Optimal Data Analysis
Vol. 4 (June 22, 2015), 87-103

2155-0182/10/$3.00

Optimal Statistical Analysis Involving a Confounding Variable

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

This paper compares maximum-accuracy statistical approaches that address the effect of a confounding variable on estimated association of class variable and attribute. An example involves modeling patient ratings of likelihood they will recommend an Emergency Department to others (class variable) on the basis of five aspects of physician care behavior (attributes), as well as patient satisfaction with the amount of time spent waiting to see the physician (confounder).

Partial UniODA is a two-step procedure which may be used to obtain a statistical model that maximizes classification accuracy (normed versus chance) achieved for a sample, by using an attribute to classify observations' actual class categories, and simultaneously "controlling for" (eliminating) the effect of a confounder.[1] In the first step of partial UniODA all the observations correctly classified by a UniODA model that treats the confounder as an attribute are dropped from the sample: the surviving observations in the reduced sample *weren't* correctly predicted by the confounder. The second step of partial UniODA assesses the non-confounded relation between the attribute and class variable using the reduced sample.[2]

Viewed conceptually, partial UniODA is an optimal (maximum-accuracy) analogue of general linear model (GLM) methods—partial correlation, analysis of covariance (ANCOVA), backward and stepwise multiple regression, and hierarchical linear model, for example—that remove *variance* attributable to a confounder (covariate) prior to assessing the relationship between the class variable and attribute.[3,4] In applications involving multiple confounders, partial UniODA is conducted iteratively—with each iteration removing surviving observations correctly classified by the current covariate for the current reduced sample. Known as structural decomposition, this procedure is analogous to principal components analysis except that the former maximizes *accuracy*, and the latter maximizes *variance*.[2,3] In the present context, structural decomposition is an optimal analogue to hierarchical and "last-to-enter" multiple regression approaches.[3-5]

In contrast to UniODA—for which no distributional assumptions are required[6-8], data must meet crucial distributional assumptions underlying GLM-based methods.[1-9] Also in contrast to UniODA, removing *variability* in an attribute that is attributable to the confounder uncommonly induces computational instability (due to the sum-of-squares for error approaching zero), or reduces statistical power in GLM methods, due to inaccuracy of GLM models—even in applications involving simple metrics

2155-0182/10/$3.00

and high R^2 values.[10-12] However, because UniODA models explicitly maximize classification *accuracy*, associated reduction in sample size occurring as correctly classified observations are removed rapidly reduces statistical power.[1,13] Partial UniODA is thus conservative unless the putative confounder is statistically benign (i.e., is unrelated to the attribute).

Viewed ecologically, partial variance and partial accuracy methods both have limited real-world applicability unless the covariate is benign, or control of the covariate is possible.

The present paper resolves limitations of partial approaches using an optimal interactive non-linear method in which level/category of a confounding ("moderating") variable influences attributes used to model corresponding sample strata.[14-23] First, an empirical example used to facilitate clarity of exposition is introduced.

Likelihood of a Patient Recommending the Emergency Department (ED) to Others

Alternative means of obtaining non-confounded optimal models are demonstrated with an example involving factors predicting the self-rated likelihood of a discharged ED patient recommending the ED to others.[24,25]

The study setting was an 800-bed urban university-based level 1 Trauma center with an annual census of 48,000 patients.[26] One week post-discharge, patients were mailed a survey assessing satisfaction with care they received in the ED (survey return rate=17%). Ratings were obtained of the likelihood of the patient recommending the ED to others, and of patient satisfaction with administrative, nurse, physician, laboratory, and family/friend aspects of care. Ratings were made using a five-point Likert-type scale with 1=*very poor*, 2=*poor*, 3=*fair*, 4=*good*, and 5=*very good*.

The dichotomous *class variable* being modeled is whether a patient is *ambivalent* (rating of likelihood of recommending ED to others=3, N=239) versus *likely to recommend* (rating of likelihood of recommending ED to others=4, N=584) the ED to others (total N=823 patients). Satisfaction ratings of aspects of care received from physicians—courtesy, took the patient's problem seriously, concern for patient comfort, explanation of test/treatment, and explanation of illness/ injury—are treated as the *attributes* (possible predictors).

