Local Control: a microaggregation methodology for performing bias-corrected reproducible observational studies across data silos while protecting patient privacy

Nicolas R. Lauve[1], Stuart J. Nelson[2], S. Stanley Young[3], Robert L. Obenchain[4], Christophe G. Lambert, PhD[5]

University of New Mexico, Albuquerque, NM, USA. [1]Dept. of Computer Science; [2]Health Sciences Library and Informatics Center;
[3]CGStat, LLC. Raleigh, NC; [4]Risk Benefit Statistics, LLC. Indianapolis, IN; [5]Center for Global Health, Dept. of Internal Medicine, Division of Translational Informatics.

Contact: {CGLambert, nlauve}@unm.edu

Abstract

We address two major challenges in the analysis of large scale observational patient studies. The first is the challenge of deriving high-quality evidence from observational data in the presence of biases and confounders, particularly with temporal data. The second is that patient privacy and other concerns prevent disclosure of individual patient-level attributes, which hinders reproducible research -- currently, there is a vast body of medical literature whose findings guide clinical practice, yet cannot be independently scrutinized. We will address these challenges through innovative extensions to the Local Control methodology:[1-5] still correcting biases while now supporting the release of question-specific microaggregates that can be shared and combined across data silos to perform reproducible treatment comparisons and comparative effectiveness studies without disclosure of individual information.

Background

- Local Control enables estimation of overall treatment effects, as well as estimation of heterogeneity of treatment effect in subpopulations.
- The key idea is to form many homogeneous patient clusters within which one can compare alternate treatments, statistically correcting for measured biases and confounders.
- After bias correction, regression or machine learning can be applied to model bias-corrected treatment differences, giving insight into what variables modify the difference in outcome from one treatment to another, unpolluted by variables that govern choice of treatment.

Cross-sectional simulation: This simulation contains a bias where treatment 1 is dosed with a higher variance than treatment 0. The adverse drug reaction (outcome variable) for both treatments is assigned using the same function: ADR = |target_dose − actual_dose| mg. Where the optimal dosage is one mg per kg of the patient's weight. We introduce the bias by modifying the variance of treatment dosages between the two groups (Figures 2, 5, 7).

The Lindner study: This data originates from a study conducted at the Ohio Heart Health Center in 1997, known as the Lindner study[6]. The study examines post-procedure expenses for patients treated with Abciximab plus usual care, compared with outcomes from patients who received usual care alone (Figures 1, 3, 4).

