{Y}_i(2)$	$\widehat{Y}_i(3)$
1	1	136	?	?	0.52	0.37	0.11	136	112	122
2	1	107	?	?	0.16	0.68	0.16	107	111	113
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1028	1	173	?	?	0.32	0.51	0.16	173	131	119
1029	2	?	112	?	0.52	0.38	0.10	–	–	–
1030	2	?	111	?	0.16	0.69	0.15	–	–	–
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1614	2	?	131	?	0.32	0.50	0.18	–	–	–
1615	3	?	?	122	0.52	0.29	0.19	–	–	–
1616	3	?	?	113	0.16	0.69	0.15	–	–	–
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1963	3	?	?	110	0.26	0.59	0.15	–	–	–

Matched Data Example

Table: Matched Dataset ($Z = 3$)

ID	Treatment (W)	$Y_i(1)$	$Y_i(2)$	$Y_i(3)$	$\widehat{r}(1\|X_i)$	$\widehat{r}(2\|X_i)$	$\widehat{r}(3\|X_i)$	$\widehat{Y}_i(1)$	$\widehat{Y}_i(2)$	$\widehat{Y}_i(3)$
1	1	136	?	?	0.52	0.37	0.11	136	112	122
2	1	107	?	?	0.16	0.68	0.16	107	111	113
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1028	1	173	?	?	0.32	0.51	0.16	173	131	119
1029	2	?	112	?	0.52	0.38	0.10	–	–	–
1030	2	?	111	?	0.16	0.69	0.15	–	–	–
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1614	2	?	131	?	0.32	0.50	0.18	–	–	–
1615	3	?	?	122	0.52	0.29	0.19	–	–	–
1616	3	?	?	113	0.16	0.69	0.15	–	–	–
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1963	3	?	?	110	0.26	0.59	0.15	–	–	–

Matched Data Example

Table: Matched Dataset ($Z = 3$)

ID	Treatment (W)	$Y_i(1)$	$Y_i(2)$	$Y_i(3)$	$\widehat{r}(1\|X_i)$	$\widehat{r}(2\|X_i)$	$\widehat{r}(3\|X_i)$	$\widehat{Y}_i(1)$	$\widehat{Y}_i(2)$	$\widehat{Y}_i(3)$
1	1	136	?	?	0.52	0.37	0.11	136	112	122
2	1	107	?	?	0.16	0.68	0.16	107	111	113
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1028	1	173	?	?	0.32	0.51	0.16	173	131	119
1029	2	?	112	?	0.52	0.38	0.10	–	–	–
1030	2	?	111	?	0.16	0.69	0.15	–	–	–
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1614	2	?	131	?	0.32	0.50	0.18	–	–	–
1615	3	?	?	122	0.52	0.29	0.19	–	–	–
1616	3	?	?	113	0.16	0.69	0.15	–	–	–
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1963	3	?	?	110	0.26	0.59	0.15	–	–	–

Matched Data Example

Table: Matched Dataset ($Z = 3$)

ID	Treatment (W)	$Y_i(1)$	$Y_i(2)$	$Y_i(3)$	$\widehat{r}(1\|X_i)$	$\widehat{r}(2\|X_i)$	$\widehat{r}(3\|X_i)$	$\widehat{Y}_i(1)$	$\widehat{Y}_i(2)$	$\widehat{Y}_i(3)$
1	1	136	?	?	0.52	0.37	0.11	136	112	122
2	1	107	?	?	0.16	0.68	0.16	107	111	113
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1028	1	173	?	?	0.32	0.51	0.16	173	131	119
1029	2	?	112	?	0.52	0.38	0.10	–	–	–
1030	2	?	111	?	0.16	0.69	0.15	–	–	–
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1614	2	?	131	?	0.32	0.50	0.18	–	–	–
1615	3	?	?	122	0.52	0.29	0.19	–	–	–
1616	3	?	?	113	0.16	0.69	0.15	–	–	–
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1963	3	?	?	110	0.26	0.59	0.15	–	–	–

Matched Data Example

Table: Matched Dataset ($Z = 3$)

ID	Treatment (W)	$Y_i(1)$	$Y_i(2)$	$Y_i(3)$	$\widehat{r}(1\|X_i)$	$\widehat{r}(2\|X_i)$	$\widehat{r}(3\|X_i)$	$\widehat{Y}_i(1)$	$\widehat{Y}_i(2)$	$\widehat{Y}_i(3)$
1	1	136	?	?	0.52	0.37	0.11	136	112	122
2	1	107	?	?	0.16	0.68	0.16	107	111	113
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1028	1	173	?	?	0.32	0.51	0.16	173	131	119
1029	2	?	112	?	0.52	0.38	0.10	–	–	–
1030	2	?	111	?	0.16	0.69	0.15	–	–	–
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1614	2	?	131	?	0.32	0.50	0.18	–	–	–
1615	3	?	?	122	0.52	0.29	0.19	–	–	–
1616	3	?	?	113	0.16	0.69	0.15	–	–	–
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
1963	3	?	?	110	0.26	0.59	0.15	–	–	–

The Matching Estimator

- The matching estimator imputes the missing potential outcomes as the average of all observed outcomes of matches from treatment group w for unit i,

$$\widehat{Y}_i(w) = \begin{cases} Y_i(w), & \text{if } T_{iw} = 1 \\ \frac{1}{m} \sum_{j \in \mathcal{M}_i^w} Y_j(w), & \text{if } T_{iw} = 0, \end{cases}$$

  where $\mathcal{M}_i^w$ is the set of indices for the closest m matches from treatment group w for unit i.

