

Implementing ODA from Within Stata: An Application to Estimating Treatment Effects using Observational Data (*Invited*)

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

In this paper, I demonstrate how treatment effects in observational data can be estimated for both binary and multivalued treatments using the new Stata package for implementing ODA. Matching and weighting techniques are implemented and ODA results are compared to those using conventional regression approaches.

A prior paper¹ introduced the new Stata package called `oda`² for implementing ODA from within the Stata environment. This Stata package is a wrapper for the MegaODA software³, therefore the MegaODA.exe file must be loaded on the computer for the `oda` package to work (MegaODA software is available at: <https://odajournal.com/resources/>). To download the `oda` package, at the Stata command line type: “`ssc install oda`” (without the quotation marks).

This paper demonstrates how the `oda` package can be used to estimate treatment effects in observational data. For binary treatments, we use two common approaches for generating comparable samples based on the propensity score—matching and weighting. For multivalued treatments, we use the generalized propensity score to generate weights. Upon completing this preliminary step (matching or weighting), we then estimate treatment effects in `oda` by specifying the treatment as the *class*

variable and the outcome as the *attribute*. For matching, we simply limit the analysis to the matched sample, whereas weighted analyses are performed by specifying the treatment as the *class* variable, the outcome as the *attribute*, and the weight using the `wt()` option.

Methods

Binary treatments – Matching

The first example uses data from a prior evaluation of a primary care-based medical home pilot program⁴ that invited patients to enroll if they had a chronic illness or were predicted to have high costs in the following year. The goal of the program was to lower health care costs for program participants by providing intensified primary care (see Linden [2011]⁴ for a more comprehensive description).

For the purpose of this empirical example, a one-to-one, propensity score-based matching approach was used. This entailed first

estimating the propensity score via the conventional approach of using logistic regression to predict program participation status using the 11 pre-intervention covariates, all entered as main effects, followed by implementation of an optimal matching algorithm⁵ to match pairs (one participant to one nonparticipating control) on the estimated propensity score, resulting in a total of 276 matched pairs.⁴

The outcome evaluated in this study was the difference in costs (post-treatment year minus pre-treatment year) between the treatment group and control group (this is commonly referred to as a “difference-in-differences”). Ordinary least squares (OLS) regression was used to estimate the treatment effect, and the results indicated a non-statistically significant difference between groups ($P=0.113$; 95% CI: -4321, 457).

We now evaluate these data using **oda** with the following syntax (see the help file for **oda** for a complete description of the syntax options):

```
oda treat diffcost, pathoda("C:\ODA\  

store("C:\ODA\output") iter(25000) loo  

seed(1234)
```

The above syntax is explained as follows: The variable “treat” is the *class* variable; the outcome variable “diffcost” 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 will be stored is “C:\ODA\output”; the number of iterations (repetitions) for computing a permutation P -value is 25,000; leave-one-out (LOO) analysis should be performed, and the seed is set to 1234 to ensure replication of the permutation results (any numeric value can be used for the seed). It should also be noted that the sample was first limited to only the matched pairs before executing **oda**. However, we could have also simply used the [if][in] qualifier to limit the analysis to those matches (see the help file for **oda**).

The **oda** package produces an extract of the total output produced by the ODA software (the complete output is stored in the specified directory with the extension “.out”).

```
ODA model:
-----
IF DIFFCOST <= 420.0 THEN TREAT = 0
IF 420.0 < DIFFCOST THEN TREAT = 1

Summary for Class TREAT Attribute DIFFCOST
-----
Performance Index          Train    LOO
-----
Overall Accuracy           54.17%  53.26%
PAC TREAT=0                69.20%  69.20%
PAC TREAT=1                39.13%  37.32%
Effect Strength PAC        8.33%   6.52%
PV TREAT=0                 53.20%  52.47%
PV TREAT=1                 55.96%  54.79%
Effect Strength PV         9.16%   7.26%
Effect Strength Total      8.75%   6.89%

Monte Carlo summary (Fisher randomization):
-----
Iterations: 25000
Estimated p: 0.289000

Results of leave-one-out analysis
-----
552 observations

Fisher's exact test (directional) classification table p = .063347
```

As seen in the **oda** output, the ODA model can be interpreted as follows: “if the difference in costs are \leq \$430, then predict that the treatment group is 0 (controls). If the difference in costs are $>$ \$430, then predict that the treatment group is 1 (treatment).”

