Optimal Data Analysis
Vol. 10 (March 31, 2021), 42-49

2155-0182/10/$3.00

Implementing ODA from Within Stata: Evaluating Test-Retest Reliability of Positive and Negative Emotional *States*, *vs*. Personality *Traits*, Assessed Using Likert Scales, for Males *vs*. Females

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

This paper illustrates testing directional hypotheses for test-retest Likert ratings of positive and negative emotional states and personality traits for males and females, using the Stata package for implementing ODA.

Recent papers[1-26] introduce the new Stata package called **oda**[27] for implementing ODA from within the Stata environment. This package is a wrapper for the MegaODA software system[28-30], so the MegaODA.exe file must be loaded on the computer for **oda** to work.[31] To download the **oda** package, at the Stata command line type: "ssc install oda" (without the quotation marks).

Using test-retest methodology to assess temporal reliability requires a minimum of two administrations of a single test, survey, or other measurement methodology. In the absence of test-specific learning a person (or any object of assessment) completing the same test twice is hypothesized to obtain identical scores on both administrations.[32,33] Discussion concerning the advantages and challenges associated with this popular methodology is presented elsewhere.[32]

This paper demonstrates using **oda** to evaluate directional hypothesis for a design assessing test-retest reliability of state (transitory) *vs*. trait (enduring) phenomena, measured using Likert scales, for combined and separate samples of male and female undergraduates.[32]

Methods

Data

Bryant and Yarnold obtained data assessing two-week temporal stability of the emotional experience for 160 undergraduates. Affect was measured using single-word descriptors of their current state of mind, which were rated using 5-point Likert items (0=not at all accurate; 4=very accurate).[34] The four items used in this example were peeved, lonely, cheerful, and friendly. The

2155-0182/10/$3.00

first and second items assess negative affect, and the third and fourth items assess positive affect. The first and third items measure temporary, rapidly changing affect (emotional *states*), whereas the second and fourth items reflect stable, slowly changing affect (personality *traits*).

Each observation in the sample received two scores for each of the four affects: for each the first assessment was indicated by placing the numeral “1” after the name of the affect (e.g., peeved1, friendly1), and the second assessment was indicated by placing the numeral “2” after the name of the affect (e.g., peeved2, friendly2). Each observation’s gender was also recorded: male=1, female=0.

<u>*Analytic Process*</u>

We separately evaluate test-retest reliability of all four affective measures, beginning with the total sample. We test the directional (“confirmatory”) alternative hypothesis that a person’s affect (i.e., Likert score) at the first testing (class variable) is identical to the corresponding affect at the second testing (attribute). The null hypothesis is this is not true. Of course, if theoretically salient, the second rating could be treated as the class variable, and the first rating as the attribute (a recommended exercise for those learning the ODA paradigm).[32,35]

The first analysis was run for “peeved” (hypothesized to reflect a transitory state) via the following **oda** syntax (see the **oda** help file for a complete description of syntax options):

```
oda peeved1 peeved2, pathoda("C:\ODA\")
store("C:\ODA\") iter(10000)
direction(< 0 1 2 3 4) sidak(4)
```

This syntax is explained as follows: peeved1 is the *class* variable; peeved2 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 directional command specifies the hypothesis that responses to the class variable and attribute should be directly consistent; and the hypothesis is evaluated using the Sidak Type I error rate[32] (*p* value corrected for four hypothesis tests, one for each of the four variables analyzed in this example).

No solution was found for this problem, so the *a priori* hypothesis was not supported. Failure to identify a solution implies that at least one class category level was empty (i.e., that no observation used one or more of the category levels for this variable), or the hypothesized directional structure is untenable given the actual responses of the observations. Since none of the class categories is empty, it is concluded that the directional hypothesis is untenable given the structure underlying observations’ responses).

