500 Class 05 (Recording)

https://thomaselove.github.io/500-2024/

2024-02-15

Today's Agenda

- The `lindner` example
- Matching with (or without) a Propensity Score
  - Matching with and without replacement
  - Greedy vs. Caliper vs. Optimal Matches
- The `dm2200` example: Executing Matching in R
  - Using `Matching` vs. `MatchIt`

Section 1

Highlights of the `lindner` example

The `lindner` example, Overview

https://thomaselove.github.io/500-examples/#the-lindner-example

The lindner data come from an observational study of 996 patients receiving an initial Percutaneous Coronary Intervention (PCI) at Ohio Heart Health, Christ Hospital, Cincinnati in 1997 and followed for at least 6 months by the staff of the Lindner Center. The patients thought to be more severely diseased were assigned to treatment with abciximab (an expensive, high-molecular-weight IIb/IIIa cascade blocker); in fact, only 298 (29.9 percent) of patients received usual-care-alone with their initial PCI. The lindner data is part of the twang package.

The analysis provided here was developed by Wyatt P. Bensken and Harry Persaud. It involves 12 separate tasks using methods for both quantitative and binary outcomes that are analogous to those in the toy example developed by Professor Love. The main difference is in style, and in the fact that 1:1 matching without replacement performs very poorly.

The `lindner` example, Introduction

The `lindner` example does 1:1 greedy matching on the linear propensity score, in this case however there are more treated subjects than controls, so the default greedy approach taken is to not match all treated subjects, since you run out of controls before you get to them all.

- 1:1 nearest neighbor matches are demonstrated, first without replacement (in Task 4) and then **with** replacement (in Task 6).
- The "with replacement" matching has some appealing features, not least of which being that now we can realistically create ATT estimates (since we match all treated subjects to a control, repeating some controls.)
  - The match quality (in terms of a Love plot) is much better with replacement in this case.
  - This does come with a cost, though. Most subjects are matched just a few times, but some are included more than 50 times!

Lindner: Setup and Data Ingest

- General R packages: broom, janitor, magrittr, naniar, patchwork, tableone
- R packages for analyses: lme4, survey, survival, twang (and broom.mixed)
- R packages for design: cobalt, Matching
- plus the tidyverse (this work predates easystats)

Data provided via csv file on our 500-data page.

- I re-uploaded the data there on 2025-02-04.

Lindner Data

- 996 rows, 12 columns

	A	B	C	D	E	F	G	H	I	J	K	L
1	studynum	lifepres	cardbill	abcix	stent	height	female	diabetic	acutemi	ejecfrac	ves1proc	sixMonthSurvive
2	lin_0001	11.6	14232	0	1	170	0	0	0	60	1	TRUE
3	lin_0002	11.6	10829	1	1	160	1	1	0	50	1	TRUE
4	lin_0003	11.6	12259	0	1	180	0	0	0	45	1	TRUE

- Sample: 996 patients receiving an initial Percutaneous Coronary Intervention (PCI) at Ohio Heart Health, Christ Hospital, Cincinnati in 1997 and followed for at least 6 months by Lindner Center staff.
- Quantitative Outcome contained in `cardbill`
- Binary Outcome contained in `sixMonthSurvive`
- Treatment Status contained in `abcix`
- Seven Covariates: `stent`, `height`, `female`, `diabetic`, `acutemi`, `ejecfrac` and `ves1proc`
- Somehow, no missing data.

Lindner Covariates

Variable	Description
height	height in cm
female	1 = female, 0 = not female
diabetic	1 = diabetes diagnosis, 0 = no
acutemi	1 = acute MI in past 7 days, 0 = no
ejecfrac	left ventricular ejection fraction, as %
stent	1 = stent deployed, 0 = no
ves1proc	vessels involved in initial PCI (0-5)

Table 1 (Covariates)

Variable	Treated	Control	p
n	698	298	–
height	171.4 (10.7)	171.5 (10.6)	0.996
sex = Male	467 (66.9)	183 (61.4)	0.111
diabetic = 1	143 (20.5)	80 (26.8)	0.034
acutemi = 1	125 (17.9)	18 (6.0)	< 0.001
stent = 1	492 (70.5)	174 (58.4)	< 0.001
ejecfrac	50.4 (10.4)	52.3 (10.3)	0.009
ves1proc	1.46 (0.71)	1.20 (0.48)	< 0.001

Lindner Outcomes and Treatment

Variable	Description
studynum	subject identifying code
lifepres	recoded into sixMonthSurvive, so we'll ignore
sixMonthSurvive	survival at six months (recoded lifepres)
cardbill	cardiac-related costs incurred within 6m of patient's initial PCI, in 1998 dollars
abcix	relabeled as treated: 1 = augmented treatment with abciximab, 0 = usual care

Stratified Outcomes

Variable	Treated	Control	p
n	698	298	–
cardbill	16127 (9384)	14614 (14514)	0.051
sixMonthSurvive = 1	687 (98.4)	283 (95.0)	0.004

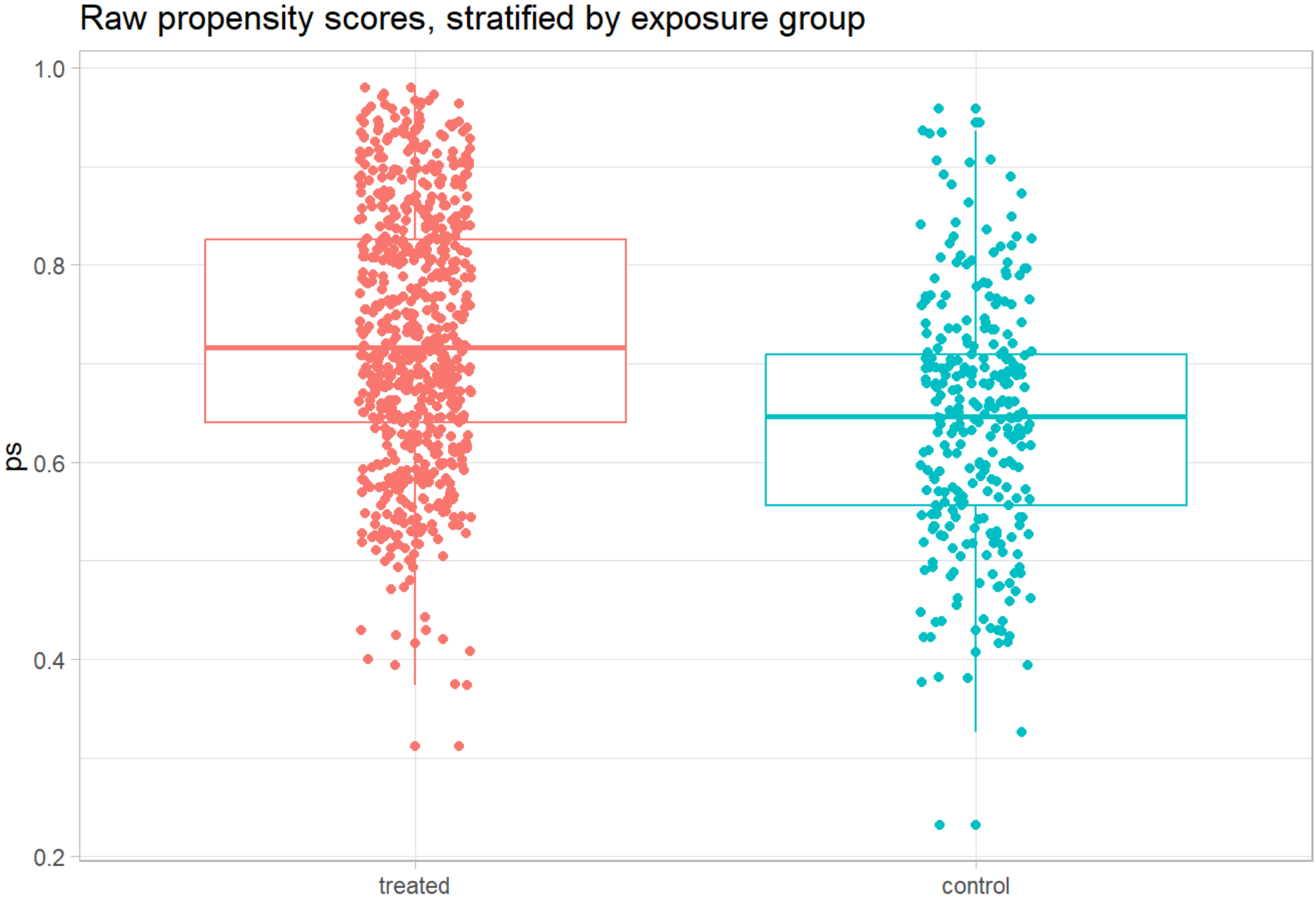
- I've rearranged some factors so they're a little easier to think about, as compared to the `lindner` example online.
- Linear Model: Treated subjects estimated to spend $1512.46 (95% CI: -3.83, 3028.76) more (unadjusted analysis).
- Logistic Model: Odds of "alive after 6m" in treated subjects are 3.31 (95% CI: 1.51, 7.48) times higher than controls (unadjusted).