Prior research reports that waiting time to see the physician—in part, a function of case mix—is a moderate-to-strong predictor of both patient satisfaction as well as recommendation ratings.[1,7,8,24-31] Because waiting time is neither directly nor reliably subject to physician control, the present research objective is to identify factors under physician control that influence patient recommendation ratings above and beyond the effect of waiting time—which is treated as the *confounding variable*.

Simple Bivariate Associations of Attributes and the Confounder with Likelihood of Recommending the ED to Others

All simple bivariate UniODA effects were consistent with respect to model *direction* (increasing satisfaction with patient-care and with waiting time predicts increasing likelihood of recommending the ED) and *optimal threshold* (critical rating value). For every simple effect the UniODA model was: if patient rating on the attribute or covariate is ≤ 3 (fair or worse) then predict patient likelihood to recommend the ED = 3 (ambivalent); otherwise, if patient rating on the attribute or covariate is >3 (good or very good) then predict patient likelihood to recommend the ED = 4 (likely to recommend ED to others). Simple bivariate associations between patient self-rated likelihood of recommending the ED to others, and patient ratings of satisfaction with waiting time and physician patient-care behaviors, are summarized in the first column of Table 1.

As seen, there was a statistically significant[2,6] relationship between patient satisfaction with waiting time and likelihood of recommending the ED (exact $p<0.0001$). The strength of

2155-0182/10/$3.00

Table 1: Empirical Comparison of Alternative Strategies for Controlling a Confounding Variable in the ODA Paradigm

Attribute	Bivariate Association with Self-Rated Likelihood of Recommending ED to Others	Alternative Strategies of Controlling a Confounding Variable in ODA (See Text for Detailed Explanation and Discussion)			
Waiting Time to See the Physician (confounder)	$p<0.0001$ ESS=30.78, D=4.5 N_3=238 (72.7, 41.8) N_4=575 (58.1, 83.7)	Discard Observations Correctly Predicted by the Confounding Variable	Treat the Confounding Variable as an Attribute in an HO-CTA Model	Treat the Confounding Variable as an Attribute in an EO-CTA Model	Treat the Confounding Variable as an Attribute in a GO-CTA Model
Physician Courtesy	$p<0.0001$ ESS=20.26, D=7.9 N_3=236 (26.3, 63.9) N_4=582 (94.0, 75.9)	$p<0.46$ ESS=6.73, D=27.7 N_3=65 (46.2, 24.0) N_4=241 (60.6, 80.7)	p's<0.0091 ESS=37.60, D=8.3 N_3=238 (68.9, 47.7) N_4=575 (68.7, 84.2)	p's<0.0093 ESS=39.66, D=7.6 N_3=236 (72.9, 47.4) N_4=575 (66.8, 85.7)	p's<0.0001 ESS=39.49, D=4.6 N_3=236 (56.4, 57.8) N_4=575 (83.1, 82.3)
Physician Took Patient's Problem Seriously	$p<0.0001$ ESS=24.83, D=6.1 N_3=237 (31.2, 66.7) N_4=579 (93.6, 76.9)	$p<0.015$ ESS=16.67, D=10.0 N_3=65 (24.6, 45.7) N_4=239 (92.1, 81.8)	p's<0.0001 ESS=40.91, D=5.8 N_3=238 (55.0, 61.8) N_4=573 (85.9, 82.1)	p's<0.0001 ESS=44.33, D=3.8 N_3=237 (61.6, 59.6) N_4=573 (82.7, 83.9)	EO-CTA and GO-CTA models were identical
Physician Concern For Patient's Comfort	$p<0.0001$ ESS=22.62, D=6.8 N_3=236 (34.3, 54.4) N_4=581 (88.3, 76.8)	$p<0.29$ ESS=8.08, D=22.8 N_3=65 (23.1, 29.4) N_4=240 (85.0, 80.3)	p's<0.031 ESS=40.16, D=7.5 N_3=238 (61.8, 54.2) N_4=574 (78.4, 83.2)	HO-CTA and EO-CTA models were identical	p's<0.0001 ESS=39.84, D=4.5 N_3=236 (61.4, 53.9) N_4=574 (78.4, 83.2)
Physician Explanation of Test/Treatment	$p<0.0001$ ESS=28.11, D=5.1 N_3=237 (41.4, 56.3) N_4=574 (86.8, 78.2)	$p<0.0006$ ESS=24.26, D=6.2 N_3=65 (36.9, 44.4) N_4=237 (87.3, 83.5)	p's<0.0001 ESS=43.36, D=6.5 N_3=237 (66.2, 54.7) N_4=568 (77.1, 84.6)	HO-CTA and EO-CTA models were identical	p's<0.0001 ESS=43.36, D=3.9 N_3=237 (66.2, 54.7) N_4=568 (77.1, 84.6)
Physician Explanation of Illness/Injury	$p<0.0001$ ESS=31.27, D=4.4 N_3=233 (48.5, 53.6) N_4=569 (82.8, 79.7)	$p<0.0028$ ESS=19.83, D=8.1 N_3=64 (39.1, 35.7) N_4=234 (80.8, 82.9)	p's<0.0001 ESS=43.54, D=3.9 N_3=233 (70.0, 52.2) N_4=564 (73.6, 85.6)	HO-CTA, EO-CTA, and GO-CTA models were identical	HO-CTA, EO-CTA, and GO-CTA models were identical

