

Executing ODA from Within Stata: Combining Random Forests and ODA to Estimate Treatment Effects for Multi-Valued Treatments (*Invited*)

Ariel Linden, Dr.P.H.
Linden Consulting Group, LLC

This paper demonstrates how the random forest algorithm can be used in conjunction with ODA to estimate treatment effects for multivalued treatments using the new Stata package for implementing ODA.

Studies in which participants are randomized to treatment are considered the gold standard for assessing causal inference since randomization putatively ensures that the study groups do not differ systematically in their characteristics, and consequently, estimated treatment effects are assumed to be unbiased.¹ If randomization is infeasible, investigators rely on statistical techniques that model treatment assignment²⁻⁸ in order to control for threats to validity which may compromise causal interpretation of the results.⁹⁻¹³

A common approach used to estimate treatment effects in studies in which there are multiple treatment arms is to first estimate propensity scores using multinomial logistic regression, then compute the inverse probability of treatment weights (IPTW), and finally to use those weights in a weighted regression to evaluate the outcomes.⁸

This paper demonstrates how this entire process can be performed using two machine learning algorithms (random forests to compute the probabilities of treatment, and ODA to evaluate the weighted outcomes), thus avoiding the many assumptions required by conventional parametric procedures.

Both procedures are implemented in Stata. **rforest** is a community-written package that implements the random forest algorithm¹⁴ and **oda** is a new Stata package¹⁵ for implementing ODA from within the Stata environment. **oda** is a wrapper for the MegaODA software system¹⁶⁻¹⁸, so the MegaODA.exe file must be loaded on the computer for the **oda** package to work.¹⁹ To download the **oda** package, at the Stata command line type: “`ssc install oda`” (without the quotation marks).

Methods

Data

The data for our empirical example come from a disease management program designed for patients with congestive heart failure and implemented in a large health plan located in the western USA.⁶ Individuals with the condition were contacted and invited to enroll in the program. Those agreeing to participate received one of the following interventions: (1) periodic telephone calls (CALL), from a nurse to discuss self-management behaviors (n=654), or (2) remote tele-monitoring (RTM), which entailed daily electronic transmission of the participant's disease-related symptoms to a database followed by a call from the nurse if symptoms appeared to indicate the onset of an acute exacerbation (n=705). Assignment to either intervention arm was conducted by the program nurse and based largely on subjective assessment of the patient's psycho-social needs,²⁰⁻²² past levels of health care utilization, and the patient's preferred level of contact. The primary goal of the intervention was to reduce avoidable hospitalizations.^{23,24} Patients with congestive heart failure, but not participating in the program, received their usual medical care and served as controls (n=6612) in this study (see [6], and [8] for a more comprehensive description).

Analytic process

The steps involved for analyzing multivalued treatments include: (1) estimating the generalized propensity score using the random forest algorithm with the multiple-level treatment variable as the outcome. Using this approach every individual receives one propensity score corresponding to the probability of assignment to each treatment, conditional on baseline characteristics. Thus, in the current data, three propensity scores were estimated for each individual corresponding to their probability of assignment to non-participation, the CALL and RTM interventions, respectively. Next (2), a single

weight is computed based on an individual's inverse probability of receiving their actual treatment. Finally (3), treatment effects are computed using weighted ODA.

The **oda** syntax for performing the outcomes analysis is as follows:

```
oda multreat hosp, pathoda("C:\ODA\  
store("C:\ ODA\output") iter(10000)  
seed(1234) wt(iptrfwt)
```

This syntax is explained as follows: the multi-treatment variable is the *class* variable; the count of hospitalizations (per person) is the *attribute*; the directory path where the MegaODA.exe file is located on my computer is "C:\ODA"; the directory path where the output and other files generated during the analysis are stored is "C:\ODA\output"; the number of iterations (repetitions) for computing a permutation *P*-value is 10,000; the seed is set to 1234 to allow replication of permutation results (any integer value can be used as seed); and the inverse probability weight is specified as the *wt()*.

As shown in the **oda** output (Table 1), the ODA model is interpreted as follows: "if the number of hospitalizations ≤ 0.5 , then predict that the treatment group is 2 (RTM). If hospitalizations are > 0.5 and ≤ 6.5 , then predict that the treatment group is 1 (CALL), and if hospitalizations are > 6.5 , predict that the treatment group is 0 (controls)." The weighted effect strength for sensitivity (wESS) is labelled in the output as "Weighted PAC" for each of the three levels of treatment. Weighted ESS was very weak for the control group (wESS=0.23%), moderate for the CALL group (wESS=29.90%), and marginally greater than the minimum criterion for a strong effect (i.e., $\geq 75\%$) for the RTM group (wESS=75.49%).²⁵

This model was not statistically significant in training analysis ($P < 0.066$) indicating that the three treatment arms could not be discriminated on the basis of the number of hospitalizations per person. In other words, there was no statistically reliable treatment effect.