Lindner Propensity Model

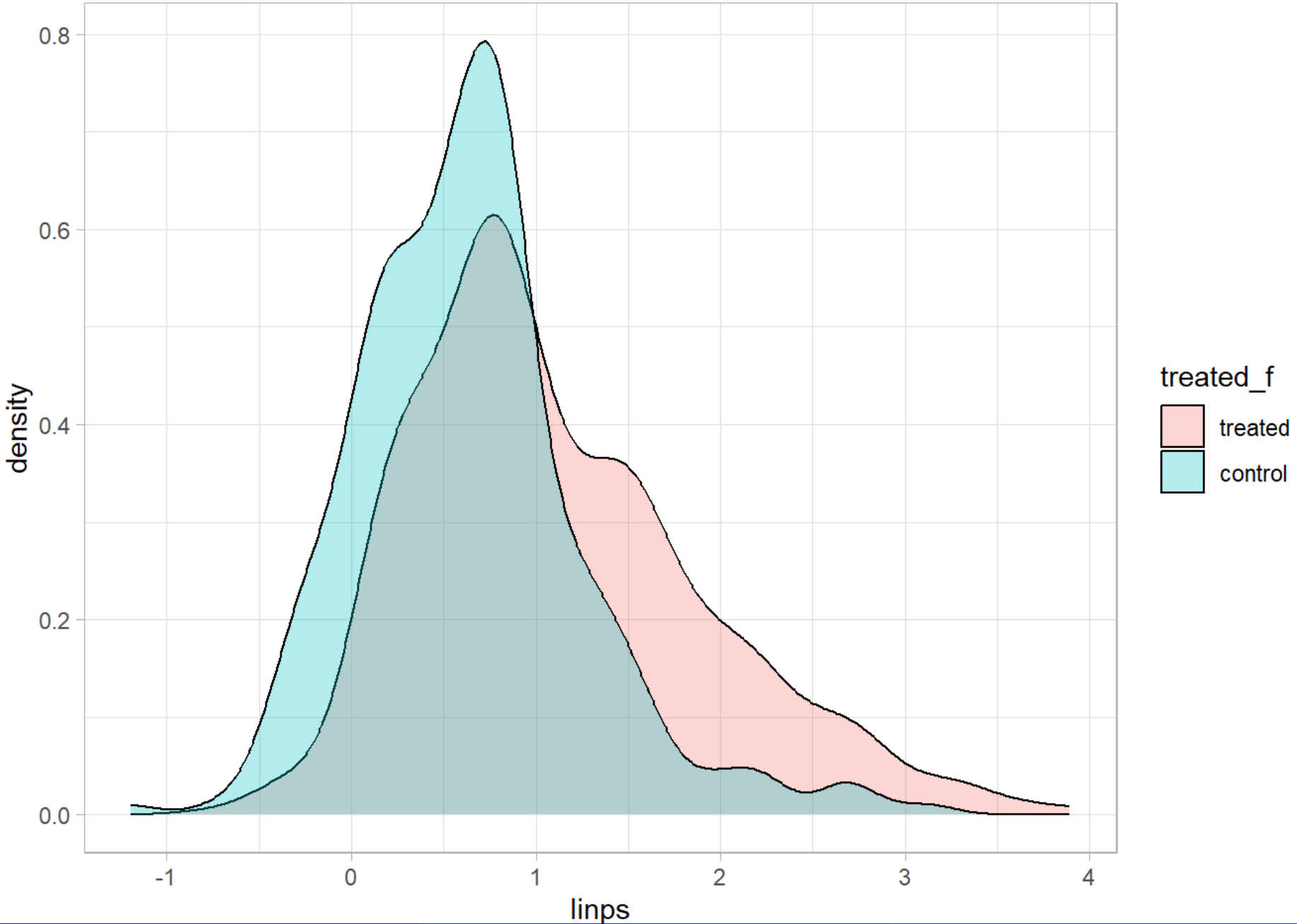
```
psmodel <- glm(treated ~ stent + height + female + diabetic +
                         acutemi + ejecfrac + ves1proc,
               family = binomial(), data =lindner_clean)

lindner_clean$ps <- psmodel$fitted
lindner_clean$linps <- psmodel$linear.predictors
```

- The estimated (raw) propensity scores range from 0.232 to 0.980, across the 996 subjects.

From Section 7.1.1 (Raw PS)



From Section 7.1.2 (Linear PS)



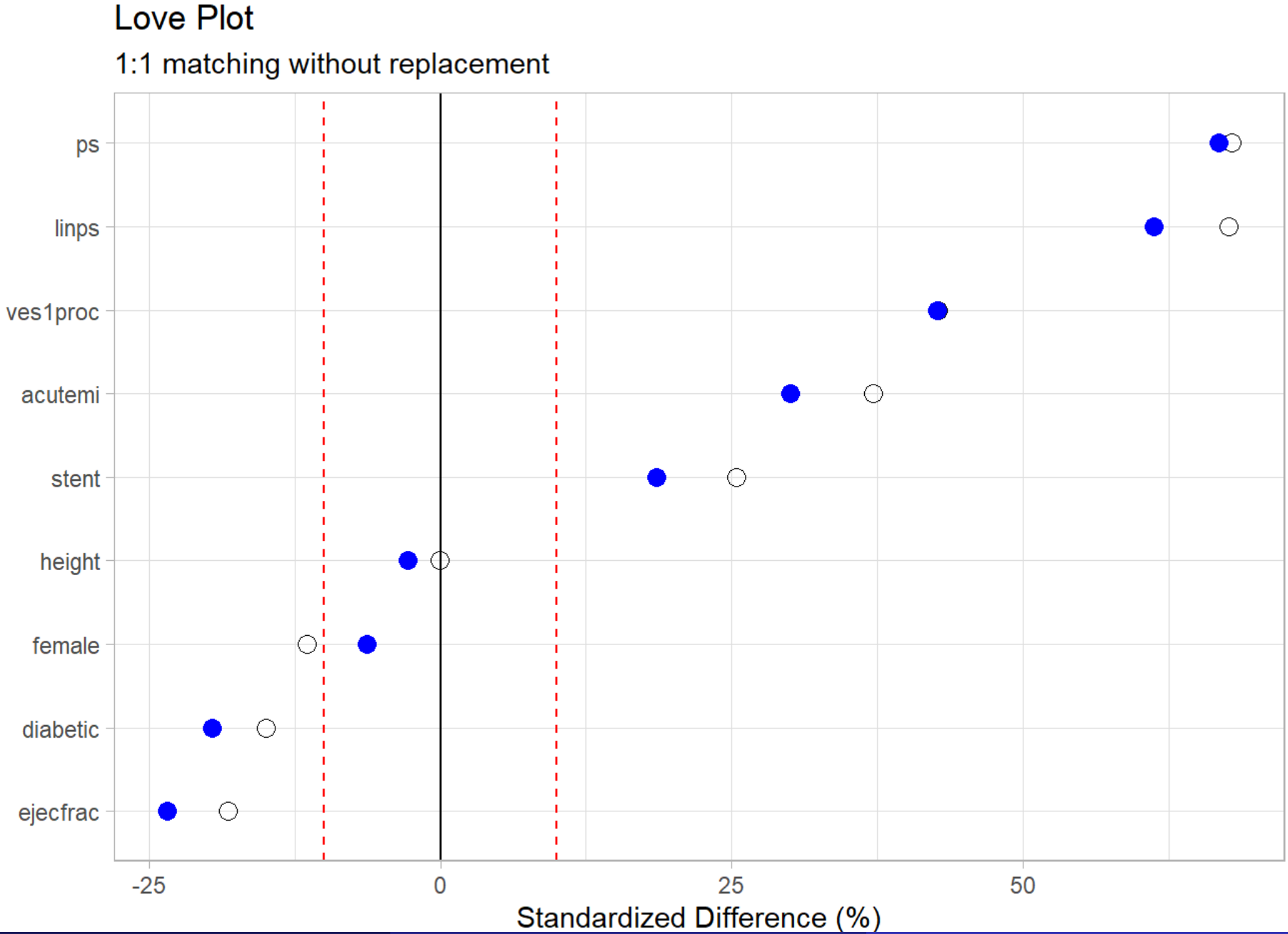
Greedy 1:1 Match on linear PS (Section 9)

Matching without replacement...

```
X <- lindner_clean$linps ## matching on the linear propensity
Tr <- as.logical(lindner_clean$treated)
match1 <- Match(Tr=Tr, X=X, M = 1, replace=FALSE, ties=FALSE)
```

> *Warning in Match(Tr = Tr, X = X, M = 1, replace = FALSE, ties = FALSE): replace==FALSE, but there are more (weighted) treated obs than control obs. Some treated obs will not be matched. You may want to estimate ATC instead.*

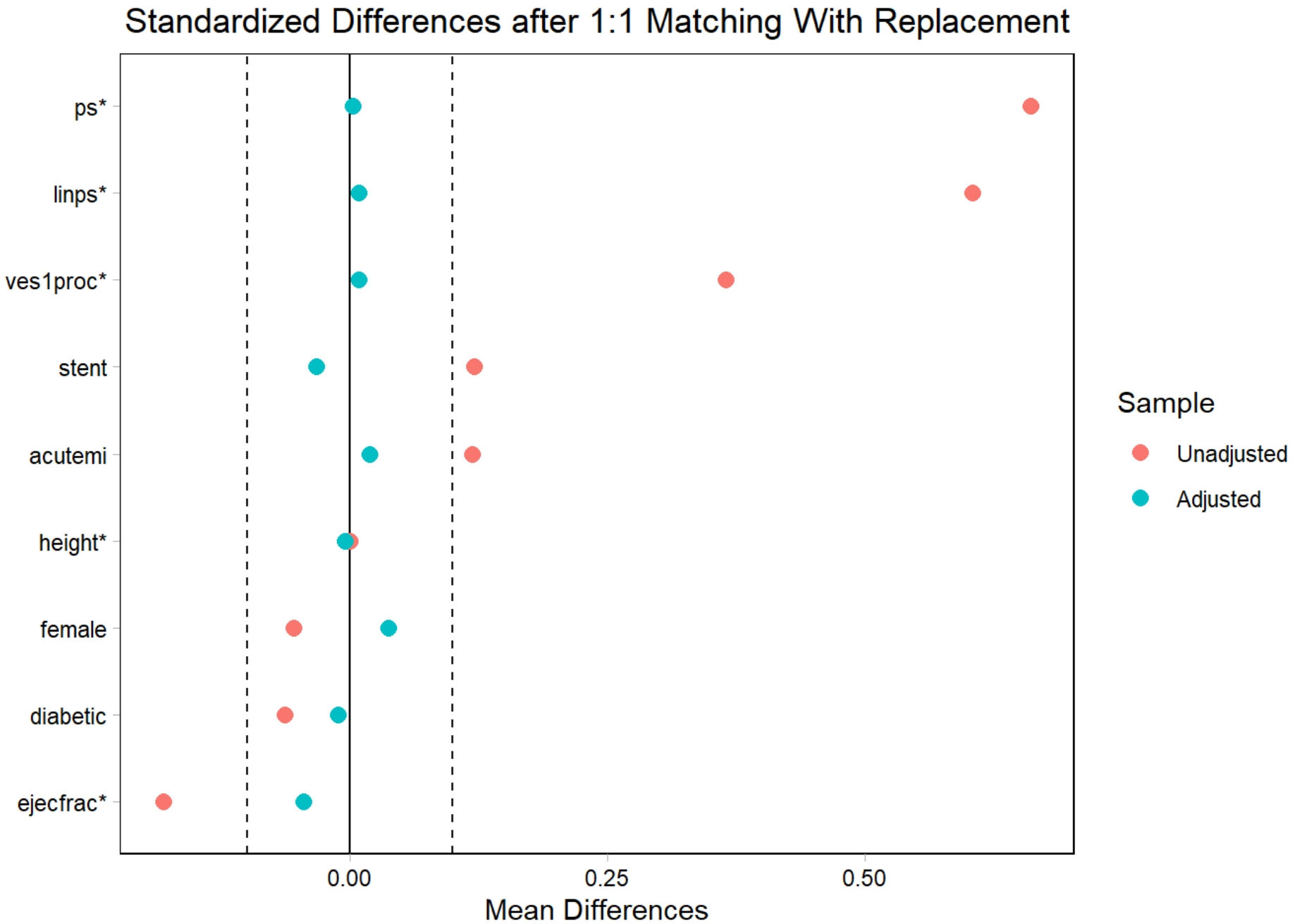
Love plot after 1:1 match without replacement (9.2)