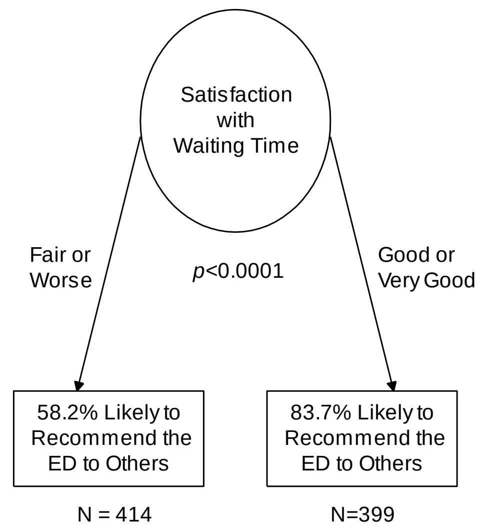
the relationship is assessed using the effect strength for sensitivity (ESS) statistic, which is normed against chance for any design, sample, and hypothesis: ESS=0 is the level of classification accuracy that is expected for the application by chance, and ESS=100 is perfect, errorless classification.[2,19,32-34] By convention, ESS<25 is considered a relatively weak effect; ESS<50 is a moderate effect; ESS<75 is a relatively strong effect; and ESS$\geq$75 is a strong effect.[2]

In the optimal data analysis (ODA) paradigm, a theoretically ideal model is defined as a perfectly accurate, maximally parsimonious model for a given application.[35,36] For any given application, the model that is closest to the theoretically ideal model (in accuracy-by-parsimony space) is selected as the best, or "globally optimal" model for the specific application.[34,35] The distance, *D*, of an empirical model from the theoretically ideal model for any given application is computed as a function of ESS and parsimony, and is presented for every model in Table 1.[37] As seen, the UniODA model for waiting time is second-closest to a theoretically ideal model versus all attributes except for physician explanation of patient illness/injury (*D*=4.5 and 4.4, respectively).

As described earlier, the UniODA model for waiting time was: if satisfaction with waiting time $\leq$3 predict likelihood to recommend ED=3; if satisfaction with waiting time >3 predict likelihood to recommend ED=4. Figure 1 presents an illustration of this UniODA model: circles represent attributes; rectangles indicate model endpoints (unique patient strata); branches show pathways through the model; values adjacent to branches are the optimal (maximum-accuracy) threshold values; values inside endpoints give the percent of class=4 observations in the strata; and numbers beneath endpoints indicate the number of observations (patients) predicted by the model to fall into all strata.[16]

Omnibus performance of the UniODA model for the total sample is summarized using a confusion matrix (Table 2).

Figure 1: UniODA Model Predicting Likelihood of Recommending ED as a Function of Satisfaction with Waiting Time



As seen, a total of (173 + 65) = 238 patients were ambivalent, and of these a total of 173 (72.7%) were correctly predicted by the UniODA model. And, a total of (241 + 334) = 575 patients were likely to recommend the ED to others, and of these a total of 334 (58.1%) were correctly predicted by the UniODA model. The percentage of observations of a given class category that are correctly predicted by the model is called the sensitivity of the model for the given class category.[2] Table 1 reports the number of observations in each class category (N_3, N_4), and the corresponding sensitivities for each class category.