Framingham heart study: The Framingham study data tracks the cardiac health of more than 4000 patients over the course of twenty-four years.[7] The effects of smoking on the time to the competing risk of either reaching death, or being diagnosed with hypertension are analyzed using Local Control (Figure 6).
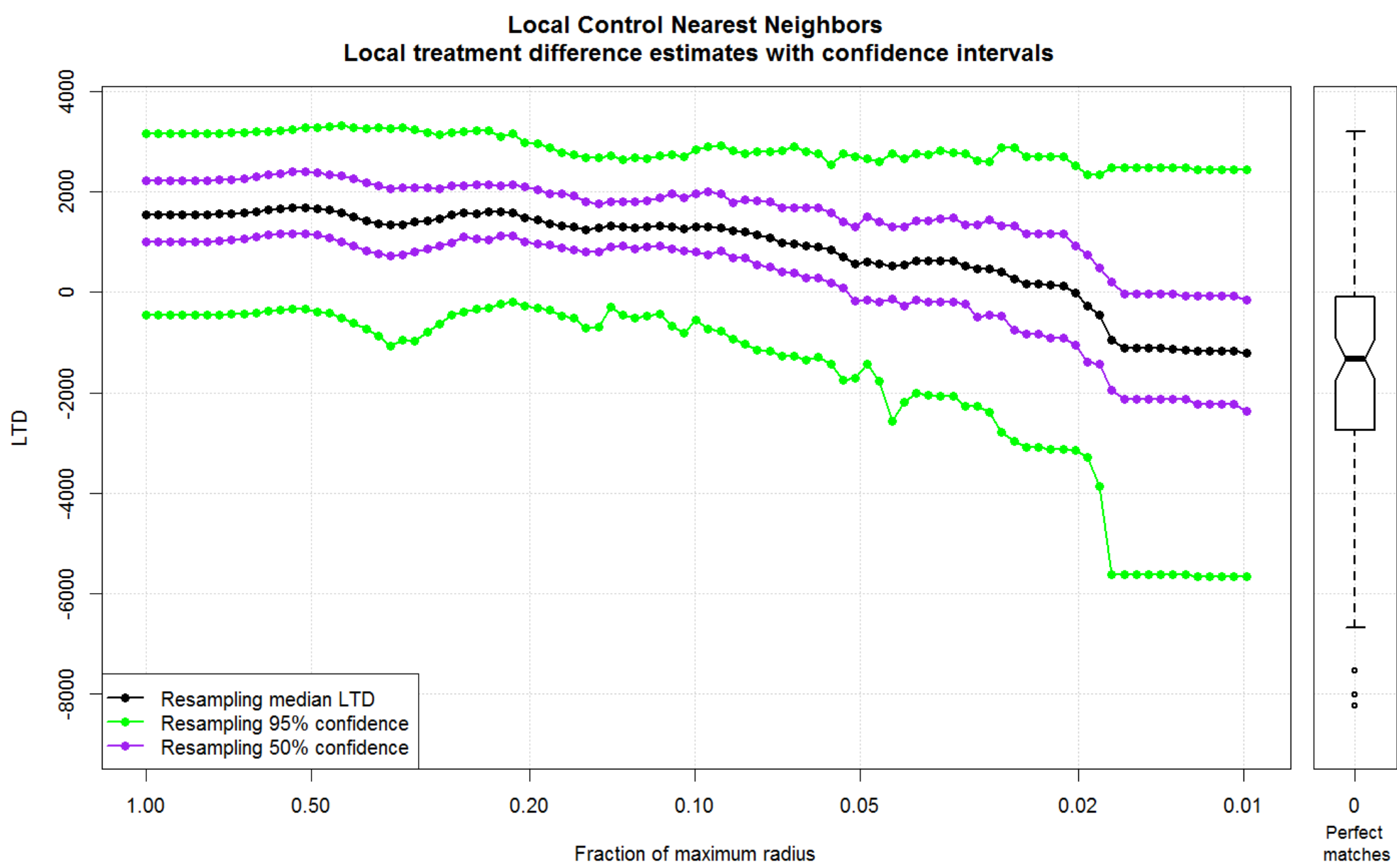

