Optimal Data Analysis
Vol. 9 (March 1, 2020), 39-44

2155-0182/10/$3.00

Implementing ODA from Within Stata: Evaluating Treatment Effects for Survival (Time-to-Event) Outcomes (*Invited*)

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

In this paper, I describe how to evaluate treatment effects in non-randomized studies with a time-to-event outcome using the new Stata package for implementing ODA.

Prior papers[1-4] introduced the new Stata package called **oda**[5] for implementing ODA from within the Stata environment. This package is a wrapper for the MegaODA software[6], 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 evaluate treatment effects in observational studies with a time-to-event outcome, such as survival, onset of disease, or hospital readmission. Time-to-event outcomes require specialized models designed to assess the influence of covariates on the outcome in the presence of *censoring*. Survival times are called censored to indicate that the study terminated before the event occurred, or that the individual was lost to follow-up at some point during the study.[7]

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.[8-13]

In **oda** we evaluate treatment effects for time-to-event outcomes by specifying the outcome indicator (e.g., dead or alive at the end of follow-up) as the *class* variable and the treatment variable as the *attribute*. To account for censoring, follow-up times are specified as a weight using the *wt*() option. To control for confounding, a propensity score weight is computed

2155-0182/10/$3.00

and then multiplied by follow-up time. This new weight is then specified in the *wt*() option instead of specifying follow-up time alone.

Methods

Data

This paper uses data from a prior evaluation of a health plan–based program intended to reduce 30-day readmission rates for patients hospitalized with one or more chronic illnesses. The intervention was modeled after that described in Linden and Butterworth,[14] which focused on behavioral change to help patients actively engage in their own health care, which in turn was expected to reduce the likelihood of readmission.[15-18] This subset of the retrospectively collected data consists of observations for 1398 participants and 7957 nonparticipants.

Ten pre-intervention characteristics available for every observation included demographic variables (age and gender), health services use in the 12 months prior to the index hospitalization (office visits, emergency department [ED] visits, hospitalizations), length of stay (LOS) for the index hospitalization, indicator variables for if the patient had congestive heart failure and/or chronic obstructive pulmonary disease, the patient's Charlson comorbidity index score (CCI),[19] and a diagnosis-based risk adjustment score. The outcome was number of days post-discharge from the index hospitalization: patients were classified as censored if they were lost to follow-up prior to 30 days, or did not experience a readmission within 30 days.

A comprehensive discussion on use of ODA (and CTA) for evaluating treatment effects for time-to-event outcomes is available elsewhere.[20] Discussion here is limited to its implementation using the **oda** package.

Analytic process

For evaluating treatment effects with no adjustment for confounding (e.g., when the data come from an RCT), **oda** can be implemented with the following syntax (see the help file for **oda** for a complete description of the syntax options):

```
oda event treat, pathoda("C:\ODA\")
store("C:\ODA\output") iter(25000)
direction(< 1 0) wt(time) seed(1234)
```

The above syntax is explained as follows: The outcome variable “event” (readmitted or censored by day 30) is the *class* variable; “treat” 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 25,000; the directional (one-sided) hypothesis is that the treatment group will have fewer events than the control group; the analysis is weighted by follow-up time; and the seed is set to 1234 to allow us to replicate permutation results (any integer value can be used as 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 TREAT <= 0.5 THEN EVENT = 1
IF 0.5 < TREAT THEN EVENT = 0


Summary for Class EVENT  Attribute TREAT
----------------------------------------

Performance Index           Train
-----------------          ------
Overall Accuracy            25.74%
Overall Wtd Accuracy        20.27%
PAC EVENT=1                 91.53%
PAC EVENT=0                 15.91%
Effect Strength PAC          7.44%
Wtd PAC EVENT=1             89.48%
Wtd PAC EVENT=0             15.91%
Effect Strength Wtd PAC      5.39%
PV EVENT=1                  13.99%
PV EVENT=0                  92.63%
Effect Strength PV           6.62%
Wtd PV EVENT=1               6.28%
Wtd PV EVENT=0              96.00%
Effect Strength Wtd PV       2.28%
Effect Strength Total        7.03%
Effect Strength Wtd Total    3.84%


Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  25000
Estimated p: 0.000000
```

2155-0182/10/$3.00

As shown in the **oda** output, the ODA model can be interpreted as follows: "if the study group <= 0.5 (i.e., control), then predict that the event is 1 (i.e., readmitted within 30 days). If the study group is > 0.5 (i.e., treatment), then predict that the event is 0 (i.e., no readmission or lost to follow-up by day 30)." The weighted effect strength for sensitivity (WESS) is labelled in the output as "Effect Strength Wtd PAC". The WESS is 5.39% (a weak effect)[21] in the analysis. The one-sided permutation *P*-value for the sample is statistically significant $P < 0.0001$. In summary, while ODA had a difficult time finding a cut-point on the outcome to adequately discriminate between treatment groups, the difference in the outcome between groups was statistically significant (i.e., the treatment group was associated with fewer readmissions in 30 days).

For evaluating treatment effects with observational data, a propensity score weighting approach must first be applied to the sample to adjust for confounding.[9-12] Here we apply the marginal mean weighting through stratification (MMWS)[11] approach (implemented in Stata using the **mmws** package). Next, this weight is multiplied by the follow-up time variable and saved as new variable called "tXwt". This analysis can then be implemented in **oda** with the following syntax:

```
oda event treat, pathoda("C:\ODA\")
store("C:\ODA\output") iter(25000)
direction(< 1 0) wt(tXwt) seed(1234)
```

The above syntax is identical to the unadjusted model described previously with the exception being that the *wt*() option specified here is the "tXwt" weight.

As shown in the **oda** output, the ODA model can be interpreted as follows: "if the study group <= 0.5 (i.e., control), then predict that the event is 1 (i.e., readmitted within 30 days). If the study group is > 0.5 (i.e., treatment), then predict that the event is 0 (i.e., no readmission or lost to follow-up by day 30)." The WESS is 3.43% (a weak effect)[21] in the analysis. The one-sided permutation *P*-value for the sample is statistically significant $P = 0.0008$. These results (and interpretation) are very similar to the model adjusting only for time, that is, while ODA had a difficult time in finding a cut-point on the outcome to adequately discriminate between treatment groups, the difference in the outcome between groups was statistically significant (i.e., the treatment group was associated with fewer readmissions in 30 days).

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


Summary for Class EVENT  Attribute TREAT
----------------------------------------

Performance Index           Train
-----------------          ------
Overall Accuracy            25.74%
Overall Wtd Accuracy        19.97%
PAC EVENT=1                 91.53%
PAC EVENT=0                 15.91%
Effect Strength PAC          7.44%
Wtd PAC EVENT=1             87.69%
Wtd PAC EVENT=0             15.73%
Effect Strength Wtd PAC      3.42%
PV EVENT=1                  13.99%
PV EVENT=0                  92.63%
Effect Strength PV           6.62%
Wtd PV EVENT=1               6.12%
Wtd PV EVENT=0              95.32%
Effect Strength Wtd PV       1.45%
Effect Strength Total        7.03%
Effect Strength Wtd Total    2.44%


Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  25000
Estimated p: 0.000800
```

Discussion

This paper demonstrates how the new Stata package **oda** can be used to evaluate treatment effects in studies with a time-to-event outcome. ODA should be considered the preferred approach over commonly-used survival models because 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.[6] Moreover, in contrast to conventional statistical models, ODA also has the distinct ability to ascertain where optimal (maximum-accuracy) cutpoints are on the outcome variable, which in

2155-0182/10/$3.00

turn, facilitates the use of measures of predictive accuracy. Moreover, ODA can perform cross-validation using LOO which assesses the cross-generalizability of the model to potentially new study participants or non-participants.[22]

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.[23-42]

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). Implementing ODA from Within Stata: Assessing Covariate Balance in Observational Studies (*Invited*). *Optimal Data Analysis*, 9, 33-38.

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

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

[7]Linden A, Adams J, Roberts N (2004). Evaluating disease management program effectiveness: an introduction to survival analysis. *Disease Management*, 7, 180-190.

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

[9]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.

[10]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.

[11]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.

[12]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.

[13]Linden A (2007). Estimating the effect of regression to the mean in health management programs. *Disease Management and Health Outcomes,*15, 7-12.

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

[15]Linden A, Roberts N (2004). Disease management interventions: What's in the black box? *Disease Management*, 7, 275-291.

[16]Linden A, Butterworth S, Roberts N (2006). Disease management interventions II: what else is in the black box? *Disease Management*, 9, 73-85.

2155-0182/10/$3.00

[17]Biuso TJ, Butterworth S, Linden A (2007). Targeting prediabetes with lifestyle, clinical and behavioral management interventions. *Disease Management*, 7, 6-15.

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

[19]Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Disease*, 40, 373-383.

[20]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.

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

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

[23]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).

[24]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.

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

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

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

[28]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.

[29]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.

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

[31]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.

[32]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.

[33]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.

[34]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

[35]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.

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

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

[38]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.

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

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

[41]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.