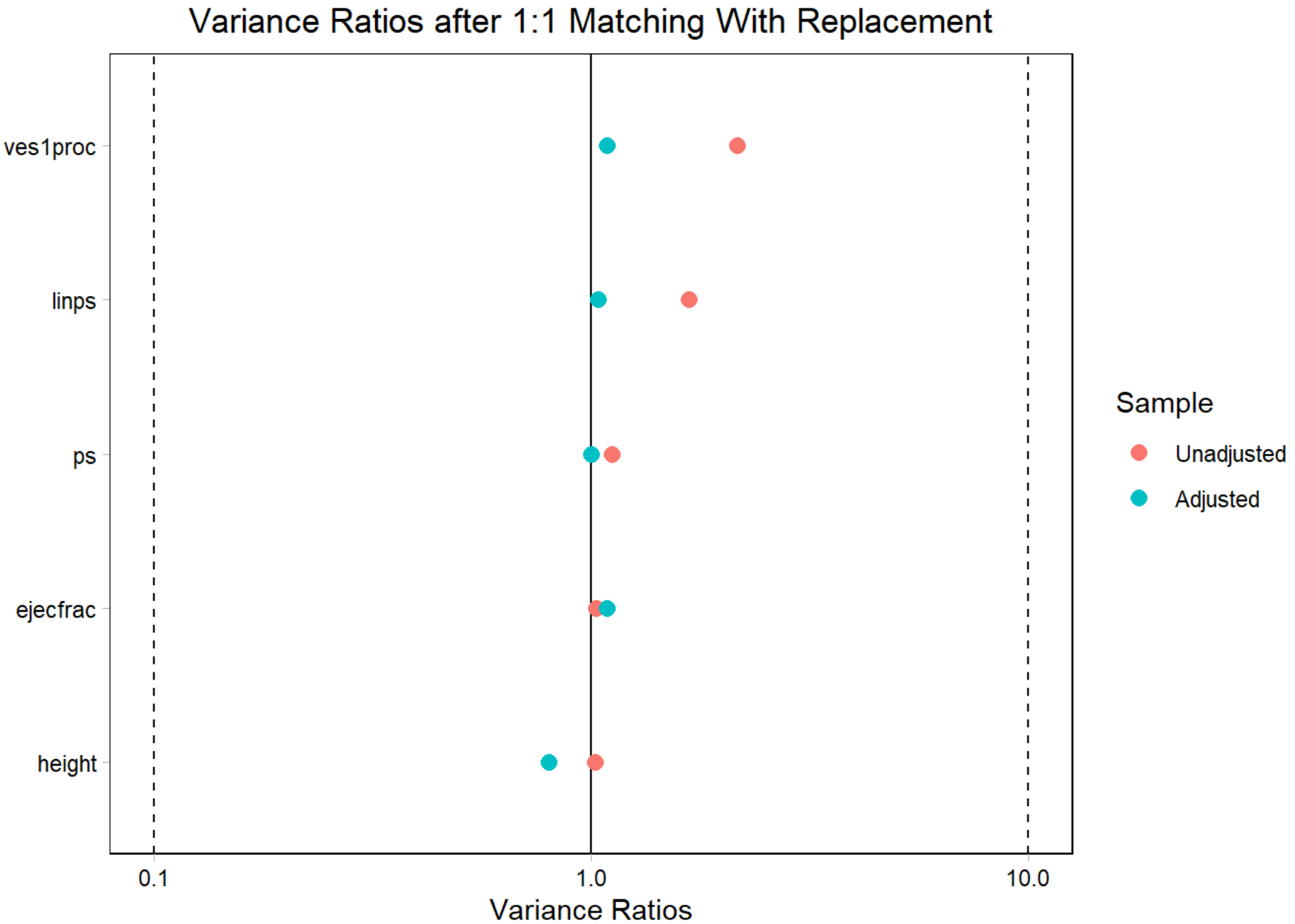
1:1 match with replacement (Section 11)

```
X <- lindner_clean$linps
Tr <- as.logical(lindner_clean$treated)
match1 <- Match(Tr=Tr, X=X, M = 1, replace=TRUE, ties=FALSE)
```

Love Plot after 1:1 with replacement (11.3)



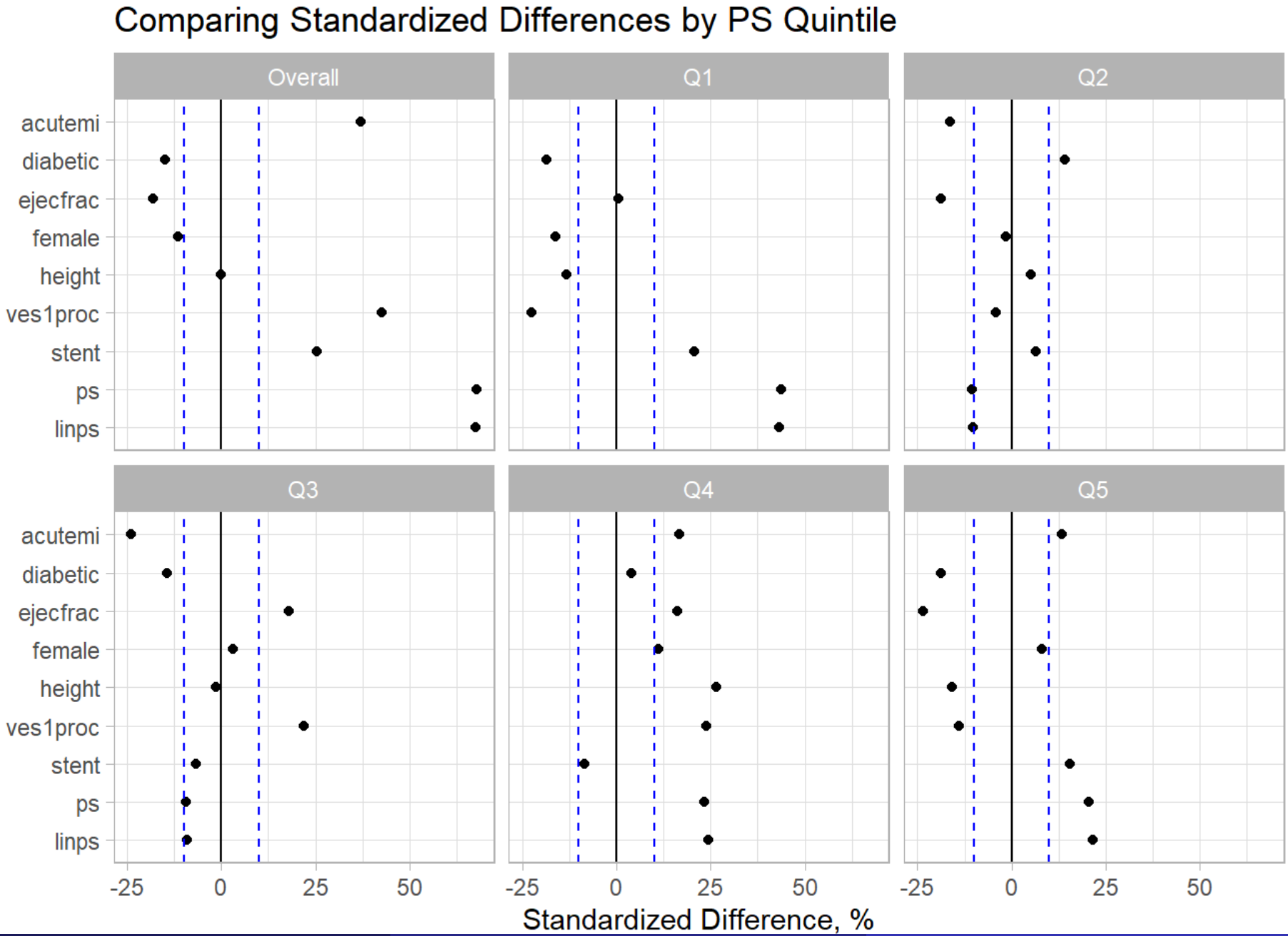
Variance Ratios after 1:1 with replacement (11.4)



