Optimal Data Analysis
Vol. 7 (July 16, 2018), 46-49

2155-0182/10/$3.00

Using ODA in the Evaluation of Randomized Controlled Trials

Ariel Linden, Dr.P.H. and Paul R. Yarnold, Ph.D.
Linden Consulting Group, LLC Optimal Data Analysis LLC

In a recent series of papers, ODA has been applied to observational data to draw causal inferences about treatment effects. In this article ODA is applied to data from a randomized controlled trial, with a reanalysis of a study by Linden and Butterworth [2014] that investigated the effect of a comprehensive hospital-based intervention in reducing readmissions for chronically ill patients. In the original analysis, negative binomial regression was used to evaluate readmission rates and emergency department visit rates at 30 and 90 days, and no treatment effects were found. However, ODA is a superior analytic approach because of its insensitivity to skewed data, model-free permutation tests to derive *P* values, identification of the threshold value which best discriminates intervention and control groups, use of a chance- and complexity-corrected indexes of classification accuracy, and cross-validation to assess generalizability of the findings.

In a comprehensive series of papers, ODA and CTA frameworks have been applied to observational data to draw causal inferences about treatment effects.[1-17] In the present paper, we turn our attention to the use of ODA for evaluating treatment effects in randomized controlled trials (RCT).

Young[18] recently used randomization tests to reanalyze results from 53 published randomized studies in which regression was originally used to evaluate effects, and found that over 10% of the originally statistically significant findings did not remain significant at the same level under randomization testing, and more than half lost their significance when subjected to additional adjustments for multiple testing. While randomization tests (also referred to as permutation tests) are only one component of ODA, these findings provide confirmation that parametric models require strong assumptions that may be easily violated. In contrast, ODA is insensitive to skewed data, it computes model-free permutation tests to derive *P* values, it identifies the threshold value which best discriminates intervention and control groups, provides a chance- and complexity-corrected index of classification accuracy, and performs cross-validation to assess generalizability of the findings.

Linden and Butterworth[19] conducted an RCT to assess whether a comprehensive hospital-based intervention could reduce readmissions for chronically ill patients versus regular care. In the original analysis, negative

2155-0182/10/$3.00

binomial regression was used to evaluate readmission rates and emergency department visit rates at 30 and 90 days and no treatment effects were found. Presently we reanalyze the results of this clinical trial using ODA.

Methods

Assessing Covariate Balance

ODA models were generated separately for each preintervention characteristic (see Table 2 in Linden and Butterworth[19]). The class variable was set as the binary treatment indicator, and preintervention characteristics were set as attributes. Non-directional (exploratory) analysis was conducted: 25,000 Monte Carlo simulations were used to compute *P* values, and leave-one-out (LOO) analysis was conducted to assess generalizability.

Evaluating Treatment Effects

ODA models were generated for each of the following four outcomes: the number of readmissions at 30 and 90 days, and the number of emergency department visits at 30 and 90 days. The class variable was set as the binary treatment indicator and each outcome was set as an attribute. This was a directional analysis (one-tailed) in which we hypothesized that outcomes would be lower in the treatment group than in the control group: 25,000 Monte Carlo simulations were used to compute *P* values, and leave-one-out (LOO) analysis was conducted to assess generalizability.

Results

Covariate Balance

The only training (total sample) analysis in this study which achieved the generalized (per-comparison) criterion for statistical significance occurred for patient activation measure (PAM) scores ($P<0.006$). The ODA model was: if PAM Score $\leq$ 46.3 then predict the observation is from the TREATMENT condition; otherwise predict the observation is from the CONTROL condition. This model correctly classified 192 of 259 (74.1%) control condition patients, and 100 of 253 (39.5%) treatment condition patients (50% accuracy is expected by chance for each condition): this level of classification accuracy corresponds to ESS = 13.7, a relatively weak effect[20,21]. This finding was stable in LOO validity analysis suggesting this result is cross-generalizable: because LOO is inherently confirmatory (one-tailed) the corresponding exact $P<0.0007$ met the criterion for experimentwise statistical significance.

Treatment Effects

ODA found no statistically significant effects for any of the four outcomes, thereby confirming the results of the original analysis using negative binomial regression.