Table 2: Confusion Table for the UniODA Model Predicting Likelihood of Recommending ED Based on Satisfaction with Waiting Time

		Predicted Recommendation	
		3	4
Actual	3	173	65
Recommendation	4	241	334

2155-0182/10/$3.00

Table 1 also reports the UniODA model predictive value for each class category.[2,34,38] As seen in Figure 1, the UniODA model predicted 414 observations had likelihood=3 and 173 (41.8%) were correctly classified, and the model predicted 399 observations had likelihood=4 and 334 (83.7%) were correctly classified. Note that Table 1 reports model predictive values after model sensitivities. Overall the UniODA model correctly classified 3 of 4 patients who in reality were ambivalent, and 3 of 5 patients who in reality were likely to recommend the ED. When the model predicted a patient was ambivalent it was correct for 2 of 5 patients, and when it predicted a patient was likely to recommend the ED it was correct for 7 of 8 patients.

Examination of simple bivariate models reveals that the confounding variable (waiting time) and the two attributes assessing physician explanations given to the patient yielded moderate effect strength, and that the other attributes yielded relatively weak effects. All UniODA models obtained for physician patient-care ratings yielded sensitivity greater than 75% for accurate classification of patients who in reality were likely to recommend the ED to others, and these five models achieved a predictive value greater than 75% for making accurate predictions regarding those patients who are likely to recommend the ED to others. Only the UniODA model obtained for satisfaction with waiting time yielded a sensitivity of approximately 75% for accurate classification of patients who in reality were ambivalent. Considered as a whole these results suggest that fair (or worse) waiting times reduce the likelihood of a positive recommendation, whereas good or very good physician patient-care behaviors increase the likelihood of a positive recommendation.

Analysis by Partial UniODA

Associations between the likelihood of recommending the ED to others and the five physician patient-care evaluations—with the effect of waiting time eliminated using partial UniODA, are summarized in the second column of Table 1. All models having p<0.05 had the same direction and optimal threshold identified in simple bivariate analysis. Performance for the models involving physician courtesy and physician concern for patient comfort were not statistically significant. Only the performance of the model involving physician explanation of test/treatment was statistically significant at the experimentwise criterion (Sidak p<0.05): the performance of the remaining models was statistically significant only at the generalized (per-comparison p<0.05) criterion.[2,19]

Of the five attributes, the performance of the non-confounded model for the physician's explanation of test/treatment was influenced the least by the confounder, yielding 10.9% lower sensitivity for class category 3 (ambivalent) and 0.6% greater sensitivity for class category 4 (likely to recommend the ED to others) versus the simple bivariate model. In contrast, the performance of the non-confounded model for physician courtesy was influenced the most by the confounder, yielding 75.7% greater sensitivity for class category 3 (the improved sensitivity of 46.2% was nevertheless 3.8% less than expected by chance), and 35.5% lower sensitivity for class category 4 (for this model ESS=6.7, a very weak effect). Non-confounded models for the other three attributes all had marginally lower sensitivity for class category 4, and substantially lower sensitivity for class category 3: the values respectively were 3.7% and 32.7% lower for the model of physician's concern for the patient's comfort; 1.6% and 19.4% lower for the model of physician took the patient's problem seriously; and 2.4% and 19.4% lower for the model of physician's explanation of illness/injury. For the three non-confounded models with generalized p<0.05 there was a mean reduction of 1.1% in sensitivity for patients likely to recommend the ED, and of 17.2% in sensitivity for ambivalent patients. As stated earlier, the model for waiting time best predicted ambivalence as a function of dissatisfactory waiting times, and

2155-0182/10/$3.00

when this effect was eliminated from models for the physician ratings the effect was to reduce the accuracy of the adjusted models for prediction of ambivalence.

Reduction in normed accuracy occurring for models corrected for confounding via partial UniODA is evident in the *D* statistic indicating distance of the empirical model from a theoretically ideal model.[37] Compared to the simple bivariate model, *D* for the corresponding non-confounded partial UniODA model was greater by between 21.6% (physician explanation of test/treatment) and 250.6% (physician courtesy). As seen in Table 1, for every attribute except for physician explanation of test/treatment, *D* for the partial UniODA model was the greatest distance that was reported for the attribute.

In addition to reducing normed accuracy, partial UniODA also greatly reduced the sample size available for assessing the non-confounded relationships. Even though the normed effect strength (ESS) of the confounder used to predict patient rating of likelihood to recommend the ED to others fell into the moderate range, the UniODA model for waiting time nevertheless correctly classified almost two-thirds of the observations, leaving only 37% of the original sample available for assessing non-confounded relationship of the patient likelihood rating and rated dimensions of physician behavior.