The possibility that a *nonlinear* relationship underlies the temporal mapping from peeved1 to peeved2 may be tested, however exposition of this methodology lies outside the scope of this article.[32,36-41]

Similarly, no solution was identified for the other state affect, cheerful (the oda syntax is identical to the syntax for peeved, except that “cheerful” is substituted for “peeved”). Thus, emotional *states* examined in this study did not demonstrate two-week test-retest reliability when combining data for males and females.

It is possible that combining data for males and females induced Simpson’s paradox, whereby the result obtained for the total sample differed from the result obtained for the separate groups.[32,38,42] This possibility is easily checked by adding the following command to the Stata **oda** program to use only the data for males:

```
if gender==1,
```

or

```
if gender==0,
```

2155-0182/10/$3.00

to use only the data for females. In both cases, as for the combined data, no solution was identified: no evidence of paradoxical confounding emerged in analysis of emotional states.

Analysis next evaluated the test-retest reliability of putatively stable personality traits: lonely (indicated as lonely1 and lonely2); and friendly (friendly1, friendly2). As before, the first analysis considered data for the total sample combining data for males and females.

Considering first the results for lonely, the **oda** package produces the following extract of the total output produced by ODA software (the complete output is stored in the specified directory with the extension ".out").

```
ODA model:
----------
IF LONELY1 <= 0.5 THEN LONELY2 = 0
IF 0.5 < LONELY1 <= 1.5 THEN LONELY2 = 1
IF 1.5 < LONELY1 <= 2.5 THEN LONELY2 = 2
IF 2.5 < LONELY1 <= 3.5 THEN LONELY2 = 3
IF 3.5 < LONELY1 THEN LONELY2 = 4


Summary for Class LONELY2  Attribute LONELY1
--------------------------------------------

Performance Index          Train
-----------------         ------
Overall Accuracy           59.38%
PAC LONELY2=0              77.11%
PAC LONELY2=1              42.86%
PAC LONELY2=2              35.00%
PAC LONELY2=3              27.27%
PAC LONELY2=4              75.00%
Effect Strength PAC        39.31%
PV LONELY2=0               77.11%
PV LONELY2=1               40.91%
PV LONELY2=2               36.84%
PV LONELY2=3               30.00%
PV LONELY2=4               75.00%
Effect Strength PV         39.96%
Effect Strength Total      39.64%


Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  10000
Estimated p: 0.000000
Sidak Adjusted (4) p:         0
```

As seen, classification performance of the ODA model corresponds to a statistically significant (conventionally reported as per-comparison $p<0.0001$) effect. Effect strength for sensitivity[43] (ESS) is labelled in the output as "Effect Strength PAC" (Percentage Accurate Classification). For the confirmatory hypothesis ESS is 39.31%, which meets the criterion ($25<\text{ESS}\leq 50$) for classification as a moderate effect.[32] An exact discrete confidence interval (CI) for this result, as well as for chance, may be computed.[44,45]

Here, PAC is "percentage accuracy in classification," also called model sensitivity.[32] Since there are five class categories, sensitivity of 100%/5 = 20% is expected by chance for each of the five categories.[43] Best performance was achieved for class 0 (not at all accurate), and for class 4 (very accurate). Exact discrete confidence intervals for sensitivity may be obtained for each class category, to ascertain which (if any) correspond to "guessing" (i.e., overlap the 95% CI for chance).[44,45]