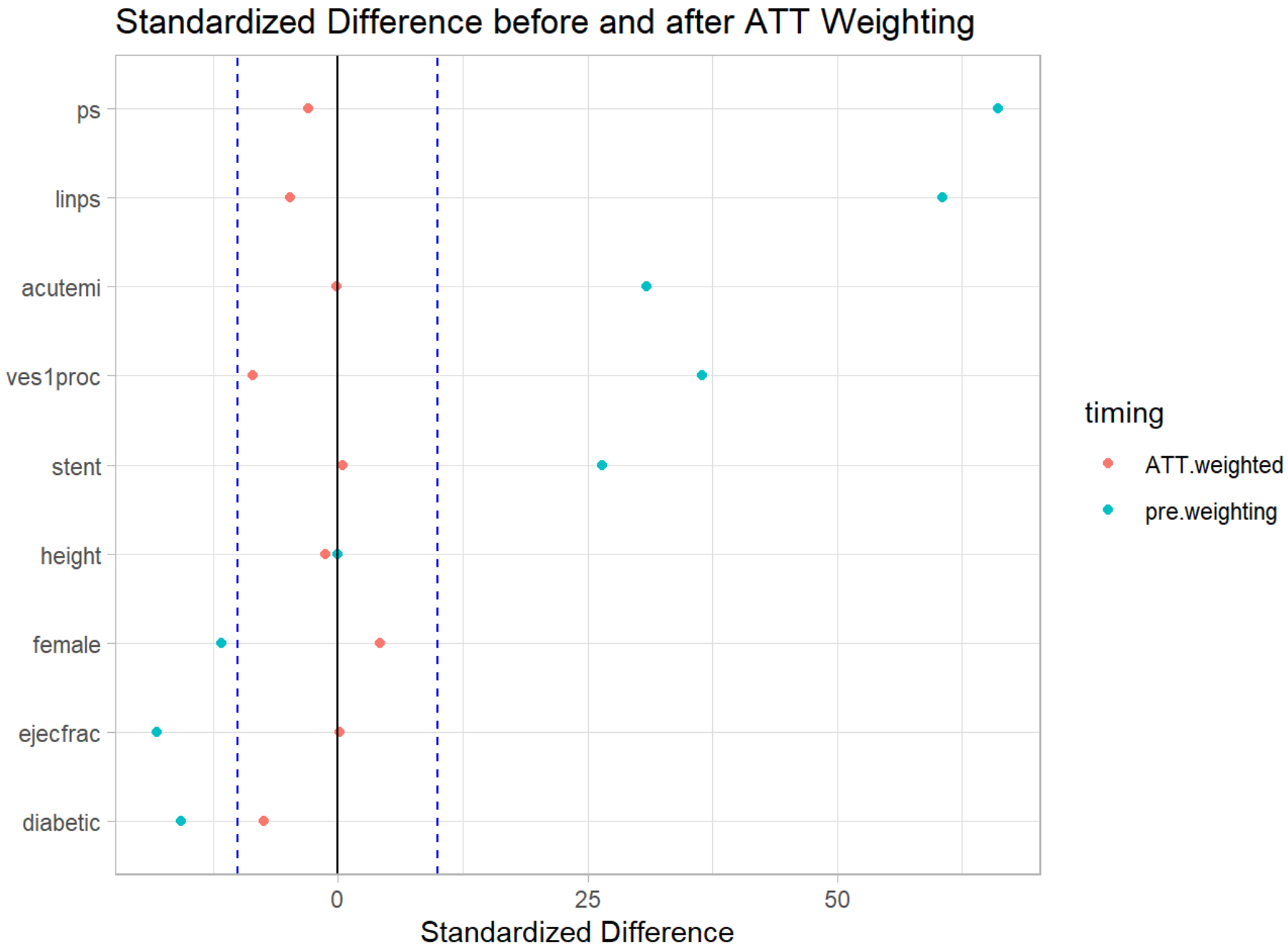
After Matching: Outcome Results

- After 1:1 matching without replacement (698 matches: lousy covariate balance)
  - so we didn't look at the outcomes
- After 1:1 matching with replacement (698 matches: good covariate balance)
  - In the matched sample, mean cardbill for treated was $16,127 and was $16,300 for controls.
  - Treated subjects were estimated to spend $-173.03 (95% CI:-1482.29, 1136.22) as compared to non-treated subjects.
  - The odds of "alive after 6m" were 5.6 times higher in treated subjects (95% CI: 2.86, 10.98).

Subclassification into 5 Quintiles



ATT Weighting



Key Outcome Results

After subclassification into 5 quintiles:

- Treated subjects estimated to spend $732.67 more (95% CI:-876.5, 2341.85) than controls.
- The odds of "alive after 6m" were 4.33 (95% CI: 1.68, 11.19) times higher in treated subjects.

After ATT (Average Treatment Effect on the Treated) Weighting

- Treated subjects estimated to spend $239.28 less (95% CI:-3019.92, 2541.37) than controls.
- The odds of "alive after 6m" were 6.5 (95% CI: 2.27, 18.65) times higher in treated subjects.

Key Outcome Results

After direct adjustment for the linear propensity score:

- Treated subjects estimated to spend $1167.90 more (95% CI:-412.22, 2748.02) than controls.
- The odds of "alive after 6m" were 4.64 (95% CI: 1.99, 11.27) times higher in treated subjects.

After ATT Weighting and direct adjustment (double robust):

- Treated subjects estimated to spend $126.72 less (95% CI:-2514.24, 2260.80) than controls.
- The odds of "alive after 6m" were 6.9 (95% CI: 2.28, 20.84) times higher in treated subjects.

Section 2

Matching Approaches (discussion built on Austin, 2014)

1:1 Greedy Matching

Greedy (or nearest neighbor) matching selects a treated subject and then selects as a matched control subject the untreated subject whose propensity score is closest to that of the treated subject. If multiple untreated subjects are equally close to the treated subject, one of these untreated subjects is selected at random, typically. Options include:

1. Select treated subjects from highest to lowest propensity score.
2. Select treated subjects from lowest to highest propensity score.
3. Select sequentially treated subjects in the order of the best possible match.
    - First treated subject is the one who is closest to an untreated subject.
    - Second treated subject is the one closest to the remaining untreated, etc.
4. Select treated subjects in a random order. Set a fixed random number seed so that the matched sample is reproducible in subsequent analyses.

Results in all treated subjects being matched to a single control.

Greedy Matching with Replacement

Matching *without* replacement means that once an untreated subject has been matched to a treated subject that untreated subject is no longer eligible for further matches to other treated subjects. As a result, each subject can be in at most one matched pair.

Now, in matching *with* replacement, we allow members of the "control" pool to be reused in the matching process.

- The process is somewhat simpler in the nearest neighbor case - just match each treated subject to the closest untreated subject.
- Because untreated subjects are recycled and thus can be included in multiple matched sets, the order in which the treated subjects are selected has no effect on the formation of matched pairs.

Matching 1:k rather than 1:1

Here, we simply try to obtain the k best matching untreated subjects for each treated subject.

- In greedy matching, it is certainly possible that the quality of matches will drop considerably with extra matches, especially near the edges of the distribution of the propensity score.
- 1:k matching is occasionally done with replacement, but of course we still want k unique matched untreated subjects for each treated subject.

Caliper Matching

Match subjects only if they fall within a pre-specified maximum distance (the caliper distance.)

- When using caliper matching, we usually match subjects on the logit of the propensity score using a caliper width as a proportion of the standard deviation of the logit of the propensity score.
- Caliper matching can be combined with other distance metrics (where, for example, a few specific covariates are targeted for more precise matching.)
- Matching with a caliper can be accomplished with or without replacement, and in 1:1 or 1:k settings.

Optimal Matching

The main distinction that matters is between optimal matching approaches and nearest-neighbor (greedy) matching approaches.

- Optimal matching forms matched pairs so as to minimize the average within-pair difference in propensity scores.
- Optimal matching is rarely the first way I run an analysis (it's a bit slow, especially with large matching problems) but this problem is disappearing as smarter people and more effective computers emerge.

Double Robust Approaches

Nothing is stopping us from using regression adjustment along with matching. It's not unusual to consider the incorporation of the linear propensity score, or an important set of prognostic covariates in a setting where we are analyzing propensity-matched subjects.

Peter Austin's (2014) Comparison

A comparison of 12 algorithms for matching on the propensity score

Peter C. Austin[a,b,c*†]

Propensity-score matching is increasingly being used to reduce the confounding that can occur in observational studies examining the effects of treatments or interventions on outcomes. We used Monte Carlo simulations to examine the following algorithms for forming matched pairs of treated and untreated subjects: optimal matching, greedy nearest neighbor matching without replacement, and greedy nearest neighbor matching without replacement within specified caliper widths. For each of the latter two algorithms, we examined four different sub-algorithms defined by the order in which treated subjects were selected for matching to an untreated subject: lowest to highest propensity score, highest to lowest propensity score, best match first, and random order. We also examined matching with replacement. We found that (i) nearest neighbor matching induced the same balance in baseline covariates as did optimal matching; (ii) when at least some of the covariates were continuous, caliper matching tended to induce balance on baseline covariates that was at least as good as the other algorithms; (iii) caliper matching tended to result in estimates of treatment effect with less bias compared with optimal and nearest neighbor matching; (iv) optimal and nearest neighbor matching resulted in estimates of treatment effect with negligibly less variability than did caliper matching; (v) caliper matching had amongst the best performance when assessed using mean squared error; (vi) the order in which treated subjects were selected for matching had at most a modest effect on estimation; and (vii) matching with replacement did not have superior performance compared with caliper matching without replacement. © 2013 The Authors. ***Statistics in Medicine*** **published by John Wiley & Sons, Ltd.**

Keywords: propensity score; matching; computer algorithms; optimal matching; Monte Carlo simulations; propensity-score matching

- You'll find this article on our Sources page.

Matching without Replacement

We first consider methods based on matching without replacement. Using this approach, we matched each untreated subject to at most one treated subject. Once an untreated subject has been matched to a treated subject, that untreated subject is no longer eligible for consideration as a match for other treated subjects.

- The primary distinction between different matching algorithms that use matching without replacement is between optimal matching and methods based on greedy matching.
  - Optimal matching forms matched pairs so as to minimize the average within-pair difference in propensity scores.
  - In contrast, greedy nearest neighbor matching selects a treated subject and then selects as a matched control subject, the untreated subject whose propensity score is closest to that of the treated subject (if multiple untreated subjects are equally close to the treated subject, one of these untreated subjects is selected at random).

Caliper Matching

A modification to greedy nearest neighbor matching is greedy nearest neighbor matching within specified caliper widths.

- In this modification to greedy nearest neighbor matching, we can match treated and untreated subjects only if the absolute difference in their propensity scores is within a prespecified maximal distance (the caliper distance).
  - When using caliper matching, we matched subjects on the logit of the propensity score using a caliper width that was defined as a proportion of the standard deviation of the logit of the propensity score.
  - We match on the logit of the propensity score because the reduction in bias due to the use of different caliper widths is better understood.

Matching with replacement

Matching with replacement permits the same untreated subject to be matched to multiple treated subjects.

- We considered nearest neighbor matching with replacement and nearest neighbor matching within specified caliper widths with replacement.
- Each of these approaches simply matches each treated subject to the nearest untreated subject (subject to possible caliper restrictions).
- Because untreated subjects are recycled or allowed to be included in multiple matched sets, the order in which the treated subjects are selected has no effect on the formation of matched pairs.

Austin's Case Study

Sample of 9107 patients discharged from 103 acute care hospitals in Ontario, Canada, with a diagnosis of AMI (or heart attack) between April 1, 1999 and March 31, 2001. We collected data on these subjects as part of the Enhanced Feedback for Effective Cardiac Treatment (EFFECT) study, an initiative intended to improve the quality of care for patients with cardiovascular disease.

