Local Control for bias correction of time-to-event observational studies

Christophe G. Lambert, PhD[1], Nicolas R. Lauve[2]

[1]Center for Global Health, Department of Internal Medicine, Division of Translational Informatics.

[2]Department of Computer Science.

University of New Mexico, Albuquerque, NM, USA.

{CGLambert, nlauve}@unm.edu

Abstract

We present to the OHDSI community the R package LocalControl, which implements novel nonparametric approaches to address biases and confounding when comparing treatments or exposures in observational studies. LocalControl is an open-source tool for researchers whose aim is to generate high quality evidence using observational health records. The package releases a family of methods for nonparametric bias correction when comparing treatments in cross-sectional, case-control, and survival analysis settings, including competing risks with censoring.

This work was supported by NIH NLM 1R21LM012389-01.

Background

Treatment comparisons can be performed in both cross-sectional and survival based settings. The problem is to find the true difference in outcomes between two treatments. A naive approach to the calculation of treatment differences in cross-sectional studies uses global averages, taking the difference in the mean outcome of each treatment. Similarly, survival-based treatment comparisons compare the cumulative risk of one or more outcomes of interest between two treatments, where time-to-event (with potential censoring) is the outcome of interest, and a naive comparison might use Kaplan-Meier curves.

In order to address biases, covariates in more complex models are typically employed. Methods include linear models and propensity scoring, the latter of which has gained wide use in correcting treatment biases[1,2], and on average, outperforms alternative methods in large scale patient records analyses[3,4]. However, a weakness of propensity scoring is that there is no guarantee about patient similarity with respect to their biasing variables, rather they only have a similar probability of treatment[5]. Thus, under the coarse matching approach of propensity scoring, if a 99-year old female had the same propensity for treatment as a 24-year old male, they might be grouped for comparison, even though this makes very little biological sense.

The Local Control method[6-9] provides a powerful and conceptually intuitive approach to statistically addressing biases and confounders in large-scale observational data. It enables estimation of overall treatment effects, as well as estimation of heterogeneity of treatment effect in subpopulations. It has been successfully used to compare treatments for major depressive disorder[8,10], and to evaluate the effect of air quality on mortality[11]. Its theoretical roots are those of propensity scoring, but it provides a tunable, finer-grained matching process for nonparametric treatment comparisons. The key idea behind Local Control is to form many homogeneous patient clusters within which one can compare alternate treatments, statistically correcting for measured biases and confounders, analogous to a randomized block design within a randomized controlled trial (RCT)[12-14]. Prior to our work, the Local Control methodology was developed only for case/control and cross-sectional studies using a hierarchical clustering approach. We introduce our new R[15] LocalControl package, which implements a nearest-neighbors clustering approach for both cross-sectional and survival/time-to-event analyses, including competing risks.

Methods

Cross-Sectional Local Control

Nearest Neighbors Local Control clusters patients for similarity on variables that are thought to be sources of bias and confounding. Each patient has a unique set of near-neighbors within a covariate hypersphere of a given radius. With the radius as a parameter, users have direct control over the degree of patient similarity within clusters. The radius can assume all real values in the range between zero, where clusters contain only perfect matches, and the maximum diameter of the covariate-space, resulting in N clusters which contain the entire population. For case/control and cross-sectional studies, Local Control calculates the global treatment difference as the average of the treatment differences across each of the neighborhoods. We illustrate this with a simulated cross-sectional dataset where two treatments have the same pharmacological properties, but a bias of a higher variance in dosage for treatment T1 versus T0 makes T1 appear to have worse outcomes. Table 1 describes more about the simulation, and Figure 1 shows a scatterplot of the outcomes as a function of variance of target dosage (as a function of weight) with actual dosage.

		T0 + T1	T0	T1	p-value
n (patients)		10000	5000	5000	--
weight (kg)	**μ**	74.76	74.72	74.80	0.804
	σ	14.97	14.99	14.94	0.800
dosage (mg)	**μ**	74.77	74.70	74.84	0.701
	σ	18.69	15.82	21.18	2.2E-16
ADR (mg)	**μ**	8.03	4.01	12.06	2.2E-16
	σ	7.86	2.99	9.07	2.2E-16

Table 1: Cross-sectional simulation cohort summary. In this simulation, we introduce 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. This table shows the distribution of weight, and dosage among the simulated patients. Using a t-test, we show that there is no statistical difference between the covariate averages in the two treatment groups. With an F-Test, we compare the variance of the two groups to show the statistical difference between the two.

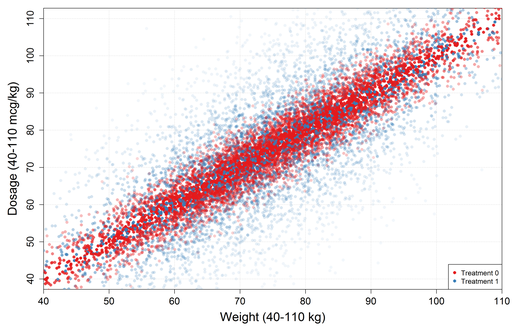

Figure 1: Distribution of covariates in the cross-sectional simulation. This plot shows the distribution of the weight, and dosage in the simulation. We use the shading of points to represent the level of adverse reaction to the treatment for each patient. The pale points indicate patients with a greater adverse reaction to the treatment, while the dark points represent those with smaller reactions.