Quantitative challenges to (sub)optimal partial methods—reduction in normed accuracy and in statistical power—represent significant statistical engineering issues. However, the *qualitative* challenge to partial methods—the ecological utility of the findings—calls into question the *theoretical significance* of these methods. In the present context, for example, after eliminating the effect of waiting time, patient ratings of the physician's explanation of the patient's illness/injury predict if the patient is likely to recommend the ED (sensitivity= 82.8), but fail to predict (at a level that exceeds the accuracy expected by chance) if the patient is ambivalent in this respect (sensitivity=48.5). Pragmatically, how is this laboratory finding translated into clinical care? Real-world differences in patient satisfaction with waiting time exist, and they cannot simply be eliminated from consideration in the clinical setting. What do these inter-patient differences in satisfaction with waiting time imply for the validity of the finding that explanation of illness/injury facilitates patient recommendations of the ED to others? Viewed from a translational perspective, rather than eliminating confounding or moderating factors from the analysis, an actionable approach involves including the confounder (moderator) in the analysis, and assessing the manner in and extent to which it influences patient decision-making.

Analysis by Classification Tree Analysis

Optimal (maximum-accuracy) classification tree analysis (CTA) is the quintessential methodology for studying moderation in multi-attribute applications.[14-23,26,30,31] CTA chains successive UniODA models, creating a multi-variable system that explicitly maximizes ESS for the sample, data geometry, and hypothesis under investigation. Three modalities of optimal CTA have been developed. The first modality is known as hierarchically-optimal CTA (HO-CTA), which enters the attribute providing the greatest ESS in the model at every step of the analysis.[14,19,39-41] The second modality is known as enumerated-optimal CTA (EO-CTA), which enumerates the first three nodes of the tree model in order to obtain the model that yields maximum ESS.[16,20,40-44] The final modality is known as globally-optimal CTA (GO-CTA), which identifies the CTA model that presents the best[41] combination of accuracy and parsimony for a given sample, data geometry, and hypothesis.[23-25,30,31,35-37]

HO-CTA enables the operator to force attributes into any desired location in the tree model.[16] For example, in the present context, it is possible to force the confounding variable (waiting time) to enter the HO-CTA at the root

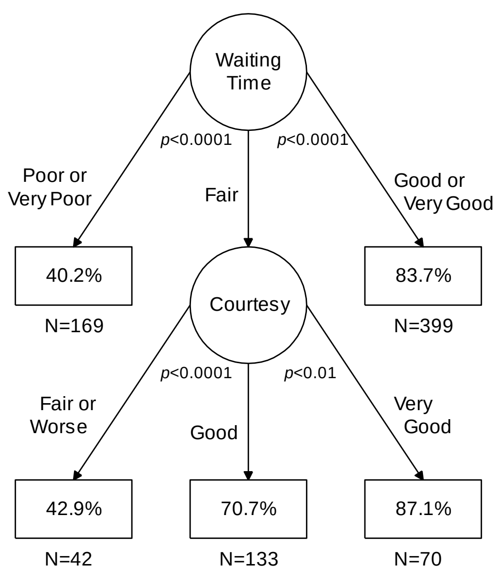
2155-0182/10/$3.00

node (the first variable in the model). This is conceptually consistent with several widely-used GLM methods, such as hierarchical linear models or stepdown ("backwards") multiple regression analysis.[3,4] As seen in Table 1, waiting time had a greater ESS than all ratings of physician patient-care behaviors except for physician explanation of illness/injury. Thus, except for the latter attribute, waiting time was the initial variable in the HO-CTA analyses. Schematic illustrations of UniODA and CTA models follow the same conventions and are subject to complementary interpretations.

Physician Courtesy

Figure 2 presents the HO-CTA model obtained by using patient ratings of the physician's concern for the patient's comfort, and satisfaction with waiting time, as attributes.