Table 1: Stata **oda** Output

```
ODA model:
-----
IF HOSP <= 0.5 THEN MULTREAT = 2
IF 0.5 < HOSP <= 6.5 THEN MULTREAT = 1
IF 6.5 < HOSP THEN MULTREAT = 0

Summary for Class MULTREAT Attribute HOSP
-----

Performance Index          Train
-----
Overall Accuracy           9.27%
Overall wtd Accuracy      12.23%
PAC MULTREAT=0            0.23%
PAC MULTREAT=1           29.97%
PAC MULTREAT=2           74.89%
Effect Strength PAC        2.54%
wtd PAC MULTREAT=0        0.23%
wtd PAC MULTREAT=1       29.90%
wtd PAC MULTREAT=2       75.49%
Effect Strength wtd PAC    2.81%
PV MULTREAT=0            88.24%
PV MULTREAT=1           12.15%
PV MULTREAT=2            8.33%
Effect Strength PV         4.36%
wtd PV MULTREAT=0        84.14%
wtd PV MULTREAT=1       15.96%
wtd PV MULTREAT=2       11.05%
Effect Strength wtd PV     5.58%
Effect Strength Total      3.45%
Effect Strength wtd Total  4.19%

Monte Carlo summary (Fisher randomization):
-----
Iterations: 10000
Estimated p: 0.066100
```

The cross-classification table for the ODA model is presented in Table 2. As seen,

the sample is highly skewed: 82.9% of the subjects were Controls, 8.2% were RTM, and 8.9% were CALL.

Discussion

This paper demonstrates how the new Stata package **oda** can be used to evaluate treatment effects in studies of multivalued treatments after the random forest algorithm estimates the generalized propensity score. This methodology should be considered the preferred approach over commonly-used parametric models since ODA avoids the assumptions required of these 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.²⁵ In contrast to conventional statistical models, ODA has the ability to ascertain where optimal (maximum-accuracy) cutpoints are on the outcome variable, which in turn, facilitates the use of measures of predictive accuracy. Moreover, ODA can perform cross-validation using LOO to assess the cross-generalizability of a statistically significant training model to potentially new study participants or non-participants.²⁵ Finally, the findings continue to support our recommendation to employ the ODA and CTA frameworks to evaluate the efficacy of health-improvement interventions and policy initiatives.²⁶⁻⁴³

Table 2: Cross-Classification Results for Training Analysis

<u>Actual Class</u>	<u>Predicted Class</u>				
	<u>Controls</u>	<u>RTM</u>	<u>CALL</u>	<u>PAC</u>	<u>wPAC</u>
Controls	15	1,241	5,356	0.23	0.23
RTM	1	196	457	29.97	29.90
CALL	1	176	528	74.89	75.49
PV	88.24	12.15	8.33		
wPV	84.14	15.96	11.05		

References

- ¹Linden A, Yarnold PR (2017). Minimizing imbalances on patient characteristics between treatment groups in randomized trials using classification tree analysis. *Journal of Evaluation in Clinical Practice*, 23, 1309-1315.
- ²Linden A, Adams J (2006). Evaluating disease management program effectiveness: an introduction to instrumental variables. *Journal of Evaluation in Clinical Practice*, 12, 148-154.
- ³Linden A, Adams JL (2010). Using propensity score-based weighting in the evaluation of health management programme effectiveness. *Journal of Evaluation in Clinical Practice*, 16, 175-179.
- ⁴Linden A, Adams JL (2010). Evaluating health management programmes over time. Application of propensity score-based weighting to longitudinal data. *Journal of Evaluation in Clinical Practice*, 16, 180-185.
- ⁵Linden A, Adams J, Roberts N (2004). Evaluating disease management program effectiveness: An introduction to survival analysis. *Disease Management*, 7, 180-190.
- ⁶Linden A (2014). Combining propensity score-based stratification and weighting to improve causal inference in the evaluation of health care interventions. *Journal of Evaluation in Clinical Practice*, 20, 1065-1071.
- ⁷Linden A, Adams J, Roberts N. Strengthening the case for disease management effectiveness: un hiding the hidden bias (2006). *Journal of Evaluation in Clinical Practice*, 12, 140-147.
- ⁸Linden A, Uysal SD, Ryan A, Adams JL (2016). Estimating causal effects for multivalued treatments: A comparison of approaches. *Statistics in Medicine*, 35, 534-552.
- ⁹Linden A, Adams J, Roberts N (2005). Evaluating disease management program effectiveness: An introduction to the bootstrap technique. *Disease Management and Health Outcomes*, 13, 159-167.
- ¹⁰Linden A (2007). Estimating the effect of regression to the mean in health management programs. *Disease Management and Health Outcomes*, 15, 7-12.
- ¹¹Linden A (2013). Assessing regression to the mean effects in health care initiatives. *BMC Medical Research Methodology*, 13, 1-7.
- ¹²Yarnold PR, Bryant FB, Soltysik RC (2013). Maximizing the accuracy of multiple regression models via UniODA: Regression away from the mean. *Optimal Data Analysis*, 2, 19-25.
- ¹³Linden A, Roberts N (2005). A Users guide to the disease management literature: Recommendations for reporting and assessing program outcomes. *American Journal of Managed Care*, 11, 81-90.
- ¹⁴Schonlau, M. and Zou, R.Y., 2020. The random forest algorithm for statistical learning. *The Stata Journal*, 20(1), pp.3-29.
- ¹⁵Linden A (2020). ODA: Stata module for conducting Optimal Discriminant Analysis. *Statistical Software Components S458728*, Boston College Department of Economics.
- ¹⁶Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Separating the chaff. *Optimal Data Analysis*, 2, 194-197.
- ¹⁷Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Harvesting the Wheat. *Optimal Data Analysis*, 2, 202-205.

