Optimal Data Analysis
Vol. 9 (February 26, 2020), 33-38

2155-0182/10/$3.00

Implementing ODA from Within Stata: Assessing Covariate Balance in Observational Studies (*Invited*)

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

In this paper, I describe how to assess whether treatment groups are comparable on observed baseline covariates (balance) in non-randomized studies using the new Stata package for implementing ODA.

Prior papers[1,2,3] introduced the new Stata package called **oda**[4] for implementing ODA from within the Stata environment. This package is a wrapper for the MegaODA software[5], and 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).

In this paper, I demonstrate how the **oda** package can be used to assess whether treatment groups are comparable (balanced) on observed baseline characteristics (covariates) in observational studies.

Studies in which participants are randomized to treatment are considered the gold standard for assessing causal inference because randomization ensures the study groups do not differ systematically in their characteristics, and consequently, treatment effects are assumed to be unbiased. When randomization is not feasible, investigators rely on statistical techniques that model the treatment assignment to control for threats to validity that may compromise causal interpretation of the results.[6-10]

In **oda** we assess covariate balance by simply specifying the treatment indicator as the *class* variable and each baseline covariate as the *attribute*. If matching was used to create comparable pairs, then the user needs only to limit the data in the analysis to those matched pairs. If weighting techniques were employed to generate comparable groups, then the user simply specifies the weight using the *wt*() option. If the study groups are indeed comparable, then we'd expect that ODA will produce a very low ESS value and non-statistically significant permutation *P*-values.

Methods

Data

This paper uses data from a prior evaluation of a primary care-based medical home pilot program[11] 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

2155-0182/10/$3.00

participants by providing intensified primary care (see Linden [2011][11] 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 11 pre-intervention covariates, all entered as main effects, followed by implementation of an optimal matching algorithm[12] to match pairs (one participant to one nonparticipating control) on the estimated propensity score, resulting in 276 matched pairs.[11] While a comprehensive discussion on the use of ODA for assessing covariate balance is available elsewhere, [13] here the discussion is limited to its implementation using the **oda** package.

Analytic process

Using a baseline covariate as an attribute, **oda** is implemented with the following (see the help file for **oda** for a complete description of the syntax options):

```
oda treat age, pathoda("C:\ODA\")
store("C:\ODA\output") iter(10000)
seed(1234)
```

The above syntax is explained as follows for assessing covariate balance of the variable "age": The variable "treat" is the *class* variable; "age" 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 should be stored is "C:\ODA\output"; the number of iterations (repetitions) for computing a permutation *P*-value is 10,000; and we set the seed to 1234 to allow us to replicate of the permutation results (any integer value can be used in the seed).

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 AGE <= 52.5 THEN TREAT = 1
IF 52.5 < AGE THEN TREAT = 0


Summary for Class TREAT  Attribute AGE
---------------------------------------

Performance Index           Train
-----------------          ------
Overall Accuracy            53.99%
PAC TREAT=0                 70.29%
PAC TREAT=1                 37.68%
Effect Strength PAC          7.97%
PV TREAT=0                  53.01%
PV TREAT=1                  55.91%
Effect Strength PV           8.92%
Effect Strength Total        8.45%


Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  10000
Estimated p: 0.214900
```

As shown in the **oda** output, the ODA model can be interpreted as follows: "if age <= 52.5, then predict that treatment group is 1 (treatment). If age is > 52.5, then predict that the treatment group is 0 (control)." The effect strength for sensitivity (ESS) is labelled in the output as "Effect Strength PAC". The ESS is 7.97% (a weak effect)[14] in the training analysis. The permutation *P*-value for the training sample was a non-statistically significant $P = 0.215$. In summary, ODA was unable to discriminate between treatment groups on the basis of age (i.e., covariate balance was achieved).

The Table below presents a comparison of all baseline characteristics of program participants and their 1:1 propensity score matched controls in this study. Values represent cut-points on the covariate, and values in parentheses represent sensitivity (for participants) and specificity (for matched controls), which are listed in the **oda** as "PAC TREAT = 1" and "PAC TREAT = 0".

2155-0182/10/$3.00

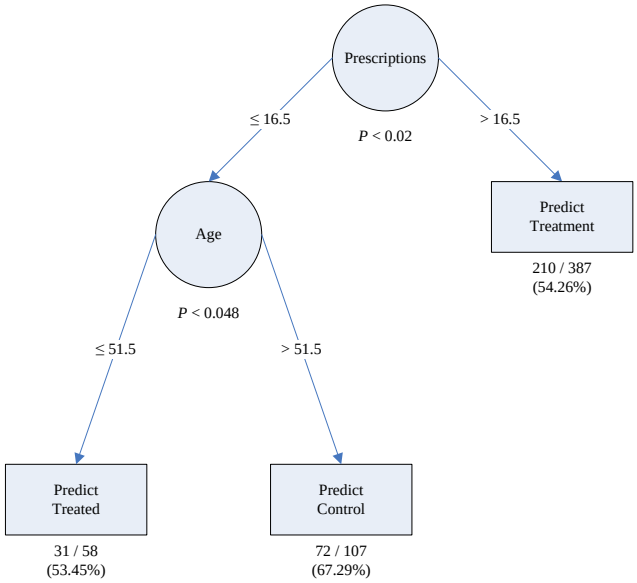
	Participants (N=276)	Matched Controls (N=276)	ESS	*P*-value
Demographic characteristics				
Age	<= 52.5 (37.68)	> 52.5 (70.29)	7.97%	0.215
Female	= 0 (45.65)	= 1 (55.07)	0.72%	0.932
Utilization and Cost				
Primary care visits	> 2.5 (96.38)	<= 2.5 (9.06)	5.43%	0.590
Other outpatient visits	> 5.5 (74.64)	<= 5.5 (35.51)	2.90%	0.727
Laboratory tests	> 3.5 (59.06)	<= 3.5 (49.64)	8.70%	0.112
Radiology tests	> 0.5 (80.80)	<= 0.5 (24.28)	5.07%	0.519
Prescriptions filled	> 16.5 (76.09)	<= 16.5 (35.87)	11.96%	0.027
Hospitalizations	> 0.5 (14.86)	<= 0.5 (87.32)	2.17%	0.546
Emergency department visits	> 0.5 (21.38)	<= 0.5 (84.06)	5.43%	0.123
Home-health visits	> 2.5 (1.45)	<= 2.5 (98.91)	0.36%	0.853
Total costs	> 4629 (49.64)	<= 4629 (61.23)	10.87%	0.079

As highlighted in the Table, the prescription fills variable is not balanced between the treatment groups in the training data.

The primary criticism of assessing covariate balance on individual covariates only is that any interactions that may exist between covariates are ignored, and indeed, these interactions may prove to confound treatment effects whereas individual covariates appear not to.[15-18]

A straightforward way to determine whether any statistically significant interactions exist in the data is to conduct a classification tree analysis (CTA).[5,15] While I will elaborate on this approach in a future article, here I will simply report the results from a CTA model using the Stata package **cta**[19] which can be downloaded from within Stata by typing at the Stata command line: "ssc install cta" (without the quotation marks).

The CTA model obtained using **cta** had an overall ESS of 13.41%, which is a relatively weak effect. The diagram below illustrates the identified interaction effect between prescription fills and age that predicts whether an individual is a treated participant, or a matched control. More specifically, the model predicts treatment if the individual had > 16.5 prescriptions filled or if the individual had ≤ 16.5 prescriptions filled and was ≤ 51.5 years of age. The model predicts an individual to be a matched control if they filled ≤ 16.5 prescriptions in the baseline period and were > 51.5 years of age.



Discussion

This paper demonstrates how the new Stata package **oda** can be used to assess covariate balance in observational studies. Further, the

2155-0182/10/$3.00

importance of testing for interactions amongst covariates to ensure that potential confounding relationships are identified was demonstrated. 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.[14] Moreover, in contrast to regression or ANOVA models, ODA also has the distinct ability to ascertain where the 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 which allows for assessing the cross-generalizability of the model to potentially new study participants or non-participants.

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.[15-36]

References

[1]Linden A (2020). Implementing ODA from Within Stata: An Application to Data From a Randomized Controlled Trial (*Invited*). *Optimal Data Analysis,* 9, 9-13.

[2]Linden A (2020). Implementing ODA from Within Stata: Implementing ODA from Within Stata: An Application to Estimating Treatment Effects using Observational Data (*Invited*). *Optimal Data Analysis,* 9, 14-20.

[3]Linden A (2020). Implementing ODA from Within Stata: An Application to Dose-Response Relationships (*Invited*). *Optimal Data Analysis,* 9, 26-32.

[4]Linden A (2020). ODA: Stata module for conducting Optimal Discriminant Analysis. *Statistical Software Components S458728, Boston College Department of Economics*.

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

[6]Linden A, Adams J (2006). Evaluating disease management program effectiveness: an introduction to instrumental variables. *Journal of Evaluation in Clinical Practice,* 12, 148-154.

[7]Linden 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.

[8]Linden 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.

[9]Linden 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.

[10]Linden A, Uysal SD, Ryan A, Adams JL (2016). Estimating causal effects for multivalued treatments: A comparison of approaches. *Statistics in Medicine, 35,* 534-552.

[11]Linden A (2011). Identifying spin in health management evaluations. *Journal of Evaluation in Clinical Practice, 17,* 1223-1230.

[12]Rosenbaum PR (1989). Optimal matching for observational studies. *Journal of the American Statistical Association,* 84, 1024–1302.

2155-0182/10/$3.00

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

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

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

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

[17]Linden 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.

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

[19]Linden A. (2020). CTA: Stata module for conducting Classification Tree Analysis. *Statistical Software Components S458729, Boston College Department of Economics*.

[20]Linden 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).

[21]Linden 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.

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

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

[24]Linden 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.

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

[26]Linden 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.

[27]Linden 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.

[28]Linden 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.

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

2155-0182/10/$3.00

[30]Linden 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.

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

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

[33]Yarnold 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.

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

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

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

Author Notes

No conflict of interest was reported.