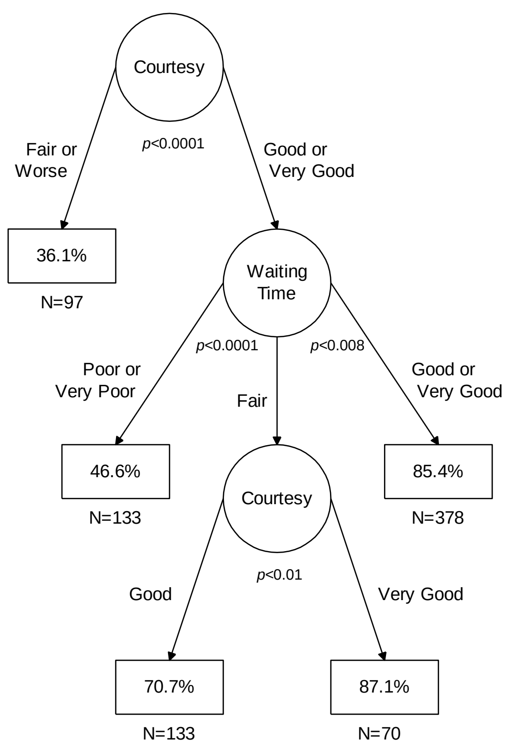
Figure 2: HO-CTA Model: Physician Courtesy



As seen, this model uses two attributes to identify five distinct patient strata. Waiting time is the root variable: 2 of 5 patients who are dissatisfied with waiting time are likely to recommend the ED to others, compared to 7 of 8 patients who are satisfied. For patients who are ambivalent about waiting time, physician courtesy is important: 2 of 5 patients ambivalent about or dissatisfied with physician courtesy are likely to recommend the ED, versus 7 of 10 patients rating physician courtesy as good, and 8 of 9 patients rating courtesy as very good (this latter difference is not statistically significant at the experimentwise criterion). Evaluated from a quantitative perspective, the *D* statistic of 8.3 for this model (Table 1) is substantially superior to the partial UniODA model (*D*=27.7), but the HO-CTA model has weak efficiency (ESS / number of strata[35]) of 7.52. The weak efficiency is in part attributable to redundant endpoint performance—the percent of (dis)satisfied patients is similar for multiple model endpoints.[41] When evaluated qualitatively the model is difficult to translate into clinical practice: an ambivalent satisfaction rating for waiting time is difficult to ascertain in a clinical setting.[28,29]

Figure 3 presents the EO-CTA model for this application.

Figure 3: EO-CTA Model: Physician Courtesy

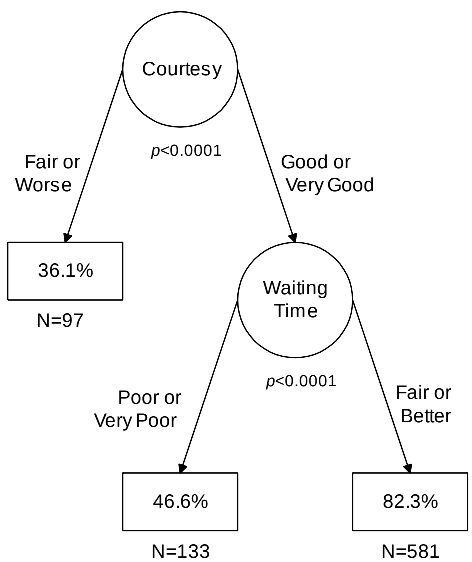


2155-0182/10/$3.00

Quantitatively this model has a marginally smaller *D* statistic (7.6) than the HO-CTA model, but it is comparably complex (five strata are identified), has comparably weak efficiency (7.93), two endpoints are redundant, and two of four Type I error rates are not statistically significant at the experimentwise criterion. Qualitatively the HO-CTA and EO-CTA models are both complex, limiting their translation into clinical practice.

Finally, Figure 4 presents the GO-CTA model for this application.