- ¹⁸Yarnold PR, Soltysik RC (2013). MegaODA large sample and BIG DATA time trials: Maximum velocity analysis. *Optimal Data Analysis*, 2, 220-221.
- ¹⁹Rhodes NJ, Yarnold PR. 2020. ODA: a package and R-interface for the MegaODA software suite. R package version 1.0.1.3. Available: <https://github.com/njrhodes/ODA>
- ¹⁹Linden A, Roberts N (2004). Disease management interventions: What's in the black box? *Disease Management*, 7, 275-291.
- ²⁰Linden A, Butterworth S, Roberts N (2006). Disease management interventions II: What else is in the black box? *Disease Management*, 9, 73-85.
- ²¹Biuso TJ, Butterworth S, Linden A (2007). A conceptual framework for targeting prediabetes with lifestyle, clinical and behavioral management interventions. *Disease Management*, 10, 6-15.
- ²²Linden A, Adler-Milstein J (2008). Medicare disease management in policy context. *Health Care Finance Review*, 29, 1-11.
- ²³Linden A, Adams J, Roberts N (October, 2003). *Evaluation methods in disease management: determining program effectiveness*. Position Paper for the Disease Management Association of America (DMAA).
- ²⁴Yarnold PR, Soltysik RC. *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books, 2005.
- ²⁵Yarnold PR, Soltysik RC (2016). *Maximizing Predictive Accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286
- ²⁶Linden A, Adams J, Roberts N (2004). The generalizability of disease management program results: getting from here to there. *Managed Care Interface*, 17, 38-45.
- ²⁷Linden A, Yarnold PR (2018). Identifying causal mechanisms in health care interventions using classification tree analysis. *Journal of Evaluation in Clinical Practice*, 24, 353-361.
- ²⁸Linden A, Yarnold PR, Nallomothu BK (2016). Using machine learning to model dose-response relationships. *Journal of Evaluation in Clinical Practice*, 22, 860-867.
- ²⁹Linden A, Yarnold PR (2018). Estimating causal effects for survival (time-to-event) outcomes by combining classification tree analysis and propensity score weighting. *Journal of Evaluation in Clinical Practice*, 24, 380-387.
- ³⁰Linden A, Yarnold PR (2016). Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice*, 22, 848-854.
- ³¹Linden A, Yarnold PR (2016). Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice*, 22, 839-847.
- ³²Linden A, Yarnold PR (2017). Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice*, 23, 703-712.
- ³³Linden A, Yarnold PR (2016). Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments. *Journal of Evaluation in Clinical Practice*, 22, 875-885.
- ³⁴Linden A, Yarnold PR (2016). Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, 22, 868-874.
- ³⁵Linden A, Yarnold PR (2017). Modeling time-to-event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice*, 23, 1299-1308.

³⁶Linden A, Yarnold PR (2018). Using machine learning to evaluate treatment effects in multiple-group interrupted time series analysis. *Journal of Evaluation in Clinical Practice*, 24, 740-744.

³⁷Yarnold PR, Linden A (2016). Theoretical aspects of the D statistic. *Optimal Data Analysis*, 22, 171-174.

³⁸Yarnold PR, Linden A (2017). Computing propensity score weights for CTA models involving perfectly predicted endpoints. *Optimal Data Analysis*, 6, 43-46.

³⁹Yarnold PR, Linden A (2016). Using machine learning to model dose-response relationships via ODA: Eliminating response variable baseline variation by ipsative standardization. *Optimal Data Analysis*, 5, 41-52.

⁴⁰Linden A, Yarnold PR (2018). The Australian gun buy-back program and the rate of suicide by firearm. *Optimal Data Analysis*, 7, 28-35.

⁴¹Linden A, Yarnold PR (2018). Using ODA in the evaluation of randomized controlled trials. *Optimal Data Analysis*, 7, 46-49.

⁴²Linden A, Yarnold PR (2018). Using ODA in the evaluation of randomized controlled trials: Application to survival outcomes. *Optimal Data Analysis*, 7, 50-53.

⁴³Linden A, Yarnold PR. (2019). Some machine learning algorithms find relationships between variables when none exist -- CTA doesn't. *Optimal Data Analysis*, 8, 64-67.

Author Notes

No conflict of interest was reported.