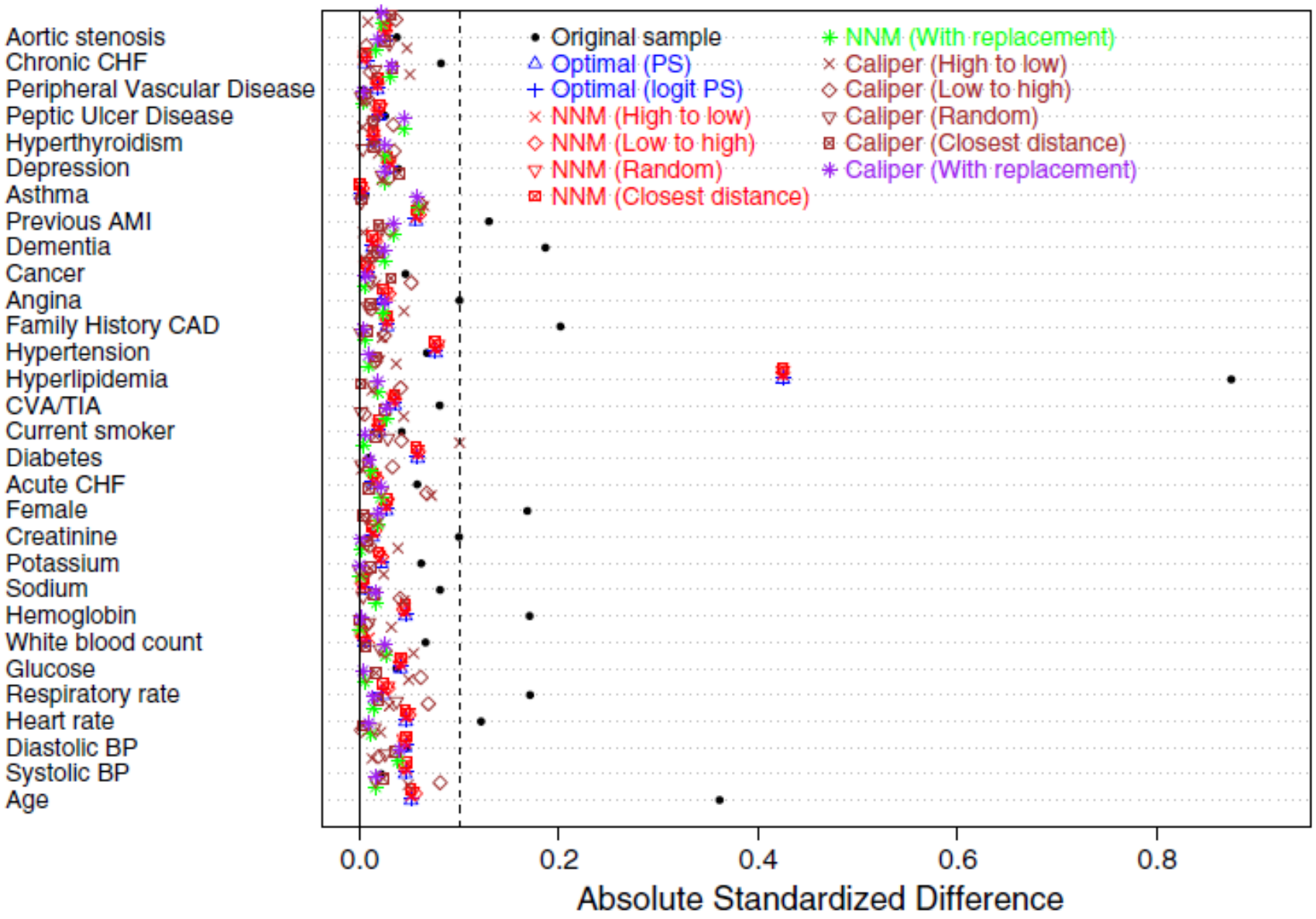
- For the current case study, the exposure was whether the patient received a prescription for a statin lipid-lowering agent at hospital discharge. 3049 (33.5%) patients received such a prescription.
- The outcome was a binary variable denoting whether the patient died within 8 years of hospital discharge. 3593 (39.5%) patients died in that window.

Case Study: Propensity Score Matching

We estimated a propensity score for statin treatment using logistic regression to regress an indicator variable denoting statin treatment on 30 baseline covariates.

- We used restricted cubic smoothing splines to model the relationship between each of the 11 continuous covariates and the log-odds of statin prescribing.
- We used each of the matching algorithms under study to form matched samples consisting of pairs of treated and untreated subjects.

Balance of Covariates Before/After Match



Details on Balance Reduction

- In the original sample, 10 of the 30 covariates had standardized differences that exceeded 0.10, with the largest standardized difference being for history of hyperlipidemia (0.88).
- Optimal matching and the nearest neighbor matching without replacement algorithms resulted in substantially improved balance. In the matched samples, 29 of the 30 covariates had standardized differences that were less than 0.10. The standardized difference for hyperlipidemia remained large (0.43) in these matched samples.
- The four algorithms based on caliper matching without replacement resulted in substantial reductions in imbalance: all standardized differences were less than 0.101. Of these algorithms, caliper matching (closest distance) resulted in the best balance (all standardized differences were less than 0.04).

Austin's conclusions re: 12 Algorithms

- Larger numbers of matched pairs (from complete optimal or complete greedy matches) yields more precise estimates than smaller numbers of matched pairs (say, when a caliper is used and only some treated subjects are matched.)
- Caliper matching often yields better "balance" and less biased estimates as compared to other algorithms.
- So we have a bias - variance tradeoff in our estimation strategies, but in terms of MSE, caliper matching usually performs pretty well.
- In terms of ordering of treated subjects for greedy matching or caliper matching, random selection is competitive with other options.
- Optimal matching is pretty comparable to nearest neighbor matching with random selection order, and in fact, it's not clearly any better than that approach.

Section 3

The dm2200 Example

The dm2200 example

This example is solely about matching, and not things like subclassification, weighting and so on. The data are simulated, again, and I'm not really focusing on the outcome models, but rather just on demonstrating how to do the matching, and how to evaluate the quality of the covariate balance.

- Simulated data on 2200 adults with diabetes: exposure is A (n = 200) or B (n = 2000), and two outcomes (BP < 140/90: binary, and BMI: quantitative.)

https://thomaselove.github.io/500-examples/dm2200_analysis.html

dm2200: First Four Matches

We demonstrate the use of both the `Matching` package and the `MatchIt` package. First, using `Matching`, which is what `toy` and `lindner` use, we show:

1. 1:1 nearest neighbor matching without replacement
2. 1:2 nearest neighbor matching without replacement
3. 1:3 nearest neighbor matching with replacement
4. 1:1 nearest neighbor matching without replacement within a caliper on the (linear) propensity score

dm2200: Four Additional Matches

Using the `MatchIt` package, we demonstrate:

1. 1:1 nearest neighbor matching without replacement
2. 1:1 nearest neighbor matching without replacement within a caliper on the (linear) propensity score
3. 1:1 optimal matching without replacement
4. 1:2 optimal matching without replacement

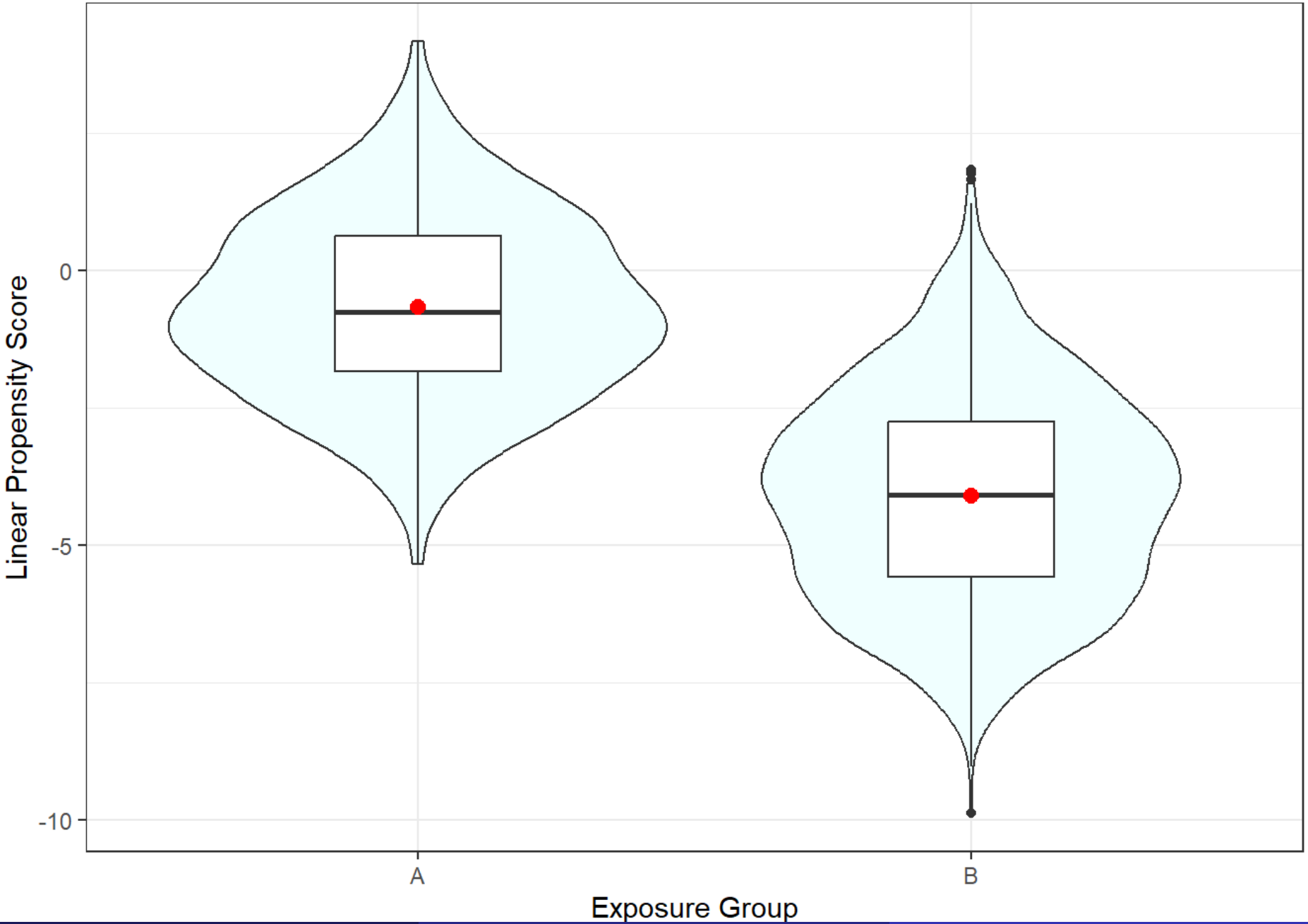
There are multiple other packages in the world for propensity matching, and `cobalt` describes and supports several of them, but these are the two I have most often used.

- `MatchIt` has some features I really like, in particular, it's easier to work with it in `cobalt`, I think, and it also has one very annoying feature, in that it's hard to get the matched data set from it.