Figure 1. Lindner treatment difference as a function of correction radius. When the maximum radius fraction is 1, the treatment difference is equal to the uncorrected global average. As the fraction decreases (left to right), the treatment difference is drawn from smaller and more similar clusters. The window on the right represents only the perfect matches contained in the data.
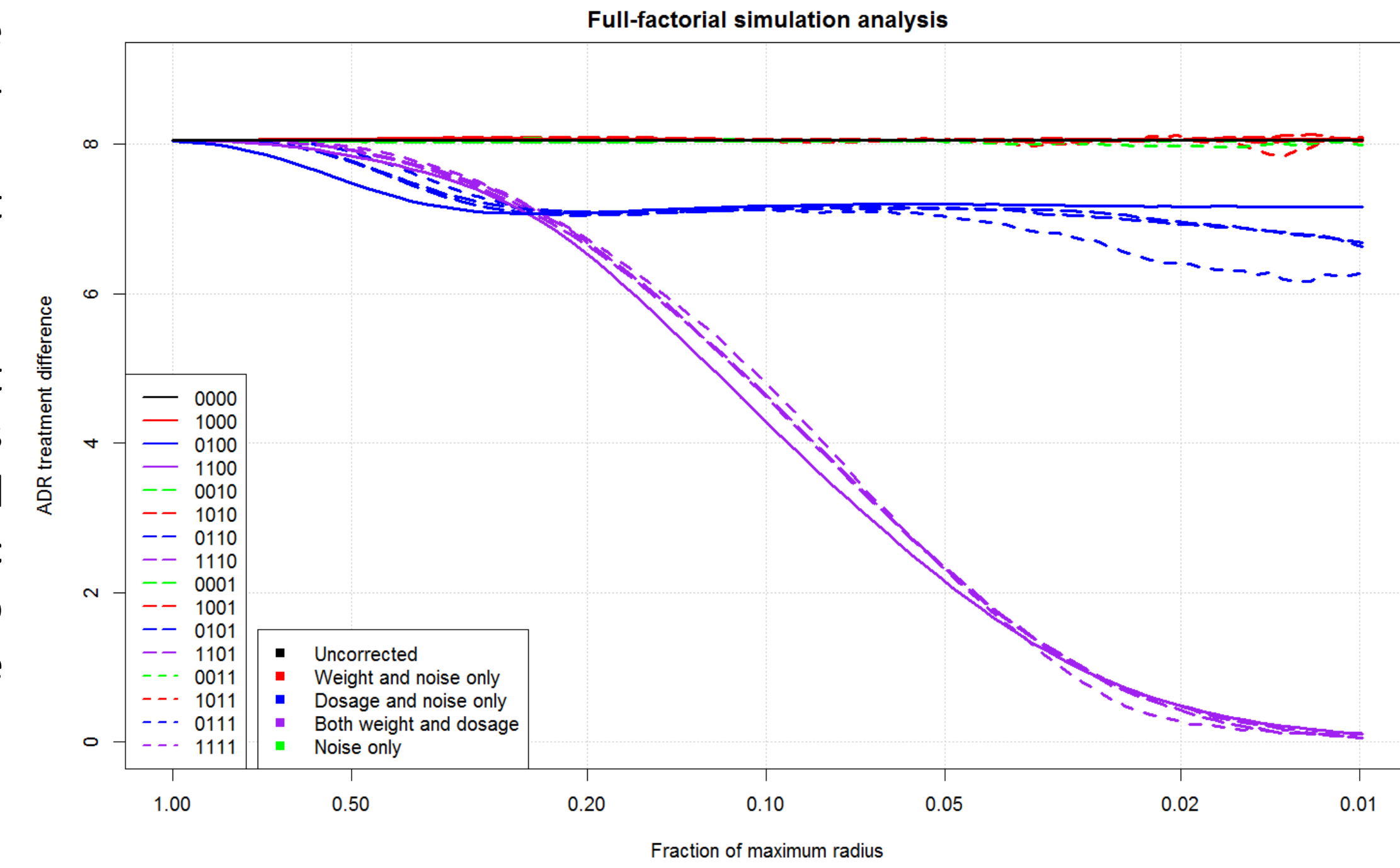
Methods