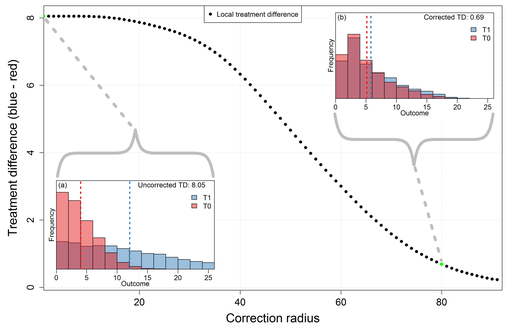

Figure 2: Treatment bias correction using Local Control on the cross-sectional simulation. We observe without correcting for bias, that the blue T1 outcome average is 8.05 units higher than T1 (lower left histogram). As the level of correction increases, corresponding to shrinking the radius of near-neighbors (closer weight and dosage), we see that the local estimate approaches the true treatment difference of zero (upper right histogram).

Survival-Based Local Control

Because one cannot simply average survival curves, the transition to longitudinal analysis required further innovation. We first devised an approach of weighting the events from each treatment in a cluster so that they sum to one, resulting in an equal contribution from all of the clusters. We then sum the series of events from all clusters, producing a global estimate which can be used with the Kaplan-Meier and competing risks counting processes. Note, that Kaplan-Meier is a special case of competing risks, the latter of which can account for multiple outcomes of interest.

We produced a simulation described in Table 2 and Figure 3, where the blue treatment naively appears to have a lower risk of adverse outcomes, but when the treatment bias (younger, lower BMI patients are more likely to receive the blue treatment) is corrected for, we see that the red treatment is actually safer. We have found in communicating results of Local Control that people struggle to intuitively grasp the notion that they are looking at a difference between treatments. With that in mind, rather than presenting survival curve differences, we created visualizations of the separate treatments before and after bias correction (Figure 3).

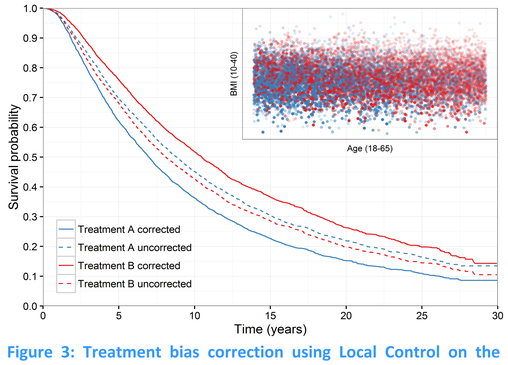

Figure 3: Treatment bias correction using Local Control on the survival simulation. Paler dots in the age vs. BMI graph indicate shorter times to a cardiac event. Because of the treatment bias, patients on A appear to have better outcomes than on B (dotted lines on Kaplan-Meier plot). However, the Local Control corrected curves (solid lines) show the true treatment effect, that treatment B actually has better outcomes than A when patients are clustered for similarity of age and BMI.

	A + B	A	B	p-value
n (patients)	1000	499	501	--
age (years)	41.17	37.95	44.39	1.31E-13
BMI (kg/m^2)	25.91	25.40	26.41	2.82E-05

Table 2: Survival simulation cohort summary. A hypothetical hypertension treatment A (blue) is prescribed more frequently to younger, healthier patients with a low body mass index (BMI), Treatment B (red) is prescribed to older patients with a higher body mass index. Significant treatment biases exist for age and BMI.

Framingham Study

To demonstrate the utility of the method, for the "treatment" of smoking vs. non-smoking we compare the risk of hypertension in the context of the competing risk of death on Framingham Heart Study Data[16], using our new competing risks Local Control Methodology.

Results

A summary of the Framingham cohort data and potential biases is shown in Table 3. We observe that smokers are more likely to be male, be younger, have a lower BMI, higher blood pressure, and have a higher resting heart rate. Figure 4 shows both uncorrected and bias-corrected comparisons of cumulative incidence of hypertension and death, illustrating how dramatically the interpretation of the data can change.