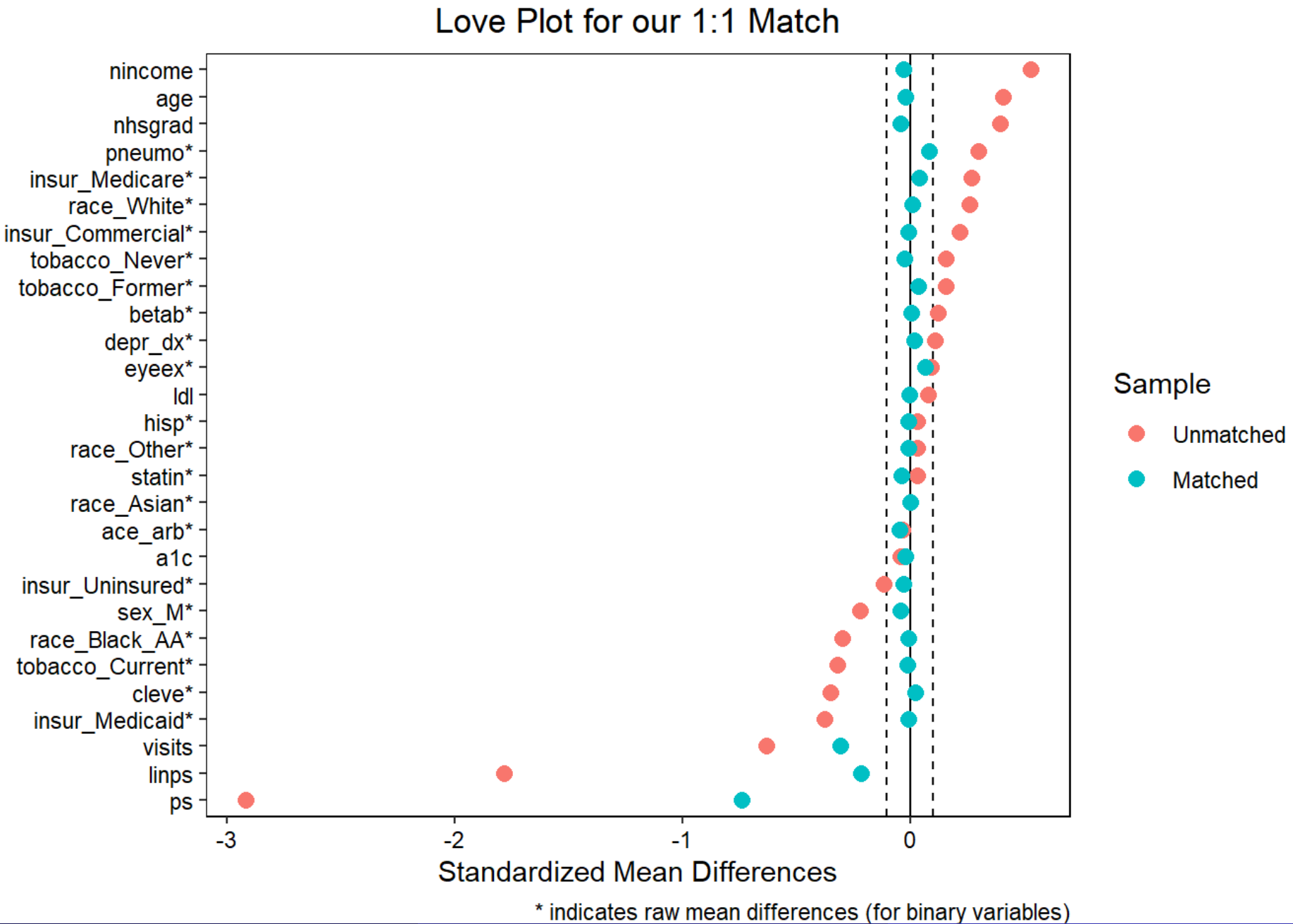
What's New in dm2200?

- Use of tabs in the HTML report (see section 1.2)
- Use of `data_codebook()` from easystats framework
- Use of Table 1 to compare A to B
- Use of `f.build()` from cobalt package
- Use of `model_parameters()` and `model_performance()` from easystats framework

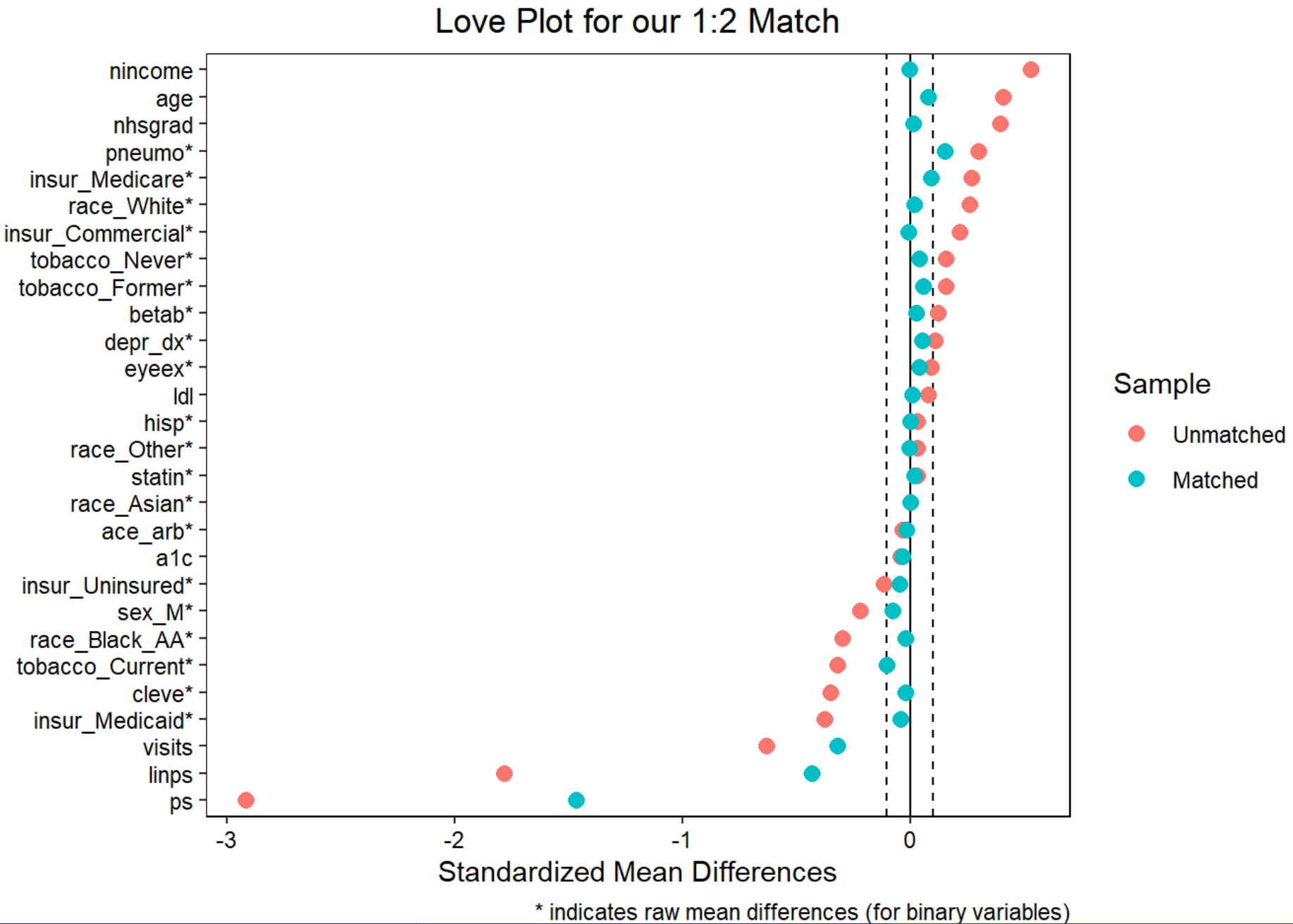
Comparing Linear PS prior to matching



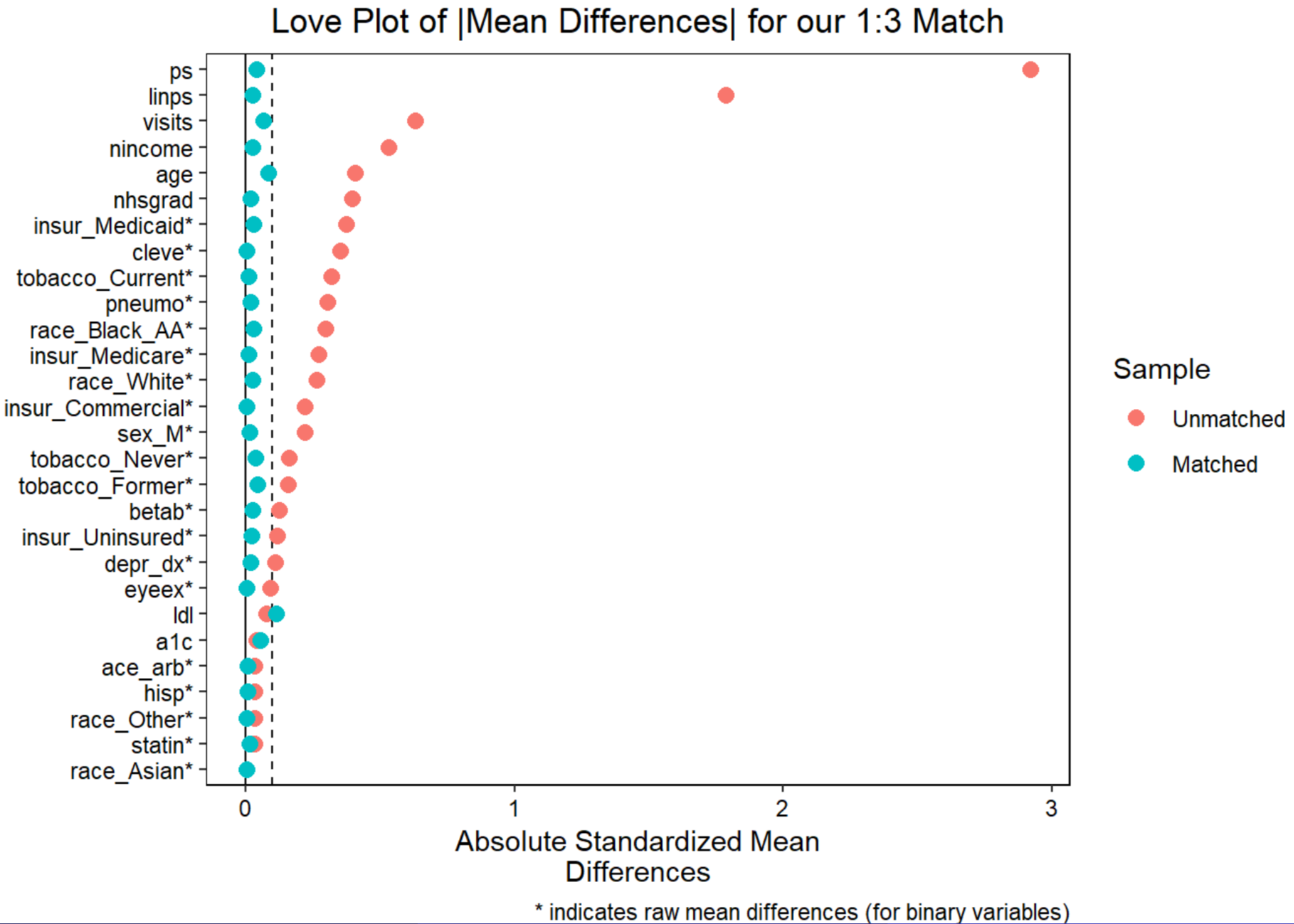
Love Plot for 1:1 greedy match (200 matches)



