Initial Balance Assessment before Matching

Jake Bowers

November 19, 2019

Contents


```
load(here::here("Analysis", "match_data_prep.rda"))
```

0.0.1 Look at baseline imbalance before matching

The following analysis shows that the neighborhoods with and without marijuana selling pharmacies are quite similar on the covariates listed below. Below we show the standardized differences (differences in means in standard deviation units) and p-values for a test of the null of no difference in means between the registered pharmacies and the non-registered pharmacies.

First, drop the observations for the placebo pharmacies

```
dat17i <- wdat17i %>% filter(!is.na(soldvsnot17))
table(dat17i$soldvsnot17, exclude = c())
```

```
  0   1
420 160
```

Looking at baseline (im)balance on individual level outcomes and covariates

```
baselineFmla <- reformulate(covs3, response = "soldvsnot17")
baselineFmlaCluster <- update(baselineFmla, . ~ . + cluster(Q56))

tmp <- model.matrix(update(baselineFmlaCluster, . ~ . - 1), data = dat17i)
uniqX <- sort(apply(tmp, 2, function(x) {
  length(unique(x))
}))
uniqX[1:10]
```

```
    dis6_i    mvd_int      sex_i    vic12_i     dis4_i     dis5_i sec_mea1_p sec_mea2_p sec_mea3_p sec_mea4_p
         2          2          2          2          3          3          3          3          3          3
```

```
anyNA <- apply(tmp, 2, function(x) {
  any(is.na(x) | is.nan(x))
})
table(anyNA)
```

```
anyNA
FALSE
   82
```

```
xb0i <- balanceTest(baselineFmlaCluster, data = dat17i, report = "all", p.adjust.method = "none")
```

```
Warning in sqrt((var.1 + var.0)/2): NaNs produced
```

```
xb0i$overall[, ]
```

```
chisquare        df   p.value
  7.12094   3.00000   0.06814
```

```
xb0ionebyone <- data.frame(xb0i$results[, , ])
xb0ionebyone$varnm <- row.names(xb0ionebyone)
xb0ionebyone <- xb0ionebyone %>% arrange(desc(abs(std.diff)))
xb0ionebyone
```