Figure 2: Full factorial simulation analysis. Local Control is run using all 16 possible combinations of the 4 covariates. Each of the curves corresponds to one of the different configurations. When both weight and dosage are included in the model (purple), the estimate of treatment difference converges to the correct answer of zero. When only one of weight or dosage is used in the model (red/blue), or neither (green), then the biases remain, and the estimate is non-zero.
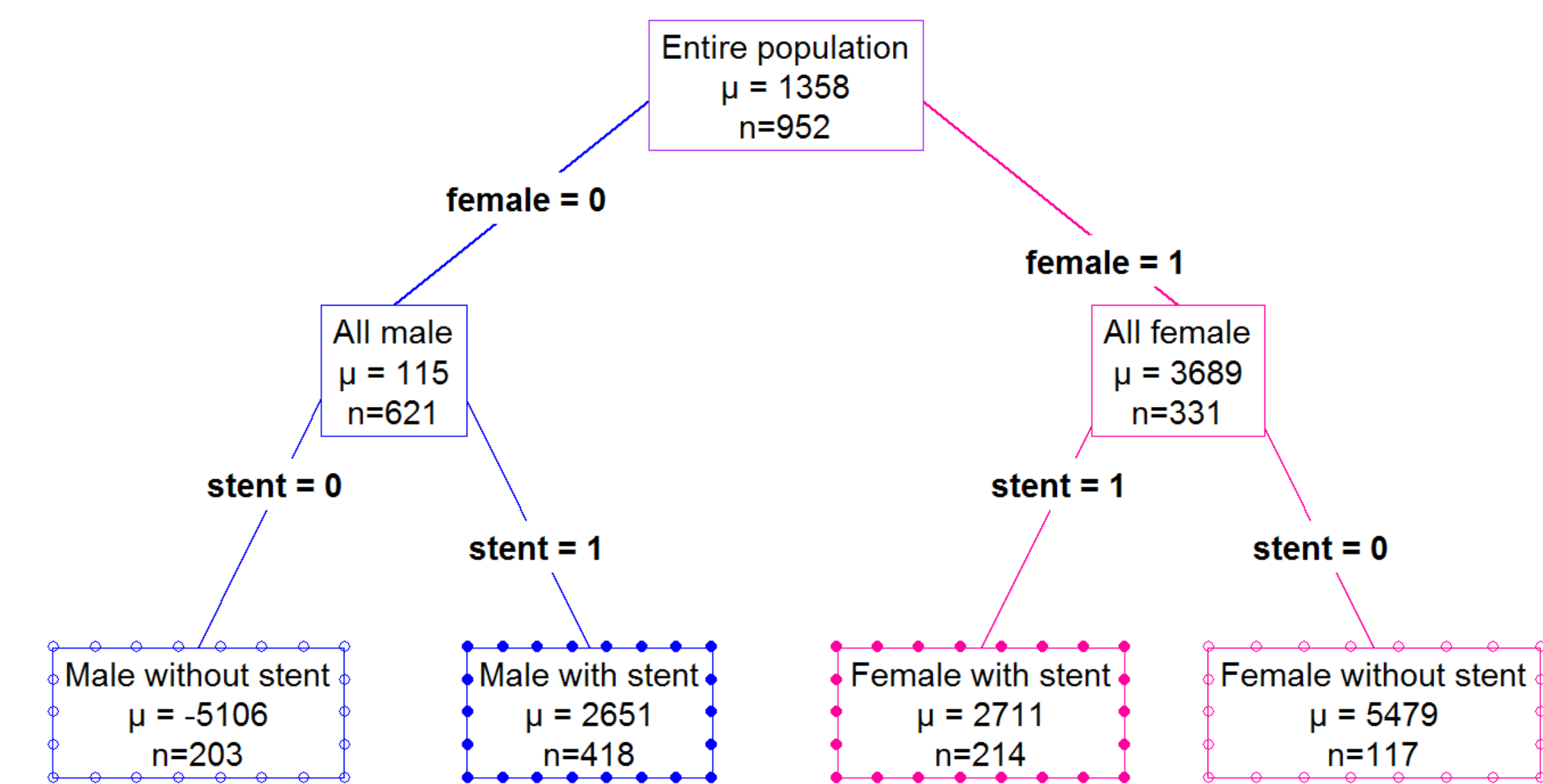