The effect strength for sensitivity (ESS) is labelled in the output as “Effect Strength PAC” (Percentage Accuracy in Classification). In the training data the ESS is 8.33% and in the LOO analysis it is 6.52% (both are very weak effects).⁶ The permutation P -value for the training sample was 0.289 and for the LOO analysis 0.063. In summary, ODA could not find a model that sufficiently discriminated between treatment groups to elicit a statistically significant treatment effect in these data.

Binary treatments – Weighting

The second example of a binary treatment uses data from an evaluation of a disease management program designed for patients with

congestive heart failure.⁷ The primary goal of the intervention was to reduce avoidable hospitalizations. Here, all 1359 program participants are compared to all 6612 non-participants on the primary outcome—the difference in post-treatment minus pre-treatment hospitalization rates between groups (i.e., difference-in-differences).

The analytic process used in this empirical example involved; (1) estimating the propensity score using logistic regression to predict program participation status using pre-intervention covariates; (2) dividing the entire sample into 5 strata of the propensity score; (3) generating weights for each individual based on their study group and strata; and (4) estimating treatment effects using weighted OLS (i.e., the weight was specified as a weight in the OLS model). The results indicated a statistically significant difference in favor of the control group ($P=0.031$; 95% CI: 0.008, 0.174). In other words, the control group had a larger decrease in hospitalizations than did the treatment group.

The **oda** syntax for analyzing these data is as follows:

```
oda treat diffhosp, pathoda("C:\ODA\")
store("C:\ODA\output") iter(25000)
seed(1234) loo wt(weight)
```

As can be seen, the syntax is nearly identical to that used in the prior example, with the only difference being the addition of the **wt()** option which is how the user specifies the propensity score weight. The **oda** output for this model is as follows:

```
ODA model:
-----
IF DIFFHOSP <= 0.5 THEN TREAT = 0
IF 0.5 < DIFFHOSP THEN TREAT = 1

Summary for Class TREAT Attribute DIFFHOSP
-----
```

Performance Index	Train	LOO
Overall Accuracy	76.58%	76.58%
Overall wtd Accuracy	75.97%	75.97%
PAC TREAT=0	89.05%	89.05%
PAC TREAT=1	15.89%	15.89%
Effect Strength PAC	4.94%	4.94%
wtd PAC TREAT=0	88.61%	88.61%
wtd PAC TREAT=1	14.49%	14.49%
Effect Strength wtd PAC	3.10%	3.10%
PV TREAT=0	83.74%	83.74%
PV TREAT=1	22.98%	22.98%
Effect Strength PV	6.72%	6.72%
wtd PV TREAT=0	83.44%	83.44%
wtd PV TREAT=1	20.74%	20.74%
Effect Strength wtd PV	4.18%	4.18%
Effect Strength Total	5.83%	5.83%
Effect Strength wtd Total	3.64%	3.64%

```
Monte Carlo summary (Fisher randomization):
-----
Iterations: 25000
Estimated p: 0.045560
```

```
Results of leave-one-out analysis
-----
7971 observations

Fisher's exact test (directional) classification table p = .402E-0006
```

As shown in the **oda** output, the ODA model can be interpreted as follows: “if the difference in hospitalizations ≤ 0.5 , then predict that the treatment group is 0 (controls). If the difference in hospitalizations are > 0.5 , then predict that the treatment group is 1 (treatment).”

In both the training and LOO data, the weighed ESS (listed as “Effect Strength Wtd PAC”) is 3.10% indicating a very weak effect.⁶ The permutation P -value for the training sample was 0.046 and for the LOO analysis <0.0001 . In summary, ODA found a model that poorly discriminated between treatment groups, and this cut-point elicited a statistically significant effect in favor of the control group. These results largely mirror those estimated using weighted regression.⁷

Multivalued treatments – Weighting

This example demonstrates how to use **oda** for estimating treatment effects in an observational study with multiple treatments.