	Control	Treatment	std.diff	adj.diff	pooled.sd	z	p	varnm
1	7.52381	1.6250	-1.045312	-5.898810	5.64311	-2.82995	0.004656	vrobb_2016
2	1.58095	1.1000	-0.805782	-0.480952	0.59688	-2.71891	0.006550	dis1_i
3	1.36368	1.0797	-0.747834	-0.284013	0.37978	-2.15897	0.030853	sec_mea3_p
4	39.82381	42.2550	0.648282	2.431190	3.75020	2.14955	0.031591	age_av
5	1.56905	1.1813	-0.643943	-0.387798	0.60222	-2.44627	0.014434	dis2_i
6	1032.11905	755.3750	-0.622365	-276.744048	444.66544	-2.08952	0.036661	pop
7	58.04762	38.8750	-0.533144	-19.172619	35.96140	-1.61854	0.105546	robb_2016
8	1.16667	1.3917	0.525500	0.225010	0.42818	1.82523	0.067966	sec_mea1_p
9	0.47619	0.2500	-0.482949	-0.226190	0.46835	-1.54971	0.121210	mvd_int
10	31.70000	35.4446	0.473959	3.744597	7.90067	1.52822	0.126459	pn_per
11	9.95881	11.6925	0.456913	1.733690	3.79436	1.75977	0.078447	educ_av
12	49.58143	45.2106	-0.455294	-4.370825	9.60000	-1.49853	0.133996	fa_per
13	2.28333	1.8313	-0.451951	-0.452083	1.00029	-1.62745	0.103641	dis3_i
14	0.13095	0.1717	0.400424	0.040764	0.10180	1.51365	0.130115	pt_per
15	2.11490	1.8563	-0.314181	-0.258646	0.82324	-2.22913	0.025805	n_sec_i
16	0.95714	0.7627	-0.309157	-0.194431	0.62891	-1.01572	0.309765	peri_per
17	1.26429	1.1008	-0.303781	-0.163438	0.53801	-1.03247	0.301851	ap_per
18	1.78571	1.8938	0.297174	0.108036	0.36354	2.60483	0.009192	vic12_i
19	12.82143	14.0450	0.295731	1.223593	4.13752	1.00932	0.312820	pc_per
20	1.20075	1.4249	0.279611	0.224161	0.80169	2.31838	0.020429	neigh3_i
21	4.35458	3.7606	-0.276394	-0.594003	2.14912	-2.21236	0.026941	pstigma3_i
22	55.26190	51.4000	-0.266478	-3.861905	14.49239	-1.01608	0.309594	h_owners
23	5.12462	4.6587	-0.233545	-0.465931	1.99504	-1.91824	0.055080	pstigma4_i
24	78.64095	81.0031	0.229855	2.362173	10.27681	0.77744	0.436897	ubn_no
25	2.70691	2.5125	-0.228366	-0.194407	0.85130	-2.08994	0.036623	c_sec_i
26	14.67333	13.2569	-0.223520	-1.416458	6.33707	-0.75155	0.452321	ubn_one
27	0.06563	-0.2126	-0.202424	-0.278272	1.37470	-1.55286	0.120457	ps1718
28	3.54286	3.2601	-0.191851	-0.282794	1.47403	-0.64098	0.521538	pi_per
29	1.92857	1.8717	-0.191760	-0.056866	0.29655	-0.68416	0.493876	sec_mea4_p
30	3.53952	3.2918	-0.185869	-0.247739	1.33287	-1.62323	0.104540	law2_i
31	3.83257	3.4114	-0.180392	-0.421208	2.33495	-1.39807	0.162091	op6_m_i
32	3.88065	3.7750	-0.167424	-0.105646	0.63101	-1.79655	0.072407	neigh6_i
33	4.25415	3.9237	-0.161451	-0.330460	2.04682	-1.33594	0.181568	pstigma2_i
34	5.81606	5.5469	-0.150701	-0.269206	1.78636	-1.48983	0.136269	op5_m_i
35	1.00952	1.0313	0.145338	0.021726	0.14949	1.41345	0.157523	dis5_i
36	2.70338	2.9932	0.144708	0.289791	2.00259	1.44497	0.148467	op3_m_i
37	3.56941	3.6875	0.143486	0.118085	0.82297	1.36605	0.171924	neigh2_i
38	2.79144	2.5169	-0.137684	-0.274539	1.99397	-1.20383	0.228656	stigma4_i
39	2.25023	2.1552	-0.137440	-0.095013	0.69130	-1.23025	0.218603	ph_impact_i
40	1.62143	1.5563	-0.132466	-0.065179	0.49204	-1.43666	0.150815	sex_i
41	1.89046	1.8080	-0.128006	-0.082435	0.64399	-1.25762	0.208529	vic12_n_i
42	2.31566	2.2277	-0.127996	-0.088005	0.68756	-1.14764	0.251116	ps_impact_i
43	1.92024	1.5150	-0.127063	-0.405238	3.18928	-0.40424	0.686039	ubn_more
44	1.00714	1.0000	-0.119809	-0.007143	0.05962	-1.08831	0.276459	dis6_i
45	2.05580	1.9678	-0.118093	-0.088019	0.74534	-1.13933	0.254564	if_impact_i
46	4.36968	4.1315	-0.114194	-0.238193	2.08587	-0.90339	0.366317	pstigma1_i
47	27.00952	25.6313	-0.111637	-1.378274	12.34599	-0.39343	0.694000	rent_per
48	0.01844	-0.1022	-0.108686	-0.120605	1.10967	-0.92168	0.356693	st1718
49	2.47074	2.2915	-0.103229	-0.179271	1.73663	-0.99195	0.321220	stigma6_i
50	1.73705	1.8646	0.097721	0.127590	1.30565	0.88029	0.378703	neigh4_i
51	1.01190	1.0208	0.093858	0.008929	0.09513	0.82143	0.411404	social_dis
52	2.97100	3.1625	0.088447	0.191500	2.16514	0.78243	0.433962	stigma1_i
53	2.70386	2.8783	0.087303	0.174401	1.99767	0.77680	0.437279	op4_m_i
54	1.92913	1.8379	-0.085803	-0.091212	1.06304	-0.70751	0.479249	rp_m1_i
55	4.26246	4.1143	-0.079961	-0.148146	1.85272	-0.69794	0.485217	pstigma6_i
56	3.40261	3.2949	-0.077806	-0.107718	1.38444	-0.71700	0.473374	law1_i
57	2.79703	2.8560	0.077672	0.059000	0.75960	0.76371	0.445041	crime_t_i
58	3.63744	3.5856	-0.075417	-0.051832	0.68727	-0.63872	0.523007	rp_m3_i
59	1.79682	1.8552	0.068236	0.058382	0.85558	0.57126	0.567826	neigh7_i
60	2178684.69048	2025986.3125	-0.060444	-152698.377976	2526268.73103	-0.19064	0.848810	cat_value
61	50.74762	49.6859	-0.057045	-1.061695	18.61155	-0.46291	0.643432	age_i
62	7912.09524	8432.1500	0.055010	520.054762	9453.82698	0.21082	0.833026	dens
63	1.01905	1.0313	0.054233	0.012202	0.22500	0.45221	0.651116	dis4_i
64	1.99475	2.0336	0.051319	0.038885	0.75770	0.45948	0.645887	dt_impact_i
65	5.19579	5.0836	-0.050685	-0.112190	2.21346	-0.46201	0.644071	op2_m_i
66	1.48095	1.4375	-0.050392	-0.043452	0.86228	-0.42999	0.667203	prev_lt_i
67	2.00283	2.0503	0.050201	0.047433	0.94486	0.53066	0.595654	ffuse_i

```
68       1.82054       1.7666 -0.036914      -0.053895       1.46000 -0.34820 0.727690        stigma3_i
69       1.77610       1.7421 -0.035938      -0.033972       0.94529 -0.24024 0.810145         neigh5_i
70       3.66256       3.5981 -0.035469      -0.064462       1.81745 -0.43012 0.667109       pstigma5_i
71       1.88095       1.8705 -0.031962      -0.010464       0.32738 -0.10849 0.913603       sec_mea2_p
72       2.66997       2.6386 -0.030738      -0.031322       1.01899 -0.24984 0.802714          rp_m2_i
73       0.67078       0.6765  0.029001       0.005675       0.19569  0.23144 0.816973 activities_index
74       5.24126       5.1924 -0.024590      -0.048908       1.98890 -0.21399 0.830555       pstigma8_i
75       5.10988       5.0798 -0.014753      -0.030046       2.03660 -0.14381 0.885649          op1_m_i
76       5.77143       5.7438 -0.012579      -0.027679       2.20040 -0.08485 0.932380           educ_i
77       2.56974       2.5921  0.011092       0.022368       2.01664  0.09913 0.921034        stigma2_i
78       2.34828       2.3330 -0.010792      -0.015293       1.41702 -0.07278 0.941985         neigh8_i
79       5.38122       5.3569 -0.009962      -0.024334       2.44258 -0.09409 0.925041       ideol_si_i
80       3.67238       3.6900  0.004787       0.017619       3.68052  0.01616 0.987109          ubn_two
81       4.57234       4.5671 -0.002798      -0.005228       1.86867 -0.02935 0.976583       pstigma7_i
```

```
## head(xb0ionebyone,n=20) ## Worst balanced
## Number of small p-values
numsmallp1 <- sum(xb0ionebyone[, "p"] <= .05)

## xb0itest <- balanceTest(baselineFmla, data = wdat17, report = "all", p.adjust.method = "none")
## xb0itest$overall[,]
```

```
summary(xb0ionebyone$std.diff)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.0453 -0.1859 -0.0754 -0.0842  0.0513  0.6483
```

```
adjps <- p.adjust(xb0ionebyone$p, method = "holm")
```

```
xb0ionebyone$thevar <- 1:nrow(xb0ionebyone)

pdf(file = here::here("media", "initial_balance_plot.pdf"))
par(oma = rep(0, 4) + .01, mar = c(3, 8, 0, 0), mgp = c(1, .5, 0))
with(xb0ionebyone, {
  plot(abs(std.diff), thevar,
    pch = 21,
    xlab = "Absolute Std. Diff of Means", ylab = "",
    bg = c("white", "black")[as.numeric(xb0ionebyone$p <= .05) + 1],
    axes = FALSE
  )
  axis(1)
  axis(2, at = thevar, labels = varnm, las = 2, tick = FALSE)
  segments(rep(0, nrow(xb0ionebyone)), thevar, abs(std.diff), thevar, lwd = .5, col = "gray")
})
dev.off()
```

```
pdf
  2
```