Figure 3. Recursive partitioning. After performing full factorial Local Control analysis with the seven Lindner variables, recursive partitioning is used to examine patient subgroups with statistically significant differences in bias-corrected treatment difference as a function of patient covariates, including the clustering variables.[4] This identifies four mutually exclusive subgroups: men and women with and without stents.
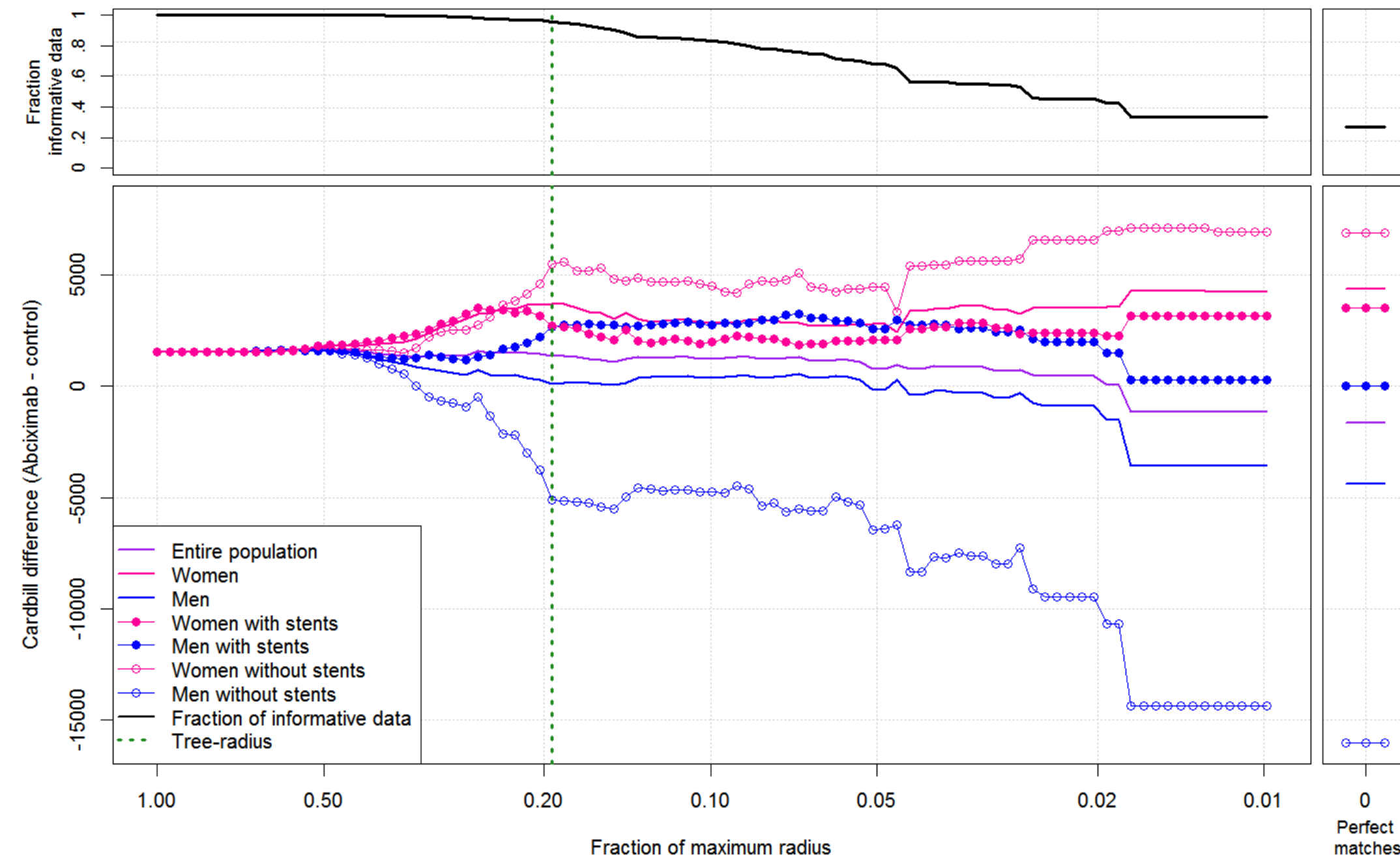

Figure 4. Lindner subgroups. After identifying significant subgroups with recursive partitioning, each of the subgroup treatment differences are graphed together. Observe that the men without stents have a much lower billing cost on Abciximab vs. control than each of the other subgroups. The top two windows display the fraction of data contributing to each of the estimates.
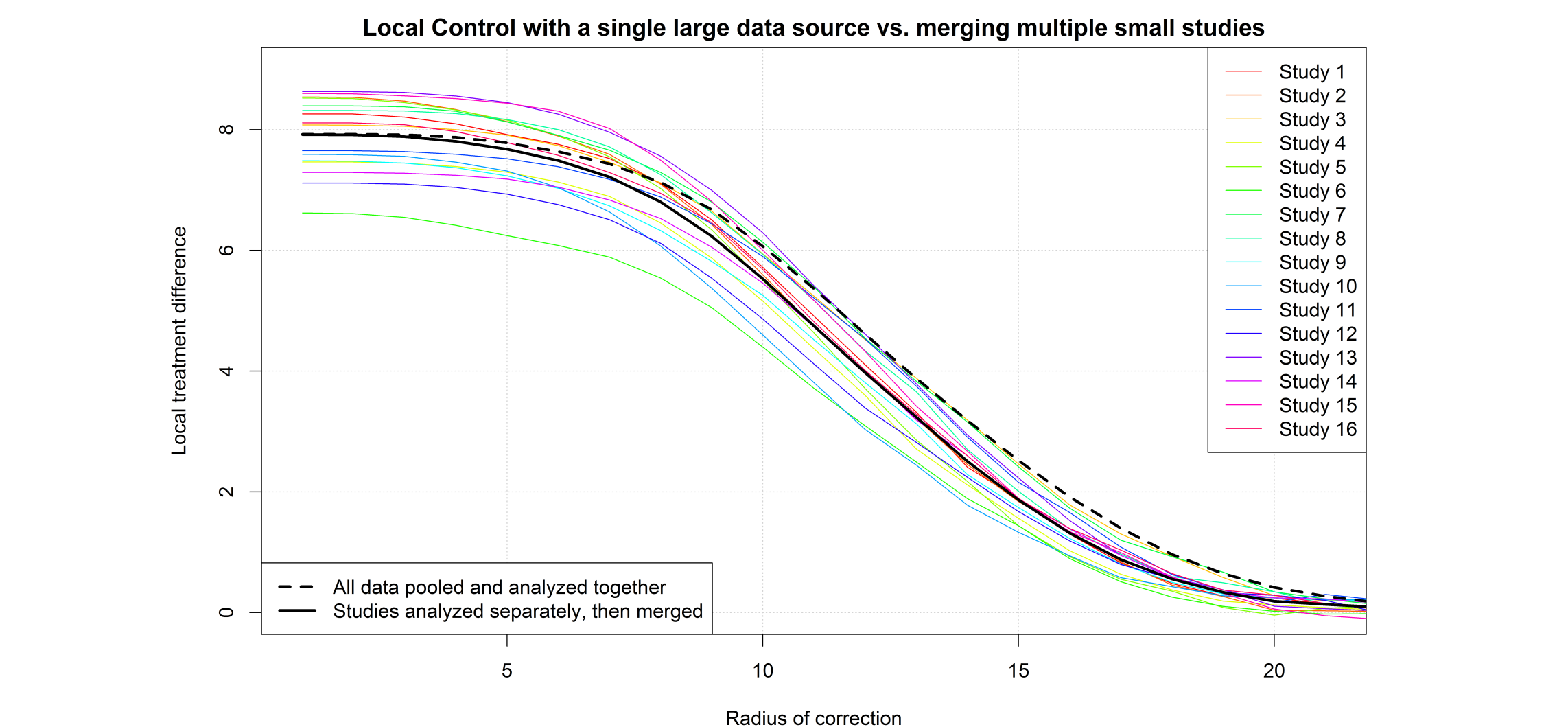
Methods