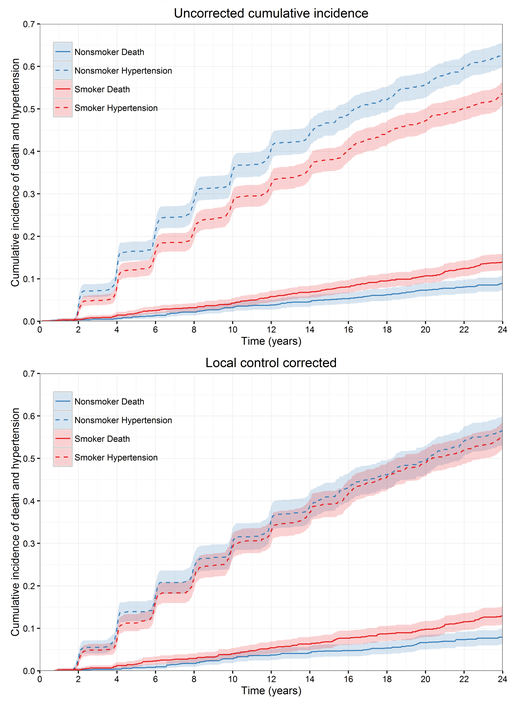

Figure 4. Framingham Heart Study: Competing risks of hypertension and death among smokers and nonsmokers. The top plot shows the cumulative incidence without any correction for covariates. This biased estimate suggests that non-smokers have a higher risk for hypertension and lower risk of death. The bottom plot displays the results from Local Control after correcting for gender, cholesterol, age, BMI, heart rate, and blood glucose level. The competing risks Local Control 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. The shaded areas represent 95% confidence interval estimates.

	All patients	Smokers	Nonsmokers	p-value
n (patients)	2291	1224	1067	--
is_male	0.44	0.52	0.35	1.27E-15
cholesterol (mg/dL)	230.37	229.13	231.8	1.29E-01
age (years)	47.42	46.1	48.93	9.46E-17
BMI (kg/m^2)	24.79	24.29	25.36	5.14E-14
blood_pressure (mm Hg)	3.07	3.32	2.78	2.45E-02
heart_rate (bpm)	74.16	74.92	73.29	5.15E-04
glucose (mg/dL)	78.53	78.14	78.99	9.43E-02

Table 3: Framingham Heart Study cohort biases. We dropped patients from the study with preexisting cardiovascular conditions. We used Fisher's exact test for the comparison of the is_male binary covariate. For the remaining continuous covariates, we used a t-test to compare the two groups. Smoking "treatment" bias was significant for sex, age, BMI, blood pressure, and heart rate.

Conclusions

With LocalControl, we have introduced a new open-source tool for the correction of bias and confounding to the OHDSI and R communities. Figure 5 shows a schema of the output from the cross-sectional function. Preliminary studies on real and simulated data with known answers have shown that LocalControl effectively corrects for measured biases in both case/control and survival/time-to-event settings.

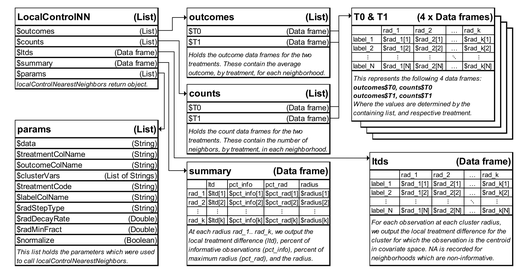

Figure 5: Output diagram for localControlNearestNeighbors. Each Local Control function returns an R list with the analysis results. Since the results differ between the three functions, each of the returned lists have S3 class names corresponding to the calling functions. In the above figure, we observe the structure of the LocalControlNN class which is returned from the localControlNearestNeighbors function. Starting from the top-left frame, we have the outermost list, the LocalControlNN object itself. The remaining rows of the frame show the names of each list element. The '$' represents one of the methods of retrieval for List elements in R, users can fetch the 'summary' data frame with LocalControlNN$summary.

References

1. Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55.
2. Rosenbaum PR, Rubin DB (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. Am Stat 39:33–38.
3. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, Welebob E, Scarnecchia T, Woodcock J (2010) Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Ann Intern Med 153:600–606.
4. Ryan PB, Madigan D, Stang PE, Overhage JM, Racoosin JA, Hartzema AG (2012) Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med 31:4401–4415.
5. Iacus SM, King G, Porro G (2012) Causal inference without balance checking: coarsened exact matching. Polit Anal 20:1–24.
6. Obenchain RL (2005) Unsupervised and supervised propensity scoring in "R." In: The Comprehensive R Archive Network. https://cran.r-project.org/src/contrib/Archive/USPS/. Accessed 8 Jun 2016.
7. Obenchain RL (2010) The local control approach using JMP. In: Faries D, Leon AC, Haro JM, Obenchain RL (eds) Analysis of observational health care data using SAS. SAS Institute, Cary, NC, pp 151–194.
8. Obenchain RL, Young SS (2013) Advancing statistical thinking in observational health care research. J Stat Theory Pract 7:456–506.
9. Lopiano KK, Obenchain RL, Young SS (2014) Fair treatment comparisons in observational research. Stat Anal Data Min 7:376–384.
10. 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. Int J Methods Psychiatr Res 22:185–194.
11. Young SS, Obenchain RL, Lambert C (2015) Bias and response heterogeneity in an air quality data set. arXiv preprint arXiv:1504. 00975
12. Gosset WS (1911) Appendix to Mercer and Hall's paper on "The experimental error of field trials". Journal of Agricultural Science 4, 128-131. Reprinted in: Pearson ES and Wishart J (eds.) (1942). Student's Collected Papers. London: University College of London, pp. 49–52.
13. Fisher RA (1926) The Arrangement of Field Experiments. Journal of the Ministry of Agriculture of Great Britain 33:503–513.
14. Addelman S (1969) The Generalized Randomized Block Design. Am Stat 23:35–36.
15. R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org.
16. 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.