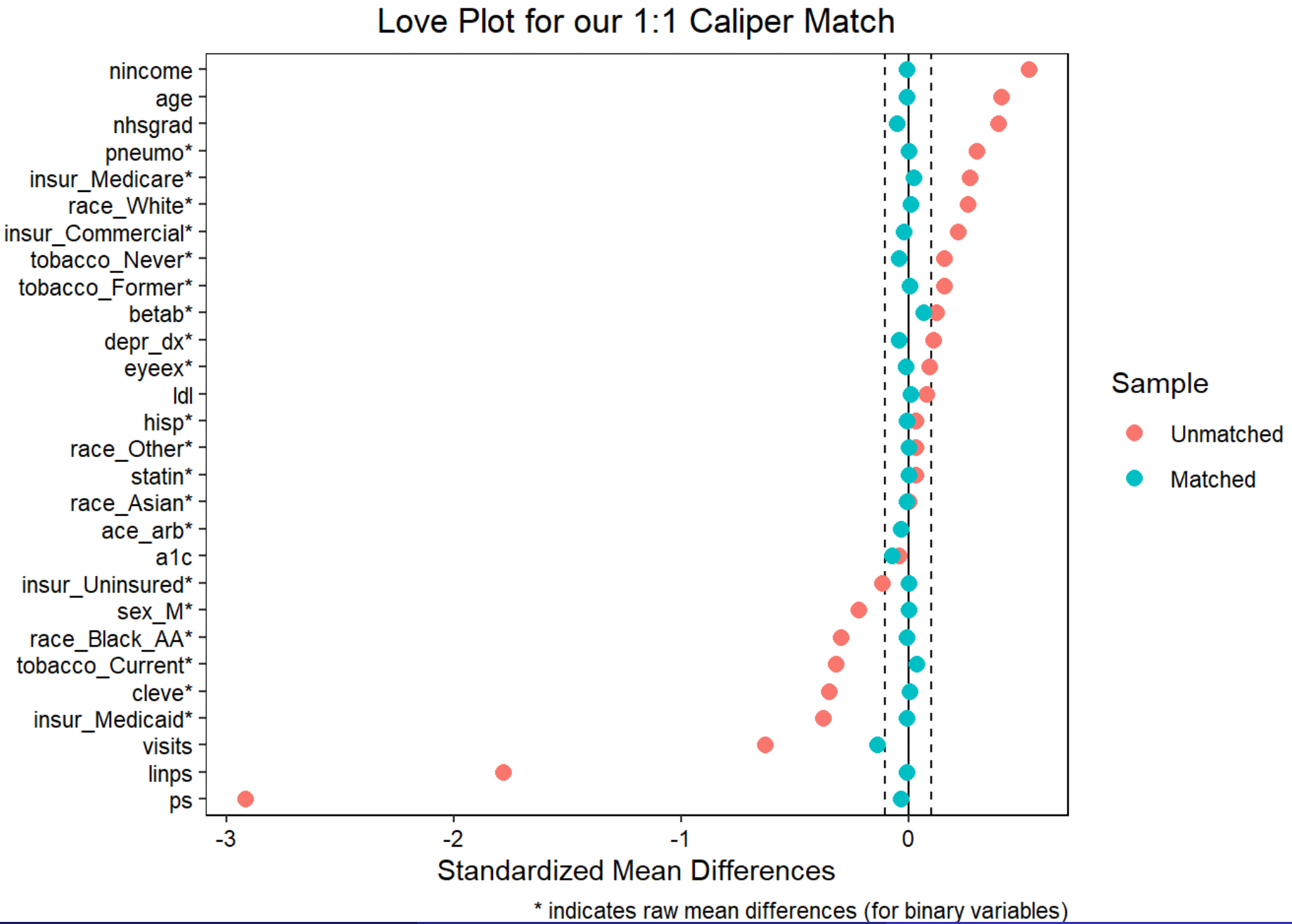
Love Plot for 1:2 greedy match (200 matched to 400)



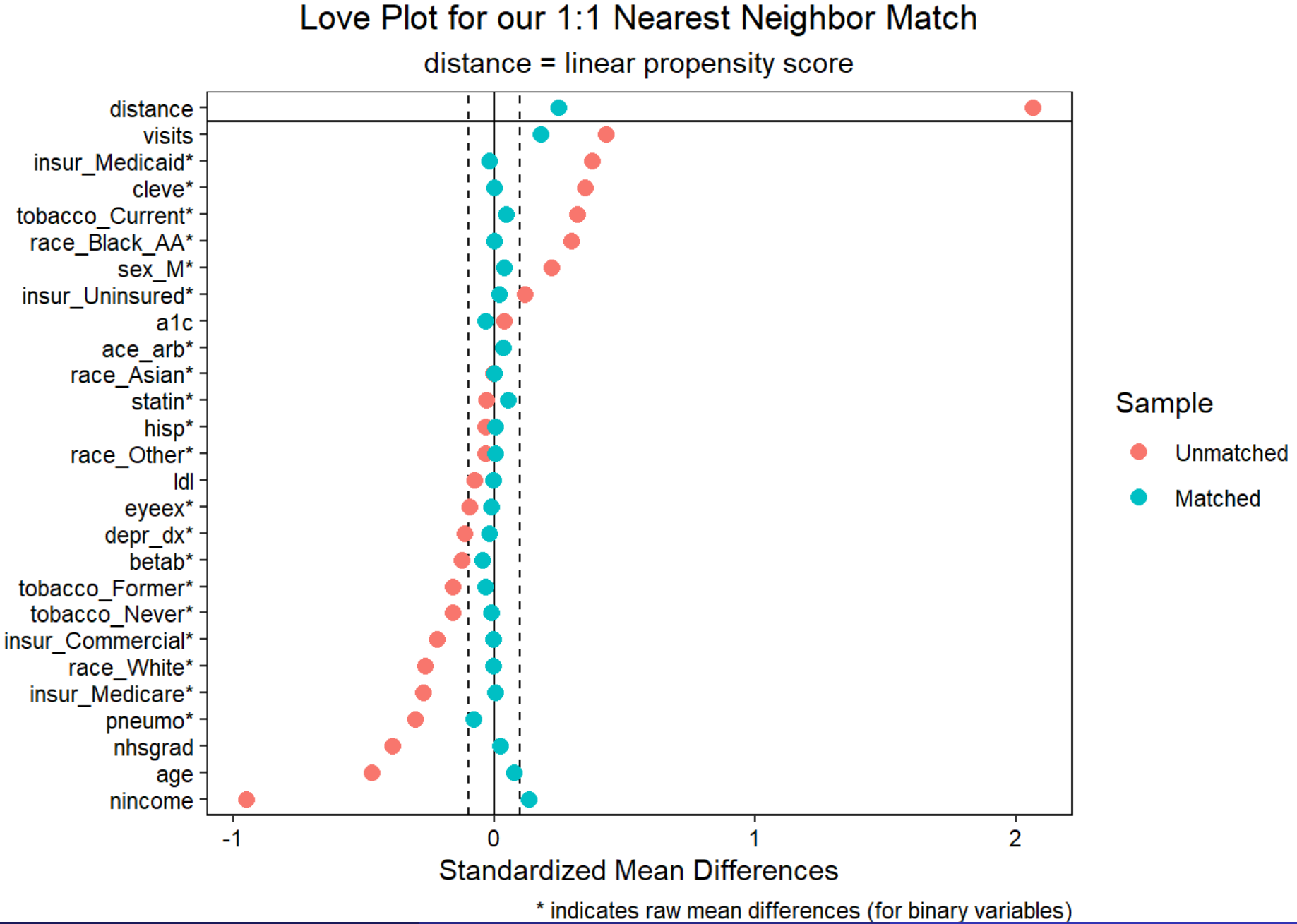
Love Plot for 1:3 greedy match (200 matched to 600)



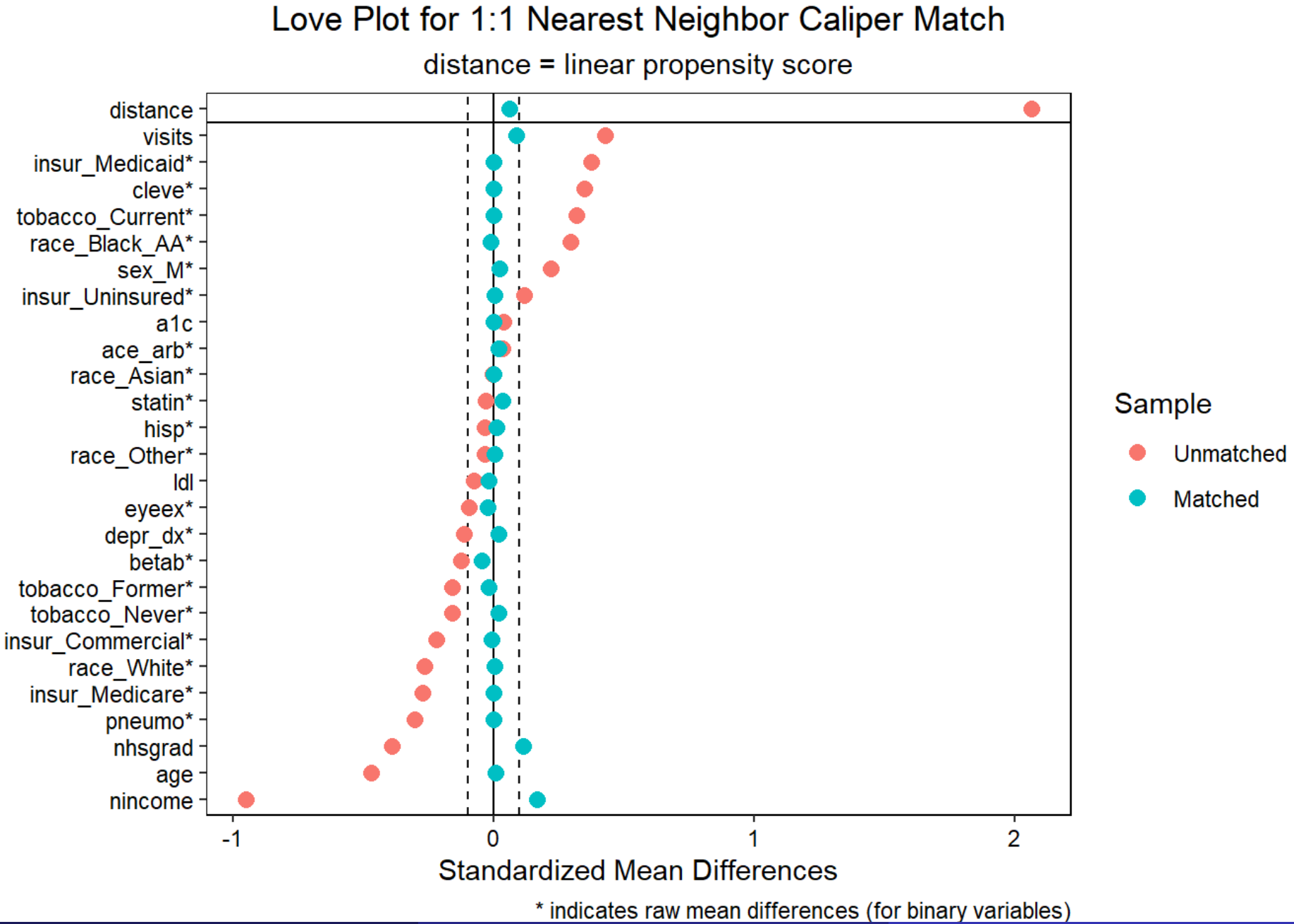
Caliper Match (0.2 SD) produces 162 matches



