Optimal Data Analysis
Vol. 4 (December 3, 2015), 148-150

2155-0182/10/$3.00

UniODA *vs.* Spearman Rank ρ: Between-Raters Reliability of Scores on the Adverse Drug Reaction Probability Scale

Paul R. Yarnold, Ph.D.
Optimal Data Analysis, LLC

The Adverse Drug Reaction Probability Scale (APS) algorithm is widely-used for rating the probability that an adverse drug event is drug-induced. Prior research (Figure 1, p. 695) presented APS ratings generated by two independent experts for $N = 129$ challenging cases.[1] Between-rater reliability of these ratings is computed by and compared between Spearman's rank-order correlation (r_S) and UniODA.

For these data $r_S = 0.79$, $p < 0.0001$: by rule-of-thumb this finding offers strong support for the *a priori* hypothesis that a statistically significant positive monotonic association exists between APS scores for the pair of raters.[1] This level of association yields modest predictive accuracy, primarily accurately predicting values close to the sample mean or median.[2-4] Eyeball analysis and an unspecified computation indicated "high weighted agreement (94.3%)" (p. 695).

The directional UniODA inter-rater reliability model[2-8] was identified using the UniODA[3,4] and MegaODA[9-11] software syntax shown here (a total of 10 class categories can be analyzed by this software, so rater 2 was treated as the class variable[1]):

```
OPEN adr.dat;
OUTPUT adr.out;
VARS r1 r2;
CLASS r2;
ATTRIBUTE r1;
DIR < 1 2 3 4 5 6 7 8;
LOO;
MCARLO ITER 25000;
GO;
```

The resulting UniODA model is presented in Table 1.

Table 1: UniODA Inter-Rater Model

If Rater #1 has APS Score of:	THEN PREDICT	That Rater #2 has APS Score of:
≤ 2		1
3		2
4		3
5		4
6		5
7		6
8		7
9		8

2155-0182/10/$3.00

Leave-one-out (LOO) validity analysis was not possible because at least two cases are needed in every class category. The confusion table for the UniODA model is presented in Table 2: hypothesized predictions are shown in **bold** along the major diagonal; predictive errors of magnitude are shown in **red** (i.e., the basic score—higher or lower than median risk—is not violated by the noted error); and errors of direction (and magnitude) are shown in **green**.

Table 2: UniODA Inter-Rater Confusion Table

		Predicted APS Score							
		1	2	3	4	5	6	7	8
	1	**1**	0	0	0	0	0	0	0
	2	**13**	**3**	**1**	0	0	0	0	0
Actual	3	**6**	**5**	**2**	0	**1**	0	0	0
APS	4	0	**3**	**3**	**3**	**1**	**1**	0	0
Score	5	0	**2**	**2**	**1**	**19**	**4**	**1**	0
	6	0	**1**	**2**	**2**	**22**	**11**	**6**	0
	7	0	0	0	0	**2**	**4**	**5**	**1**
	8	0	0	0	0	0	0	0	**1**

This between-rater agreement level was statistically significant ($p < 0.0001$): it reflects high-moderate predictive accuracy for overall classification (*ESS* = 41.6), but weak-moderate accuracy for point predictions (*ESP* = 26.5).[12]

The findings suggest that the lowest possible APS score, and the highest possible APS score, may be overly precise (out-of-reach) in most complex real-world cases. It is unlikely that non-expert users of the APS algorithm will produce highly reliable (or valid) rating scores.

References

[1]Lanctôt KL, Naranjo CA (1995). Comparison of the Bayesian approach and a simple algorithm for assessment of adverse drug events. *Clinical Pharmacology and Therapeutics, 58,* 692-698. DOI: 10.1016/0009-9236(95)90026-8

[2]Yarnold 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. URL: http://optimalprediction.com/files/pdf/V2A3.pdf

[3]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*. Washington, DC: APA Books.

[4]Yarnold PR, Soltysik RC (In Review). *Maximizing predictive accuracy*. Chicago, IL: ODA Books.

[5]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. URL: http://optimalprediction.com/files/pdf/V3A5.pdf

[6]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. URL: http://optimalprediction.com/files/pdf/V3A15.pdf

[7]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. URL: http://optimalprediction.com/files/pdf/V3A16.pdf

[8]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. URL: http://optimalprediction.com/files/pdf/V4A6.pdf

2155-0182/10/$3.00

[9]Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Separating the chaff. *Optimal Data Analysis*, 2, 194-197. URL: http://optimalprediction.com/files/pdf/V2A29.pdf

[10]Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Harvesting the Wheat. *Optimal Data Analysis*, 2, 202-205. URL: http://optimalprediction.com/files/pdf/V2A31.pdf

[11]Yarnold PR, Soltysik RC (2013). MegaODA large sample and BIG DATA time trials: Maximum velocity analysis. *Optimal Data Analysis*, 2, 220-221. URL: http://optimalprediction.com/files/pdf/V2A35.pdf

[12]Yarnold PR (2013). Standards for reporting UniODA findings expanded to include *ESP* and all possible aggregated confusion tables. *Optimal Data Analysis*, *2*, 106-119. URL: http://optimalprediction.com/files/pdf/V2A19.pdf

Author Notes

The study analyzed de-individuated data and was exempt from Institutional Review Board review. No conflict of interest was reported.

Mail: Optimal Data Analysis, LLC
6348 N. Milwaukee Ave., #163
Chicago, IL 60646
USA