The data used here are the same as those used in the prior example (binary treatment with weighting) but the treatments are as follows: Those agreeing to participate received one of the following two interventions based on the subjective assignment by a program nurse: (1) periodic telephone calls from a nurse to discuss self-management behaviors (n=654); or (2) remote tele-monitoring (RTM) that 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). The control group received usual medical care only (n=6612). These data were originally analyzed using several different methodological approaches, with consistent results between models.⁸ The control group had a statistically lower hospitalization rate than the call group but was not statistically different than the RTM group. The RTM group had a statistically lower rate than the call group.

The steps involved for analyzing multivalued treatments, include: (1) estimating the generalized propensity score GPS using multinomial logistic regression with the multiple-level treatment variable as the outcome. Using this approach, each 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 telephonic intervention and RTM, respectively. (2) A single weight is computed based on an individual’s probability of receiving their actual treatment (there are several different weighting techniques used in practice, but nearly all are variations of this basic method). Finally, (3) treatment effects are estimated using weighted regression, and Bonferroni-adjusted pairwise contrasts are performed to determine where effects are found between groups.

The **oda** syntax for analyzing these data is as follows:

```
oda multreat hosp, pathoda("C:\ODA\")
store("C:\ODA\output") iter(25000)
seed(1234) loo wt(multwt)
```

As can be seen, the general syntax is identical to that used in the binary weighting example (however the treatment variable and weight here is for the multivalued treatments). The **oda** output for this model is as follows:

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

Summary for Class MULTREAT Attribute HOSP
-----
```

Performance Index	Train	LOO
Overall Accuracy	69.67%	69.58%
Overall wtd Accuracy	68.86%	68.78%
PAC MULTREAT=0	81.00%	81.00%
PAC MULTREAT=1	29.97%	28.90%
PAC MULTREAT=2	0.14%	0.14%
Effect Strength PAC	5.56%	5.02%
wtd PAC MULTREAT=0	80.19%	80.19%
wtd PAC MULTREAT=1	28.45%	27.43%
wtd PAC MULTREAT=2	0.12%	0.12%
Effect Strength wtd PAC	4.38%	3.87%
PV MULTREAT=0	84.47%	84.47%
PV MULTREAT=1	12.15%	11.77%
PV MULTREAT=2	5.88%	4.17%
Effect Strength PV	1.25%	0.20%
wtd PV MULTREAT=0	83.75%	83.75%
wtd PV MULTREAT=1	11.47%	11.11%
wtd PV MULTREAT=2	4.68%	3.42%
Effect Strength wtd PV	-0.05%	-0.86%
Effect Strength Total	3.40%	2.61%
Effect Strength wtd Total	2.16%	1.50%

```

Monte Carlo summary (Fisher randomization):
-----
Iterations: 25000
Estimated p: 0.001480

Results of leave-one-out analysis
-----
7971 observations
(P-values are computed for binary class variables only)
```

As seen in the **oda** output, the ODA model can be interpreted as follows: “if hospitalizations ≤ 0.5 , then predict that the treatment group is 0 (controls). If hospitalizations are between 0.5 and 6.5, then then predict the

treatment group is 1 (calls). And, if hospitalizations > 6.5, then predict that the treatment group is 2 (RTM).”

In both the training and LOO data, the weighted ESS is under 4% indicating a very weak effect.⁶ The permutation P -value for the training sample was 0.001. No P -value can be computed in LOO analyses with > 2 treatment (*class* variable) levels.

Given that a statistically significant effect has been computed, we can further investigate between which groups the effect is found. We can do this by reissuing the `oda` syntax for each pair of the treatments, for example:

