[2017 OHDSI Symposium – Poster presentation]

Local Control for patient level prediction and heterogeneity of treatment effect

Nicolas R. Lauve[1*], Stuart J. Nelson[2,3], S. Stanley Young[4], Robert L. Obenchain[5], Christophe G. Lambert[3,6*]

[1]University of New Mexico (UNM) Dept. of Computer Science; [2]UNM Health Sciences Library and Informatics Center; [3]UNM Dept. of Internal Medicine, Division of Translational Informatics; [4]CGStat, LLC; [5]Risk Benefit Statistics, LLC; [6]UNM Dept. of Internal Medicine, Center for Global Health.

*Corresponding authors' emails: {nlauve, cglambert}@unm.edu

Abstract

Innovations to the R package LocalControl are presented. The package implements nonparametric approaches to address biases when comparing treatments or exposures in observational studies. This work first illustrates a factorial design of experiments approach to identifying variables for bias correction when performing treatment comparisons within the LocalControl framework. Additionally, methods are presented for estimating bias-corrected treatment outcome differences for sub-populations as well as for patient level prediction, conditioned on individual characteristics.

Introduction

The Local Control method[1-3] provides a powerful and conceptually intuitive approach to statistically addressing biases and confounders in large-scale observational data. 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. Local Control can be used alongside, or as an alternative to other methods of treatment comparisons, such as regression, or propensity scoring. The package was introduced during the 2016 Symposium and is now available on the Central R Archive Network (https://CRAN.R-project.org/package=LocalControl), and on the OHDSI github repository (https://github.com/OHDSI/LocalControl).

Feature selection

One of the open areas for research in Local Control is how to choose the relevant covariates for bias correction. One approach that is viable for a modest number of covariates is a full factorial regression analysis of how significant each covariate is in modeling the treatment difference. This work describes the full factorial approach, but note that for more variables, a fractional factorial approach could be employed for greater efficiency[4]. A full factorial design of experiments approach first runs all 2^k combinations of including or excluding each of the k covariates in the Local Control model. One can then model with linear regression the change in treatment difference estimates as a function of the binary variables (main effects and interactions) that designate which cluster variables were employed in the Local Control runs.

Patient-level prediction and heterogeneity of treatment effect

Estimates of bias-corrected treatment differences are useful in making generalizations that one treatment may be safer or more effective than another on average. However, they do not answer the question of what is the expected outcome from a given treatment for a particular patient. Patient level prediction recognizes that there may be heterogeneity of treatment effect, namely that patients can have very different outcomes depending on patient characteristics. Traditional approaches will use regression models or machine learning on the covariates to predict patient outcomes. While these approaches can provide patient level predictions, the interpretation of such models could be distorted by the biasing variables. With local treatment difference as the dependent variable, and patient covariates as the independent variables, recursive partitioning (Figure 1) can be used to classify patient subgroups[2,5-6] which are unpolluted by variables that govern choice of treatment.

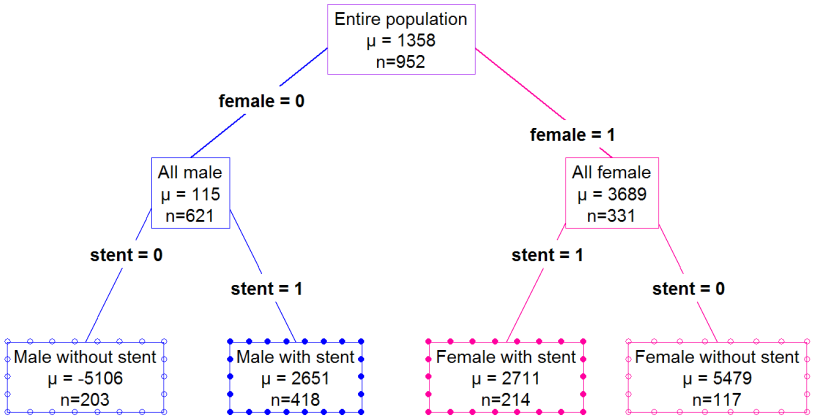

Figure 1. Abciximab recursive partitioning tree. This example uses data from a study conducted at the Ohio Heart Health Center in 1997, known as the Lindner study[7]. The study examines post-procedure effects of the treatment, Abciximab, plus usual care, compared with outcomes from patients who received usual care alone. After performing full factorial Local Control analysis with seven variables, recursive partitioning identifies four mutually exclusive subgroups with significant difference in treatment outcomes: men and women with and without stents.