- The point estimate for τ_{jk}^t is:

$$\widehat{\tau}_{jk}^t = \frac{1}{n_t} \sum_{W_i = t} \left(\widehat{Y}_i(j) - \widehat{Y}_i(k) \right) = \frac{1}{n_t} \sum_{i=1}^{n} (T_{ij} - T_{ik}) \left(T_{it} + \frac{\psi_{it}}{m} \right) Y_i^{obs},$$

  where ψ_{it} is the number of times that unit i serves as a match to other units in treatment group t.

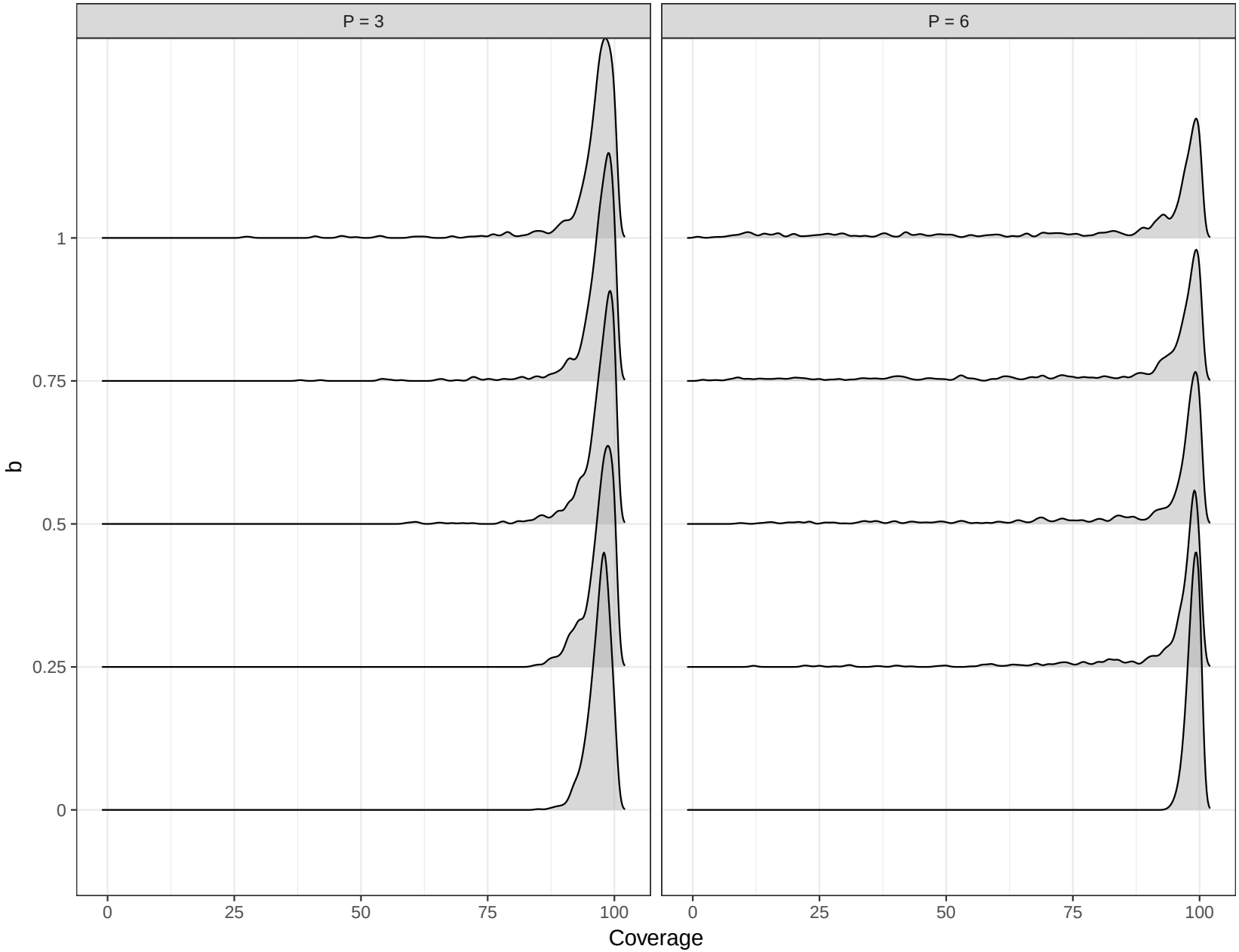
Simulation Results

Table: Median, 25% percentile, and 75% percentile of the 95% coverage for τ^1

	% Configurations over 90% coverage	Median	25%	75%
B-N	86	0.98	0.93	0.99

- Simulation factors included:
  - Covariate bias (b), number of covariates (P), treatment group sizes, covariate distribution, etc.
- Also found:
  - Region coverage decreases as b and P increase.

Simulation Results



Discussion

- Simulations demonstrate that estimation using matching yields close to nominal coverage.
- Results were obtained for multiple, continuous covariates with a continuous outcome, by matching for nominal treatment.
- Future work:
  - Identify number of matches per unit m with optimal operating characteristics.
  - Variable ratio matching for multiple treatments, with different m for each unit

Contact

- Email: anthony_scotina@brown.edu
- Website: scotinastats.github.io
- Twitter: @ScotinaStats

Contact

- Email: anthony_scotina@brown.edu
- Website: scotinastats.github.io
- Twitter: @ScotinaStats

Thank you!