```
oda multreat hosp if inlist(multreat, 0,1),  
pathoda("C:\ODA\") store("C:\ODA\output")  
iter(25000) seed(1234) loo wt(multwt)
```

Here, “if inlist(multreat, 0,1)” indicates that the analysis should be limited to treatments 0 and 1. The results of this pairwise analysis indicate no statistical difference between controls and RTM ($P= 0.955$), a statistical difference between controls and calls ($P < 0.0001$), and a statistical difference between RTM and calls ($P= 0.004$). We did not adjust for multiple comparisons, but doing so will not alter the findings presently.

Discussion

This paper demonstrated how the new Stata package `oda` can be used to evaluate treatment effects in observational data with binary or multivalued treatments. 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.³ Moreover, in contrast to regression models, ODA also has the distinct ability to

ascertain where optimal (maximum-accuracy) cutpoints lie on the outcome variable, which in turn facilitates the use of measures of predictive accuracy. Moreover, ODA can perform cross-validation using LOO which enables assessing cross-generalizability of the model to potentially new study participants or non-participants.

Finally, the findings continue to support the recommendation to employ the ODA and CTA frameworks to draw causal inferences regarding treatment effects in observational data, and in data from randomized controlled trials.⁹⁻²⁵ A large, ever-increasing body of evidence supports the use of ODA and CTA to evaluate the efficacy of health-improvement interventions and policy initiatives.²⁶⁻²⁹

References

- ¹Linden A (2020). Implementing ODA from Within Stata: An Application to Data From a Randomized Controlled Trial (*Invited*). *Optimal Data Analysis*, 9, 9-13.
- ²Linden A (2020). ODA: Stata module for conducting Optimal Discriminant Analysis. *Statistical Software Components S458728, Boston College Department of Economics*.
- ³Yarnold PR, Soltysik RC (2016). *Maximizing Predictive Accuracy*. Chicago, IL: ODA Books. DOI: 10.13140/RG.2.1.1368.3286
- ⁴Linden A (2011). Identifying spin in health management evaluations. *Journal of Evaluation in Clinical Practice*, 17, 1223-1230.
- ⁵Rosenbaum PR (1989). Optimal matching for observational studies. *Journal of the American Statistical Association*, 84, 1024–1302.
- ⁶Yarnold PR, Soltysik RC (2005). *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books.

⁷Linden 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.

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, 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.

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

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

¹²Linden 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.

¹³Linden A, Yarnold PR (2016). Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice*, 22, 839-847.

¹⁴Linden A, Yarnold PR (2016). Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice*, 22, 848-854.

¹⁵Linden A, Yarnold PR (2016). Using machine learning to identify structural breaks in single-group interrupted time series designs. *Journal of Evaluation in Clinical Practice*, 22, 855-859.

¹⁶Linden A, Yarnold PR, Nallomothu BK (2016). Using machine learning to model dose-response relationships. *Journal of Evaluation in Clinical Practice*, 22, 860-867.

¹⁷Yarnold 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.

¹⁸Linden 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.

¹⁹Linden 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.

²⁰Linden A, Yarnold PR (2017). Modeling time-to-event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice*, 23, 1299-1308.

²¹Linden A, Yarnold PR (2018). The Australian gun buy-back program and the rate of suicide by firearm. *Optimal Data Analysis*, 7, 28-35.

²²Linden A, Yarnold PR (2018). Identifying causal mechanisms in health care interventions using classification tree analysis. *Journal of Evaluation in Clinical Practice*, 24, 353-361.

²³Linden 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.

²⁴Linden A, Yarnold PR (2018). Using ODA in the evaluation of randomized controlled trials. *Optimal Data Analysis*, 7, 46-49.

²⁵Linden A, Yarnold PR (2018). Using ODA in the evaluation of randomized controlled trials: Application to survival outcomes. *Optimal Data Analysis*, 7, 50-53.

²⁶Linden 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).

²⁷Linden 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.

²⁸Biuso TJ, Butterworth S, Linden A (2007). A conceptual framework for targeting prediabetes with lifestyle, clinical and behavioral management interventions. *Disease Management*, 10, 6-15.

²⁹Linden A, Adler-Milstein J (2008). Medicare disease management in policy context. *Health Care Finance Review*, 29, 1-11.

Author Notes

No conflict of interest was reported.