Results obtained conducting the same analysis separately for males indicates a moderate ESS of 41.27%

```
ODA model:
----------
IF LONELY1 <= 0.5 THEN LONELY2 = 0
IF 0.5 < LONELY1 <= 1.5 THEN LONELY2 = 1
IF 1.5 < LONELY1 <= 2.5 THEN LONELY2 = 2
IF 2.5 < LONELY1 <= 3.5 THEN LONELY2 = 3
IF 3.5 < LONELY1 THEN LONELY2 = 4


Summary for Class LONELY2  Attribute LONELY1
--------------------------------------------

Performance Index          Train
-----------------         ------
Overall Accuracy           58.62%
PAC LONELY2=0              82.22%
PAC LONELY2=1              40.00%
PAC LONELY2=2              28.57%
PAC LONELY2=3              14.29%
PAC LONELY2=4             100.00%
Effect Strength PAC        41.27%
PV LONELY2=0               77.08%
PV LONELY2=1               38.10%
PV LONELY2=2               36.36%
PV LONELY2=3               20.00%
PV LONELY2=4               50.00%
Effect Strength PV         30.39%
Effect Strength Total      35.83%


Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  10000
Estimated p: 0.000000
Sidak Adjusted (4) p:         0
```

Conducting this analysis separately for females indicates moderate ESS of 45.79%.

Since the ESS obtained for the separate groups both exceed the ESS obtained for the combined sample, there is evidence of paradoxical confounding: thus, data should not be combined.[42]

Copyright 2021 by Optimal Data Analysis, LLC
2155-0182/10/$3.00

```
ODA model:
----------
IF LONELY1 <= 0.5 THEN LONELY2 = 0
IF 0.5 < LONELY1 <= 1.5 THEN LONELY2 = 1
IF 1.5 < LONELY1 <= 2.5 THEN LONELY2 = 2
IF 2.5 < LONELY1 <= 3.5 THEN LONELY2 = 3
IF 3.5 < LONELY1 THEN LONELY2 = 4

Summary for Class LONELY2  Attribute LONELY1
--------------------------------------------

Performance Index          Train
-----------------         ------
Overall Accuracy           60.27%
PAC LONELY2=0              71.05%
PAC LONELY2=1              45.45%
PAC LONELY2=2              50.00%
PAC LONELY2=3              50.00%
PAC LONELY2=4              66.67%
Effect Strength PAC        45.79%
PV LONELY2=0               77.14%
PV LONELY2=1               43.48%
PV LONELY2=2               37.50%
PV LONELY2=3               40.00%
PV LONELY2=4              100.00%
Effect Strength PV         49.53%
Effect Strength Total      47.66%

Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  10000
Estimated p: 0.000000
Sidak Adjusted (4) p:         0
```

Unlike emotional states which failed to demonstrate stability, the personality trait “lonely” is moderately consistent across time.

A different finding emerged in analysis of the test-retest stability of the trait “friendly.” For the combined sample of males and females, moderate ESS=29.88.

```
ODA model:
----------
IF FRIEND1 <= 0.5 THEN FRIEND2 = 0
IF 0.5 < FRIEND1 <= 1.5 THEN FRIEND2 = 1
IF 1.5 < FRIEND1 <= 2.5 THEN FRIEND2 = 2
IF 2.5 < FRIEND1 <= 3.5 THEN FRIEND2 = 3
IF 3.5 < FRIEND1 THEN FRIEND2 = 4

Summary for Class FRIEND2  Attribute FRIEND1
--------------------------------------------

Performance Index          Train
-----------------         ------
Overall Accuracy           40.00%
PAC FRIEND2=0              45.16%
PAC FRIEND2=1              35.14%
PAC FRIEND2=2              36.84%
PAC FRIEND2=3              35.71%
PAC FRIEND2=4              66.67%
Effect Strength PAC        29.88%
PV FRIEND2=0               50.00%
PV FRIEND2=1               38.24%
PV FRIEND2=2               31.82%
PV FRIEND2=3               50.00%
PV FRIEND2=4               33.33%
Effect Strength PV         25.85%
Effect Strength Total      27.86%

Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  10000
Estimated p: 0.000000
Sidak Adjusted (4) p:         0
```