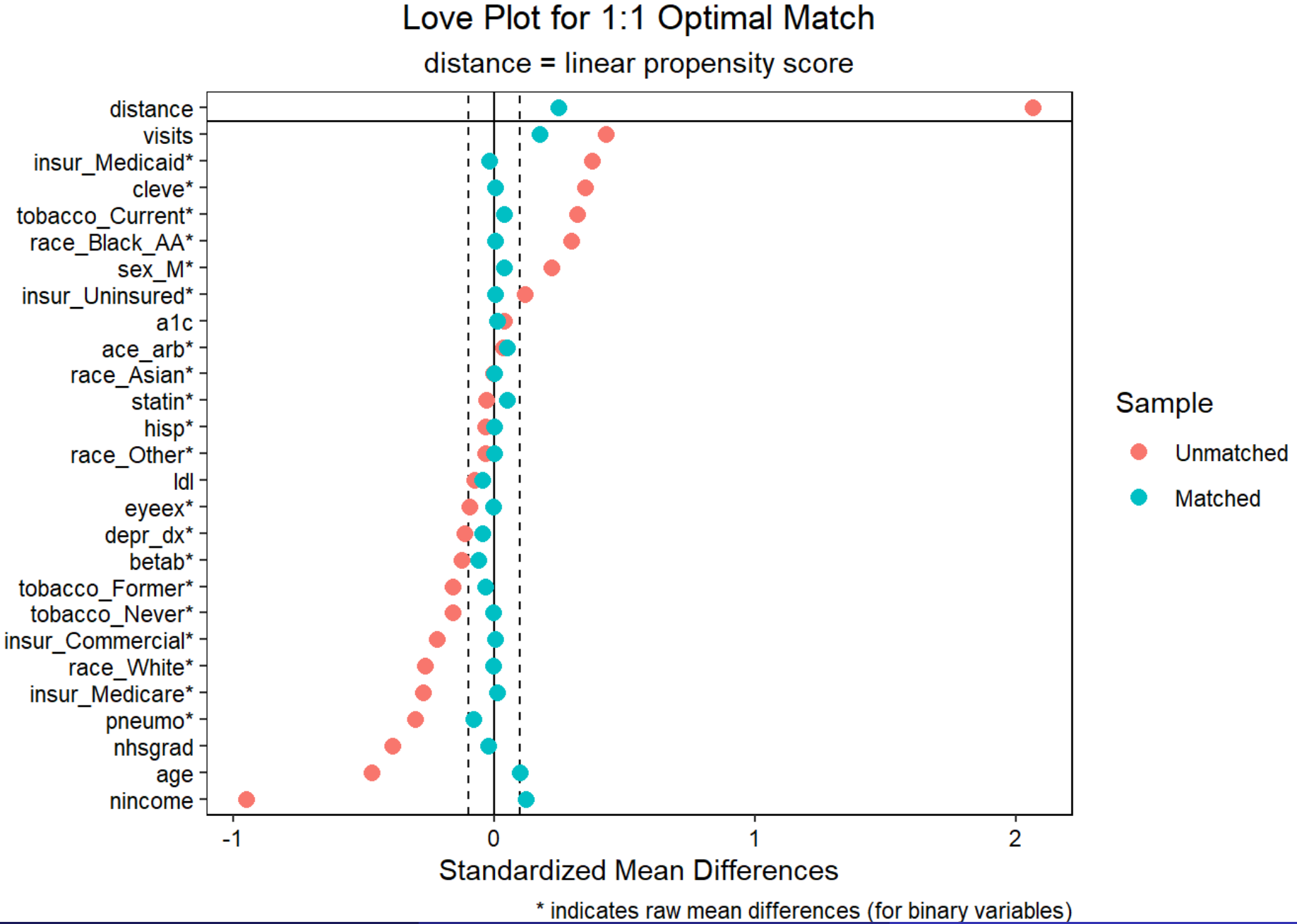
MatchIt nearest neighbor 1:1 (200 matches)



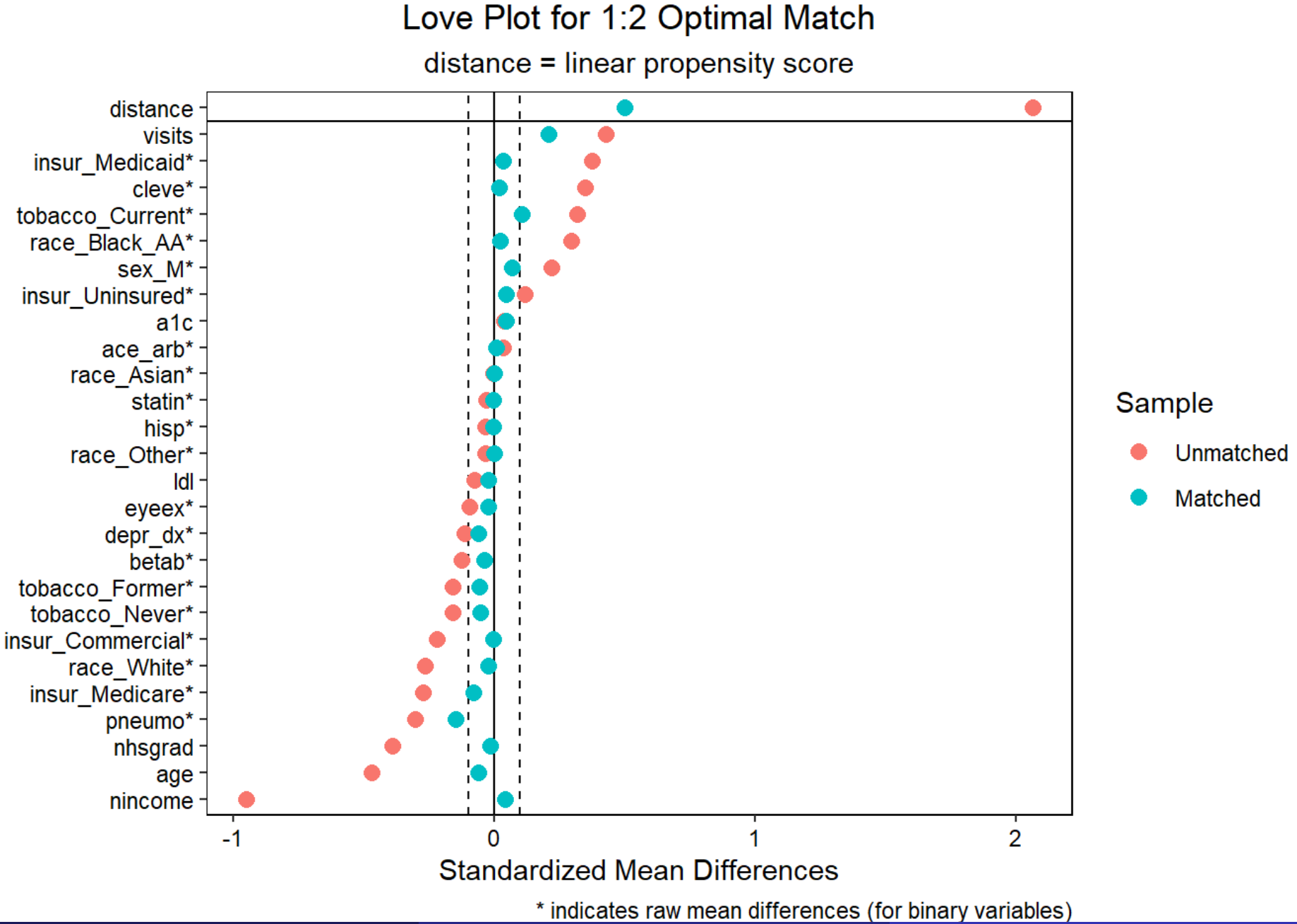
MatchIt nearest neighbor caliper 1:1 (172 matches)



MatchIt 1:1 Optimal Match (200 matches)



MatchIt 1:2 Optimal Match (200 matched to 400)



Outcome Models

Unadjusted results:

- Odds of "good BP" appear lower (OR = 0.56, 95% CI: 0.42, 0.76) in "A" subjects than in "B" subjects
- BMI is lower in "A" subjects by 2.26 kg/m^2 (95% CI: -3.45, -1.07) than in "B" subjects

Example discusses outcome models for each of the 8 matches.

- Some matches produce higher quality balance than others, of course, on the variables included in the propensity score, and thus, we'd be inclined to trust them more.
- Some matches use more of the data than others, too.
- Nothing wrong with considering non-matched analyses (weighting, etc.) too, in this situation, but the example is restricted to matching.

Across the 8 matches

- Odds of "good BP" if exposed to A rather than B

Match	1	2	3	4	5	6	7	8
OR	0.56	0.42	0.57	0.65	*DNC*	*DNC*	2.00	1.07

DNC = Did not converge

- Difference (A-B) in mean BMI

Match	1	2	3	4	5	6	7	8
BMI diff	-2.3	-1.8	-1.7	-1.6	1.7	1.1	-3.2	-3.8

Agenda for Zoom Call

- Zoom for class 5 will be 2025-02-13 at 10 AM
- Feinstein's Model for Research Architecture
  - 7 Key Aspects for Making Fair Judgments about Causation
- A Published Example of My Early Propensity Score Work
  - M.I. Ahmed et al. International Journal of Cardiology 154 (2012) 128–133.
- Reviewing the Project Proposals
- Discussing Rosenbaum, Causal Inference, Chapter 6 (Quasi-Experimental Designs)