Figure 5: Microaggregation. Local Control is run independently on 16 different subsets of the original data simulating a multi-silo analysis. The results from each subset vary due to small sample sizes. Merging the microaggregates from each of the smaller studies produces a result which converges to the known answer with a larger radius than that of the original analysis.
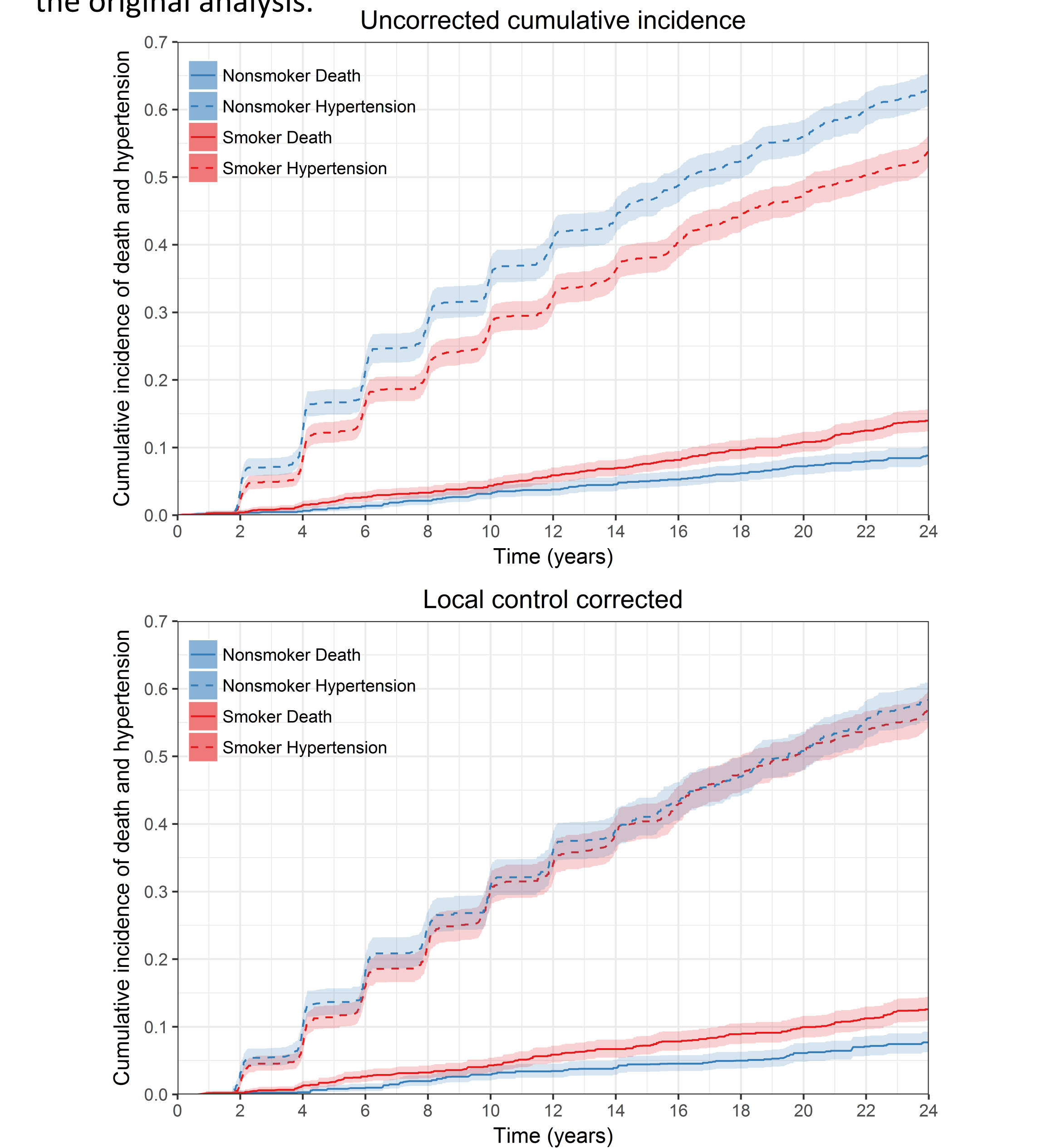