Separate analysis of the trait of friendly yielded ESS=36.24 for males.

```
ODA model:
----------
IF FRIEND1 <= 0.5 THEN FRIEND2 = 0
IF 0.5 < FRIEND1 <= 1.5 THEN FRIEND2 = 1
IF 1.5 < FRIEND1 <= 2.5 THEN FRIEND2 = 2
IF 2.5 < FRIEND1 <= 3.5 THEN FRIEND2 = 3
IF 3.5 < FRIEND1 THEN FRIEND2 = 4

Summary for Class FRIEND2  Attribute FRIEND1
--------------------------------------------

Performance Index          Train
-----------------         ------
Overall Accuracy           43.68%
PAC FRIEND2=0              47.06%
PAC FRIEND2=1              48.00%
PAC FRIEND2=2              31.58%
PAC FRIEND2=3              35.00%
PAC FRIEND2=4              83.33%
Effect Strength PAC        36.24%
PV FRIEND2=0               53.33%
PV FRIEND2=1               52.17%
PV FRIEND2=2               30.00%
PV FRIEND2=3               53.85%
PV FRIEND2=4               31.25%
Effect Strength PV         30.15%
Effect Strength Total      33.20%

Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  10000
Estimated p: 0.000000
Sidak Adjusted (4) p:         0
```

In contrast, separate analysis of the trait of friendly for females yielded ESS=19.91 (a weak effect[32]).

```
ODA model:
----------
IF FRIEND1 <= 0.5 THEN FRIEND2 = 0
IF 0.5 < FRIEND1 <= 1.5 THEN FRIEND2 = 1
IF 1.5 < FRIEND1 <= 2.5 THEN FRIEND2 = 2
IF 2.5 < FRIEND1 <= 3.5 THEN FRIEND2 = 3
IF 3.5 < FRIEND1 THEN FRIEND2 = 4

Summary for Class FRIEND2  Attribute FRIEND1
--------------------------------------------

Performance Index          Train
-----------------         ------
Overall Accuracy           35.62%
PAC FRIEND2=0              42.86%
PAC FRIEND2=1               8.33%
PAC FRIEND2=2              42.11%
PAC FRIEND2=3              36.36%
PAC FRIEND2=4              50.00%
Effect Strength PAC        19.91%
PV FRIEND2=0               46.15%
PV FRIEND2=1                9.09%
PV FRIEND2=2               33.33%
PV FRIEND2=3               47.06%
PV FRIEND2=4               37.50%
Effect Strength PV         18.28%
Effect Strength Total      19.10%

Monte Carlo summary (Fisher randomization):
-------------------------------------------
Iterations:  10000
Estimated p: 0.002600
Sidak Adjusted (4) p: .01035951
```

2155-0182/10/$3.00

These findings indicate that males and females have approximately comparably stable traits of loneliness, however the two-week stability of the personality trait of friendliness is substantially greater for males than for females among the college undergraduates studied.

We believe ODA should be considered the preferred statistical approach *vs*. alternative methods since it avoids statistical assumptions required of conventional 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.[32] Compared to other methods, only ODA can identify the optimal (maximum-accuracy) assignments (categorical attributes) or cutpoints (ordered attributes) which exist for the attribute, that in turn facilitates use of measures of predictive accuracy. ODA can evaluate model reproducibility using multiple methods, allowing assessment of potential cross-generalizability of the model when applied to classify independent random samples.[32] We therefore recommend that researchers employ the ODA framework to evaluate the statistical hypotheses which are explored in their laboratory and field research endeavors.[46-64]

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). Implementing ODA from within Stata: Evaluating treatment effects for survival (time-to-event) outcomes (*Invited*). *Optimal Data Analysis*, *9*, 39-44.

[6]Linden A (2020). Implementing ODA from within Stata: Evaluating treatment effects in multiple-group interrupted time series analysis (Invited). *Optimal Data Analysis*, *9*, 45-50.

[7]Linden A (2020). Implementing ODA from within Stata: identifying structural breaks in single-group interrupted time series designs (Invited). *Optimal Data Analysis*, *9*, 51-56.