Patients can be divided into the identified subgroups to examine the average local treatment differences per subgroup (Figure 2). In the Lindner example, the data suggests over a wide range of radii of bias correction (smaller radii correspond to more homogeneous patient clusters), that men without stents result in lower cost of care on Abciximab, but that all other subgroups have a lower or neutral cost of treatment on usual care alone. Machine learning approaches applied to bias-corrected treatment differences can be employed for patient level prediction conditioned on individual characteristics.

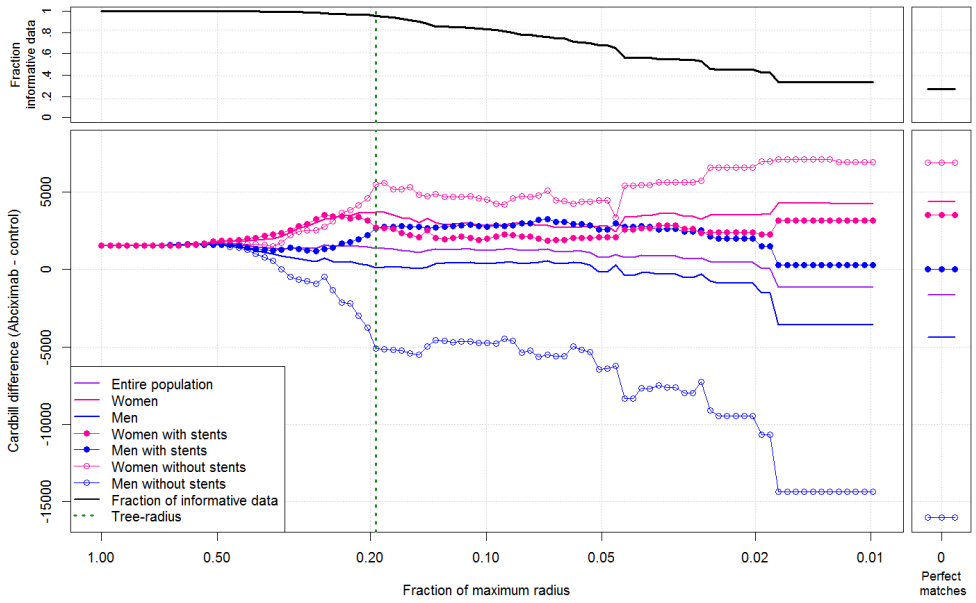

Figure 2. Lindner subgroups treatment difference as a function of correction radius. When the maximum radius fraction is 1, the treatment difference is equal to the uncorrected global average. As the fraction decreases (left to right), the treatment difference is drawn from smaller and more similar clusters. The top two windows display the fraction of data contributing to the estimate. The windows on the right represent only the perfect matches contained in the data. 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 across a wide range of correction radii.

Conclusion

In large data sets it can be true that an ”average/overall” effect is meaningless. For example a drug might work for women, but not for men – individual people do not have a fractional sex. When there is treatment response heterogeneity, a recommendation of one-size-fits-all is problematic and even a bias-corrected overall effect is misleading. LocalControl enables the analysis of both the bias-corrected average effect, as well as creates insight into subgroup outcome heterogeneity.

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. Box G, Hunter J, Hunter W (2005). Statistics for Experimenters: Design, Innovation, and Discovery. Wiley series in probability and statistics. Wiley-Interscience. ISBN 9780471718130. URL https://books.google.com/books?id=oYUpAQAAMAAJ.
5. 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.
6. Young SS, Obenchain RL, Lambert CG (2016). “A Problem of Bias and Response Heterogeneity.” In A Alan Moghissi, G Ross (eds.), Standing with giants: A collection of public health essays in memoriam to Dr. Elizabeth M. Whelan, pp. 153–169. American Council on Science and Health, New York, NY.
7. 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.