

Maximizing Overall Percentage Accuracy in Classification: Discriminating Study Groups in the National Pressure Ulcer Long-Term Care Study (NPULS)

Paul R. Yarnold, Ph.D.

Optimal Data Analysis, LLC

UniODA may be used to identify two different types of (weighted) maximum-accuracy models.^{1,2} First, ODA can identify models that explicitly maximize overall percentage accuracy in classification or PAC—that is, the percentage of the total sample that is correctly classified by the model. Second, ODA can identify models that explicitly maximize the predictive accuracy of the model normed against chance using the effect strength for sensitivity (ESS) statistic, that is both chance-corrected (0 = the predictive accuracy expected by chance for the application) and maximum-corrected (100 = perfect, errorless classification). Because comparatively little is known about optimal models that maximize PAC, this research note initiates a literature on the matter. The present exposition involves assessing if clinical and demographic factors can be discriminated on the basis of study group using a UniODA model that explicitly maximizes PAC.

In the original research study groups were aggregated *a priori* on the basis of risk for developing a pressure ulcer: Group 1 is the people at risk of developing a pressure ulcer at the start of the study (N = 1,293); Group 2 is people who developed a pressure ulcer during the study (N = 457); Group 3 is people having an existing pressure ulcer at the start of the study (N = 534); and Group 4 is people with new and existing pressure ulcers (N = 136).³ Statistical results obtained by chi-square versus UniODA models that maximized ESS, used to assess compara-

bility of the four study groups on various attributes, were recently reported.⁴ The analyses conducted using UniODA in the prior study are replicated herein, except that presently PAC was explicitly maximized (using MegaODA⁵⁻⁷ and UniODA¹ software this entails *not* weighting observations by prior odds vis-à-vis the control syntax: PRIORS OFF;).

Proceeding in the same order as the prior study⁴ a so-called degenerate model—failing to classify observations into one or more categories of the class variable^{1,2}—was obtained when

UniODA maximized PAC obtained using Group to discriminate (i.e., predict, classify) gender. Specifically, all observations were classified as being female, yielding overall PAC = 70.2%.

It is always true that $p < 0.99$ (and ESS = 0, that is, the level of predictive accuracy that is expected for the application by chance) for a degenerate model involving a two-category class variable.^{1,2}

Degenerate models were also obtained using Group to predict (PAC is given within parentheses): enteral feeding (79.9%), modular nutritional products (87.2%), vitamin/mineral supplements (69.0%), poor meal intake in initial four weeks of study (60.8%), cognitive impairment (73.4%), mobility issues (85.2%), incontinence (91.2%), hospitalization and emergency room visits (83.0%), and mortality (92.8%): for all models, ESS = 0, $p < 0.99$.

Non-degenerate models that were not statistically reliable^{1,2,8} were identified for oral nutritional supplement (PAC = 52.8%, ESS = 6.0, $p < 0.071$); BMI < 22 kg/m² (PAC = 54.5%, ESS = 4.1, $p < 0.16$); and weight loss over the 12-week-long study (PAC = 52.3%, ESS = 5.3, $p < 0.068$).

No statistically significant UniODA models that maximized overall PAC were identified in this application.

References

¹Yarnold PR, Soltysik RC. *Optimal data analysis: Guidebook with software for Windows*. Washington, DC: APA Books, 2005.

²Yarnold PR, Soltysik RC (2016). *Maximizing predictive accuracy*. Chicago, IL: ODA Books.

³Horn SD, Bender SA, Bergstrom N, Cook AS, Ferguson ML, Rimmasch HL, Sharkey SS, Smout RJ, Taler GA, Voss AC (2002). *Journal of the American Geriatrics Society*, 50, 1816-1825.

⁴Yarnold PR (2016). ODA vs. chi-square: Describing baseline data from the National Pressure Ulcer Long-Term Care Study (NPULS). *Optimal Data Analysis*, 2, 194-197.

⁵Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Separating the chaff. *Optimal Data Analysis*, 2, 194-197.

⁶Soltysik RC, Yarnold PR (2013). MegaODA large sample and BIG DATA time trials: Harvesting the wheat. *Optimal Data Analysis*, 2, 202-205.

⁷Yarnold PR, Soltysik RC (2013). MegaODA large sample and BIG DATA time trials: Maximum velocity analysis. *Optimal Data Analysis*, 2, 220-221.

⁸Yarnold PR, Soltysik RC (1991). Theoretical distributions of optima for univariate discrimination of random data. *Decision Sciences*, 22, 739-752.

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