[8]Linden A (2020). Implementing ODA from within Stata: Finding the optimal cut-point of a diagnostic test or index (Invited). *Optimal Data Analysis*, *9*, 74-78.

[9]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Exploratory hypothesis, binary class variable, and binary attribute. *Optimal Data Analysis*, *9*, 94-98.

[10]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Confirmatory hypothesis, binary class variable, and binary attribute. *Optimal Data Analysis*, *9*, 99-103.

[11]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Exploratory hypothesis, binary class variable, and binary attribute. *Optimal Data Analysis*, *9*, 104-108.

[12]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Exploratory hypothesis, binary class variable, and ordinal (rank) attribute. *Optimal Data Analysis*, *9*, 109-113.

[13]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: confirmatory hypothesis, binary class variable, and ordinal attribute. *Optimal Data Analysis*, *9*, 128-132.

[14]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Exploratory hypothesis, binary class variable, categorical ordinal attribute. *Optimal Data Analysis*, *9*, 133-136.

[15]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Nondirectional hypothesis, binary class variable, categorical ordinal attribute. *Optimal Data Analysis*, *9*, 137-140.

[16]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Directional hypothesis, binary class variable, ordinal attribute. *Optimal Data Analysis*, *9*, 141-145.

[17]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Confirmatory hypothesis, binary class variable, continuous attribute. *Optimal Data Analysis*, *9*, 146-151.

[18]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Nondirectional, multi-categorical class variable, multicategorical attribute. *Optimal Data Analysis*, *9*, 152-156.

[19]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Directional hypothesis, multicategorical class variable and attribute. *Optimal Data Analysis*, *9*, 157-161.

[20]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: Directional hypothesis, multicategorical class variable, ordinal attribute. *Optimal Data Analysis*, *9*, 162-166.

[21]Yarnold PR, Linden A (2020). Implementing ODA from within Stata: *A Priori* hypothesis, three-category class variable, four-level (integer) attribute. *Optimal Data Analysis*, *9*, 167-171.

[22]Linden A, Yarnold PR (2020). Implementing ODA from within Stata: A reanalysis of the National Supported Work Experiment. *Optimal Data Analysis*, *9*, 178-182.

[23]Yarnold PR, Linden A (2021). Implementing ODA from within Stata: Exploratory hypothesis, three-category class variable, continuous attribute. *Optimal Data Analysis*, *10*, 3-9.

[24]Yarnold PR, Linden A (2021). Implementing ODA from within Stata: Confirmatory and exploratory inter-rater reliability hypothesis with a three-category ordinal rating. *Optimal Data Analysis*, *10*, 12-17

[25]Yarnold PR, Linden A (2021). Implementing ODA from within Stata: Assessing parallel-forms reliability using a binary and an ordered attribute. *Optimal Data Analysis*, *10*, 24-29.

[26]Yarnold PR, Linden A (2021). Implementing ODA from within Stata: Assessing split-half reliability using a polychotomous attribute. *Optimal Data Analysis*, *10*, 30-35.

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

[28]Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Separating the chaff. *Optimal Data Analysis*, *2*, 194-197.

[29]Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Harvesting the Wheat. *Optimal Data Analysis*, *2*, 202-205.

[30]Yarnold PR, Soltysik RC (2013). MegaODA large sample and BIG DATA time trials: Maximum velocity analysis. *Optimal Data Analysis*, *2*, 220-221.

2155-0182/10/$3.00

[31]Rhodes 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

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

[33]Yarnold PR (2014). UniODA *vs*. kappa: Evaluating the long-term (27-year) test-retest reliability of the Type A Behavior Pattern. *Optimal Data Analysis*, *3*, 14-16.

[34]Bryant FB, Yarnold PR (1990). The impact of Type A behavior on subjective life quality: Bad for the heart, good for the soul? *Journal of Social Behavior and Personality*, *5*, 369-404.

