Optimal Data Analysis
Vol. 3 (March 28, 2014), 12-13

2155-0182/10/$3.00

UniODA *vs.* Weighted Kappa: Evaluating Concordance of Clinician and Patient Ratings of the Patient's Physical and Mental Health Functioning

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

This study investigated the concordance between clinician and patient assessments of patient's *physical* and *mental* functioning, made using 4-category ordinal scales, for a consecutive sample of 166 outpatients with rheumatoid arthritis. Weighted kappa isn't a normed statistic, but the respective weighted kappa statistic obtained for the assessments, 0.39 and 0.30, are believed to reflect fair concordance. Data were also analyzed via confirmatory UniODA models which hypothesized that physician and patient ratings agreed. Findings indicated relatively strong concordance for the physical health ratings (ESS=55.5), and moderate concordance for mental health ratings (ESS=43.3).

Table 1 is the inter-rater cross-classification table for physical health ratings.[1] A weighted kappa statistic was used to assess concordance between clinician and patient ratings. Weighted kappa has been criticized on many grounds[2] and its validity presently can be called into question. In Table 1 the null cell entries compromise the validity of estimated *p*, and the disproportionate row and column marginal totals compress the maximum attainable upper limit of the statistic, for example. For these data, weighted kappa was reported as 0.39, with estimated $p<0.0001$.

Detailed discussion and examples of the application of UniODA in reliability analysis is available elsewhere.[2] A confirmatory UniODA analysis was conducted presently to test the

Table 1: Agreement matrix comparing clinician versus patient ratings of patient's *physical* health status (Kwoh et al., 1992).

Patient Rating	Clinician Rating			
	Complete	Adequate	Limited	Incapacitated
Complete	11	12	0	0
Adequate	12	65	28	0
Limited	0	13	21	3
Incapacitated	0	0	0	1

Note: Tabled are frequency counts.

a priori hypothesis that the clinician and patient ratings agree—that is, that ratings fell into the major diagonal running from the upper left-hand

2155-0182/10/$3.00

corner to the lower right-hand corner of the inter-rater cross-classification table.[3] The model achieved relatively strong[2] ESS of 55.5, and was statistically significant (p<0.0001). Off-diagonal entries indicate disagreements between clinician and patient ratings, reflecting "local instability" along the domain of the rating scale.[2]

Table 2 presents the inter-rater cross-classification table for mental health ratings.[1] For these data, weighted kappa was reported as 0.30, with estimated p<0.0001. For these data a confirmatory UniODA analysis yielded a moderate ESS of 43.3, which was statistically significant (p<0.0001).

Table 2: Agreement matrix comparing clinician versus patient ratings of patient's *mental* health status (Kwoh et al., 1992).

Patient Rating	**Clinician Rating**			
	Complete	Adequate	Limited	Incapacitated
Complete	46	28	2	1
Adequate	19	40	6	0
Limited	2	18	2	1
Incapacitated	0	0	0	1

Note: Tabled are frequency counts.

In contrast to weighted kappa, UniODA tests the *a priori* hypothesis that raters agree; UniODA provides an exact Type I error rate; and the index of UniODA performance is normed—for *every* UniODA analysis the ESS which is expected by chance is 0, and errorless, prefect classification is 100. There is no apparent rational reason to use weighted kappa in applications such as the present.

References

[1]Kwoh CK, O'Connor GT, Regan-Smith MG, Olmstead EM, Brown LA, Burnett JB, Hochman RF, King K, Morgan GJ (1992). Concordance between clinician and patient assessment of physical and mental health status. *Journal of Rheumatology*, *19*, 1031-1037.

[2]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: A guidebook with software for Windows*. Washington, DC: APA Books. See p. 124 for discussion of weighted kappa. For local instability see pp. 137-138, and Figure 7.1.c.

[3]UniODA analysis of the physical rating data was accomplished using the following code: commands are indicated in red.[2]

```
OPEN DATA;
OUTPUT Kwoh.out;
VARS doctor patient;
CLASS patient;
ATTR doctor;
DIRECTIONAL < 1 2 3 4;
MCARLO ITER 25000;
DATA;
1 1 (11 times)
2 1 (12 times)
1 2 (12 times)
2 2 (65 times)
3 2 (28 times)
2 3 (13 times)
3 3 (21 times)
4 3 (3 times)
4 4
END DATA;
GO;
```

Author Notes

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

eMail: Journal@OptimalDataAnalysis.com