Discussion

In this paper we have demonstrated that ODA can perform all the necessary analyses required in an RCT, namely assessing chance imbalances in baseline characteristics and evaluating treatment effects. ODA should be considered the preferred approach over commonly-used parametric models because ODA avoids the assumptions required of parametric models, is insensitive to skewed data or outliers, and has the ability to handle any variable metric including categorical, Likert-type integer, and real number measurement scales[21]. Moreover, in contrast to regression models, ODA also has the distinct ability to ascertain where the optimal (maximum-accuracy) cutpoints are on the time series, which in turn, facilitates the use of measures of predictive accuracy.

Finally, ODA has the capability to use cross-validation methods such as LOO which was employed presently, in addition to hold-out, multiple-sample, test-retest, and bootstrap validation methods[20,21] for assessing the cross-generalizability of the model to other states with

2155-0182/10/$3.00

similar characteristics planning to implement a similar set of initiatives.[22] Again, there is no equivalent in the regression-based framework, failing to provide insight into the likelihood that any observed intervention effect would cross-generalize.

References

[1]Linden A, Yarnold PR. Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice 2016;22:839-847.*

[2]Linden A, Yarnold PR. Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice 2016;22:848-854.*

[3]Linden A, Yarnold PR. Using machine learning to identify structural breaks in single-group interrupted time series designs. *Journal of Evaluation in Clinical Practice 2016;22:855-859.*

[4]Linden A, Yarnold PR, Nallomothu BK. Using machine learning to model dose-response relationships. *Journal of Evaluation in Clinical Practice 2016;22:860-867.*

[5]Yarnold PR, Linden A. Using machine learning to model dose-response relationships via ODA: eliminating response variable baseline variation by ipsative standardization. *Optimal Data Analysis* 2016(5):41-52.

[6]Linden A, Yarnold PR. Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice 2016;22:868-874.*

[7]Linden A, Yarnold PR. Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments. *Journal of Evaluation in Clinical Practice 2016;22:875-885.*

[8]Yarnold PR, Linden A. Novometric analysis with ordered class variables: The optimal alternative to linear regression analysis. *Optimal Data Analysis* 2016;22:65-73.

[9]Yarnold PR, Linden A. Theoretical aspects of the D statistic. *Optimal Data Analysis* 2016;22:171-174.

[10]Linden A, Yarnold PR. Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice 2017;23:703-712.*

[11]Yarnold PR, Linden A. Computing propensity score weights for CTA models involving perfectly predicted endpoints. *Optimal Data Analysis* 2017;6:43-46.

[12]Linden A, Yarnold PR. Modeling time-to-event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice* 2017;23:1299-1308.

[13]Linden A, Yarnold PR. Minimizing imbalances on patient characteristics between treatment groups in randomized trials using classification tree analysis. *Journal of Evaluation in Clinical Practice 2017;23:1309-1315.*

[14]Linden A, Yarnold PR. The Australian gun buy-back program and the rate of suicide by firearm. *Optimal Data Analysis* 2018;7:28-35.

[15]Linden A, Yarnold PR. Identifying causal mechanisms in health care interventions using classification tree analysis. *Journal of Evaluation in Clinical Practice 2018;24:353-361.*

2155-0182/10/$3.00

[16]Linden A, Yarnold PR. Estimating causal effects for survival (time-to-event) outcomes by combining classification tree analysis and propensity score weighting. *Journal of Evaluation in Clinical Practice 2018;24:380-387.*

[17]Linden A, Yarnold PR. Using machine learning to evaluate treatment effects in multiple-group interrupted time series analysis. *Journal of Evaluation in Clinical Practice DOI:* 10.1111/jep.12966

[18]Young A. Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. Working paper, MIT; 2015.

[19]Linden A, Butterworth SW. A comprehensive hospital-based Intervention to reduce readmissions for chronically ill patients: A randomized controlled trial. *American Journal of Managed Care* 2014;20(10):783-792.

[20]Yarnold PR, Soltysik RC. *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books, 2005.

[21]Yarnold PR, Soltysik RC. *Maximizing Predictive Accuracy*. Chicago, IL: ODA Books, 2016. DOI: 10.13140/RG.2.1.1368.3286

[22]Linden A, Adams J, Roberts N. The generalizability of disease management program results: getting from here to there. *Managed Care Interface* 2004;17:38-45.

Author Notes

No conflict of interest was reported by either author.