[35]Bryant FB, Harrison PR (2013). How to create an ASCII input data file for UniODA and CTA software (*Invited*). *Optimal Data Analysis*, *2*, 2-6.

[36]Yarnold PR (2014). How to assess inter-observer reliability of ratings made on ordinal scales: Evaluating and comparing the Emergency Severity Index (Version 3) and Canadian Triage Acuity Scale. *Optimal Data Analysis*, *3*, 42-49.

[37]Yarnold PR (2014). How to assess the inter-method (parallel-forms) reliability of ratings made on ordinal scales: Evaluating and comparing the Emergency Severity Index (Version 3) and Canadian Triage Acuity Scale. *Optimal Data Analysis*, *3*, 50-54.

[38]Yarnold PR (2015). Estimating inter-rater reliability using pooled data induces paradoxical confounding: An example involving Emergency Severity Index triage ratings. *Optimal Data Analysis*, *4*, 21-23.

[39]Yarnold, PR (2015). UniODA *vs*. Spearman rank ρ: Between-raters reliability of scores on the Adverse Drug Reaction Probability Scale. *Optimal Data Analysis*, *4*, 148-150.

[40]Yarnold, PR (2016). Novometric *vs*. ODA reliability analysis *vs*. polychoric correlation with relaxed distributional assumptions: Inter-rater reliability of independent ratings of plant health. *Optimal Data Analysis*, *5*, 179-183.

[41]Yarnold PR (2019). Regression *vs*. novometric-based assessment of inter-examiner reliability. *Optimal Data Analysis*, *8*, 107-111.

[42]Yarnold PR (1996). Characterizing and circumventing Simpson's paradox for ordered bivariate data. *Educational and Psychological Measurement*, *56*, 430-442.

[43]Yarnold PR (2018). Visualizing application and summarizing accuracy of ODA models. *Optimal Data Analysis*, *7*, 85-89.

[44]Yarnold PR (2018). Comparing exact discrete 95% CIs for model *vs*. chance ESS to evaluate statistical significance. *Optimal Data Analysis*, *7*, 82-84.

[45]Rhodes JN, Yarnold PR (2020). Generating novometric confidence intervals in R: Bootstrap analyses to compare model and chance ESS. *Optimal Data Analysis*, *9*, 172-177.

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

[47]Yarnold PR, Linden A. (2016). Novometric analysis with ordered class variables: The optimal alternative to linear regression analysis, *Optimal Data Analysis*, *5*, 65-73.

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

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

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

[51]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.

[52]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.

[53]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.

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

[55]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.

[56]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.

[57]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.

[58]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.

[59]Rhodes NJ (2020). Statistical power analysis in ODA, CTA and Novometrics (Invited). *Optimal Data Analysis*, *9*, 21-25.

[60]Yarnold PR, Brofft GC (2013). ODA range test *vs*. one-way analysis of variance: Comparing strength of alternative line connections. *Optimal Data Analysis*, *2*, 198-201.

[61]Yarnold PR (2013). ODA range test *vs*. one-way analysis of variance: Patient race and lab results. *Optimal Data Analysis*, *2*, 206-210.

[62]Yarnold PR (2014). How to assess the inter-method (parallel-forms) reliability of ratings made on ordinal scales: Evaluating and comparing the Emergency Severity Index (Version 3) and Canadian Triage Acuity Scale. *Optimal Data Analysis*, *3*, 50-54

[63]Yarnold PR (2016) Causality of adverse drug reactions: The upper-bound of arbitrated expert agreement for ratings obtained by WHO and Naranjo algorithms. *Optimal Data Analysis*, *5*, 37-40.

[64]Yarnold PR (2016). Novometric *vs*. ODA reliability analysis *vs*. polychoric correlation with relaxed distributional assumptions: Inter-rater reliability of independent ratings of plant health. *Optimal Data Analysis*, *5*, 179-183.

Author Notes

No conflicts of interest were reported.