Relationship between pharmacies and baseline perception of risk:

```
table(dat17i$treat, dat17i$n_sec_i, exclude = c())
```

```
     1   2 2.12818532818533   3   4
  0 107 183                2 101  27
  1  58  71                0  27   4
```

```
boxplot(n_sec_i ~ treat, data = dat17i)
stripchart(n_sec_i ~ treat, data = dat17i, vertical = TRUE, add = TRUE)
```

twidths

Hansen and Bowers (2008) suggested that an observational study could be judged, in part, by comparing it to a randomized experimental study of the same covariates and design. The preceding test makes this comparison. If we had randomly assigned pharmacies to register to sell marijuana and we had assessed treatment versus control mean differences in 100 variables, we would have expected 5 variables to have p less than .05 **just through chance**. That is, 5 small p-values out

of 100 would not impugn the design of an experiment — in fact it would be expected. In this case, we see 11 such small p-values — suggesting an overall inconsistency with the experimental standard (not surprising since this is observational data). The omnibus or overall p above attempts to direct attention away from the individual p-values and to focus on the collection of differences. And, we could also have used a multiple testing adjustment for the p-values (which would show no statistically significant differences).

We can also show that, using unadjusted p-values, that these covariates-to-marijuana selling relationships depart somewhat from the patterns of a randomized design by just counting up the number of significant p-values and comparing that number to the expected number under a randomized design.

```
## It looks pretty balanced at least on means!
## Recall the number of p-values less than .05 that we'd expect by chance:
nrow(xb0ionebyone) * .05
```

```
[1] 4.05
```

```
sum(xb0ionebyone[, "p"] <= .05)
```

```
[1] 11
```

```
## So perhaps some imbalance but not a lot.
```

Save products

```
save(xb0i, baselineFmla, baselineFmlaCluster, file = "initial_balance.rda")
```

0.1 References

References

Hansen, B.B. and J. Bowers (2008). "Covariate Balance in Simple, Stratified and Clustered Comparative Studies". In: *Statistical Science* 23, p. 219.