Figure 6. Competing risks of hypertension and death among smokers and nonsmokers in the Framingham study. The top plot shows the cumulative incidence without any correction for covariates. This biased estimate suggests that non-smokers have a higher risk for hypertension and lower risk of death. The bottom plot displays the results from Local Control after correcting for gender, cholesterol, age, BMI, heart rate, and blood glucose level. The bias-corrected curves show us that, among comparable patients, there is almost no difference in the rate of hypertension over time, but that the greater risk of death remains for smokers.

Results

- Fully function CRAN and OHDSI package, "LocalControl"
  - https://github.com/OHDSI/LocalControl
- Variable selection with full- or fractional-factorial sensitivity analyses
- Bias corrected survival analysis with competing risks supported
- Bias corrected subgroup analysis
- Microaggregates can be shared and combined across network studies to perform reproducible treatment comparisons
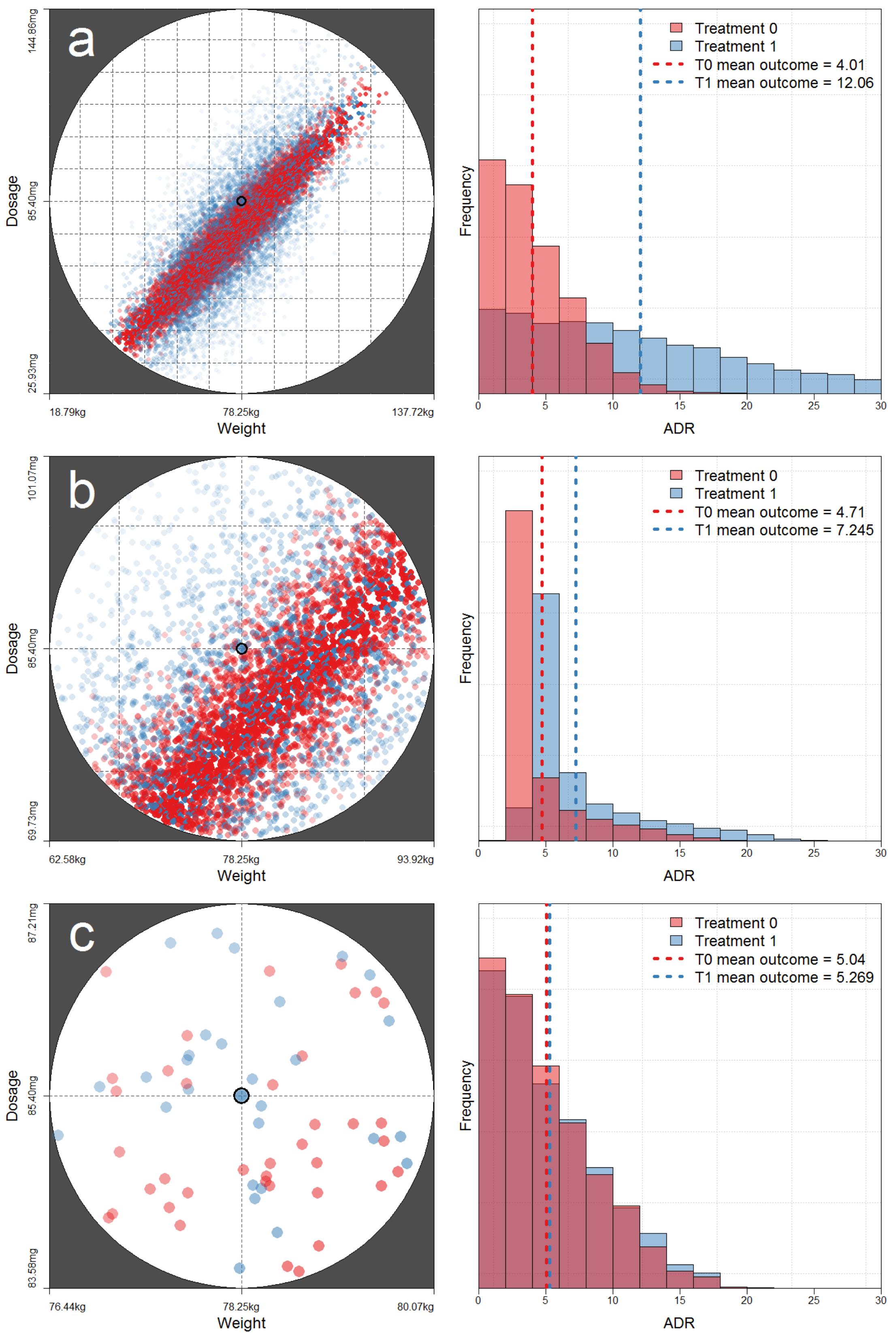

Figure 7: Local Control clustering. Observe that without correcting for bias, the blue T1 outcome average is 8.05 units higher than T0 (top histogram). As the level of correction increases, corresponding to shrinking the radius of near-neighbors (closer weight and dosage), we see that the local estimate approaches the true treatment difference of zero.

References

1) Obenchain RL (2010). "The Local Control Approach Using JMP." In D Faries, AC Leon, JM Haro, RL Obenchain (eds.), Analysis of Observational Health Care Data Using SAS, pp. 151–194. SAS Institute, Cary, NC.

2) Obenchain RL, Young SS (2013). "Advancing Statistical Thinking in Observational Health Care Research." Journal of Statistical Theory and Practice, 7(2), 456–506.

3) Lopiano KK, Obenchain RL, Young SS (2014)."Fair Treatment Comparisons in Observational Research." Statistical Analysis and Data Mining, 7(5), 376–384.

4) Faries DE, Chen Y, Lipkovich I, Zagar A, Liu X, Obenchain RL (2013). "Local Control for Identifying Subgroups of Interest in Observational Research: Persistence of Treatment for Major Depressive Disorder." International Journal of Methods in Psychiatric Research, 22(3), 185–194.

5) Nicolas R. Lauve, Stuart J. Nelson, S. Stanley Young, Robert L. Obenchain and Christophe G. Lambert (2018). LocalControl: Nonparametric Methods for Generating High Quality Comparative Effectiveness Evidence. R package version 1.2.0. https://CRAN.R-project.org/package=LocalControl

6) Kereiakes DJ, Obenchain RL, Barber BL, Smith A, McDonald M, Broderick TM, Runyon JP, Shimshak TM, Schneider JF, Hattemer CR, Roth EM, Whang DD, Cocks D, Abbottsmith CW (2000). "Abciximab Provides Cost-Effective Survival Advantage in High-Volume Interventional Practice." American Heart Journal, 140(4), 603–610.

7) Dawber TR, Meadors GF, Moore FE Jr (1951) Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health 41:279–281.