Figure 4: GO-CTA Model: Physician Courtesy

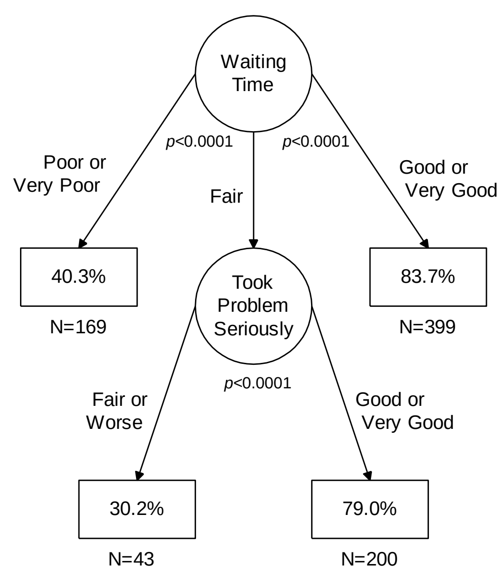


Quantitatively the GO-CTA model has a substantially smaller *D* statistic (4.6) than all other models in this application; it is the least complex of all CTA models (only three strata are identified), even though the model achieved ESS nearly as strong as achieved by the more complex EO-CTA model); it has relatively weak efficiency (13.16) which, nevertheless, is substantially stronger than was achieved by other models; and both Type I error rates are statistically significant at the experimentwise criterion. Weaknesses of the model include sensitivity for predicting ambivalence that is only marginally greater than expected by chance (see Table 1); and two endpoints are redundant. Qualitatively the model is promising. The primary factor influencing the likelihood that a patient will recommend the ED to others is the actionable physician behavior of expressing courtesy to the patient: interventions may be employed to address deficiencies and to establish a baseline level of competency in this regard.[45,46] For physician-patient interactions rated as reflecting good or very good courtesy, the secondary factor influencing a positive recommendation is waiting time, which can be ambivalent or better and still motivate a positive outcome from 7 of 8 patients. Satisfaction with waiting time may be achieved as a synergy between nurse management of patient waiting time expectations[27,28] and information systems developed to help health care workers to keep abreast of patient actual waiting times.[47]

Physician Takes Patient's Problem Seriously

Figure 5 presents the HO-CTA model obtained by using satisfaction with waiting time, and patient ratings of satisfaction with the degree to which the physician approached the patient's problem seriously, as attributes.

Figure 5: HO-CTA Model: Physician Took Patient's Problem Seriously

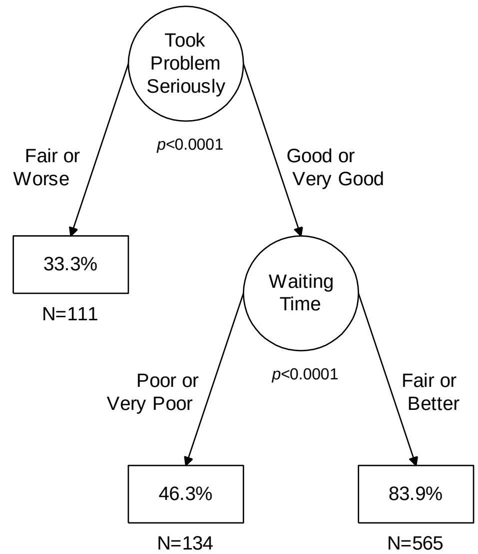


2155-0182/10/$3.00

As seen, this model identifies four patient strata. Waiting time is the root variable: 2 of 5 patients who are dissatisfied with waiting time are likely to recommend the ED to others, compared to 7 of 8 patients who are satisfied. For patients who are ambivalent about waiting time, the perceived degree to which the physician approached the patient's problem is important: 3 of 10 patients ambivalent about or dissatisfied with physician serious problem-solving are likely to recommend the ED, versus 8 of 10 patients satisfied with their perception of the physician's serious problem-solving approach (all effects were statistically significant at the experimentwise criterion). Evaluated quantitatively, the *D* statistic of 5.8 for this model (Table 1) is substantially superior to the partial UniODA model (*D*=10.0), but the HO-CTA model has relatively weak efficiency of 10.23. The weak efficiency is in part attributable to redundant endpoint performance—the percent of (dis)satisfied patients is similar for endpoints on the left-hand side and on the right-hand side of the model. Qualitatively the model is difficult to translate into clinical practice because, again, an ambivalent satisfaction rating for waiting time is difficult to ascertain in a clinical setting.

Figure 6 gives EO-CTA and GO-CTA models—which were identical in this application. These models have the same structure and optimal thresholds as the GO-CTA model for physician courtesy. The *D* statistic of 3.8 is the lowest identified in this study. Little research addresses non-serious physician treatment of a patient's problem. It is reported that cultural influences and embedded cultural implications of some diseases (e.g., sexual, mental illness, obesity, alcoholism, drug addiction) may render some physicians to be a poor match for some patients.[48] Poorly understood diseases, such as fibromyalgia[49] and hypermobility syndrome[50], can frustrate both patients and care-givers. The present finding, and the paucity of research in this area, suggests that refining this theoretical construct and its measurement are warranted.

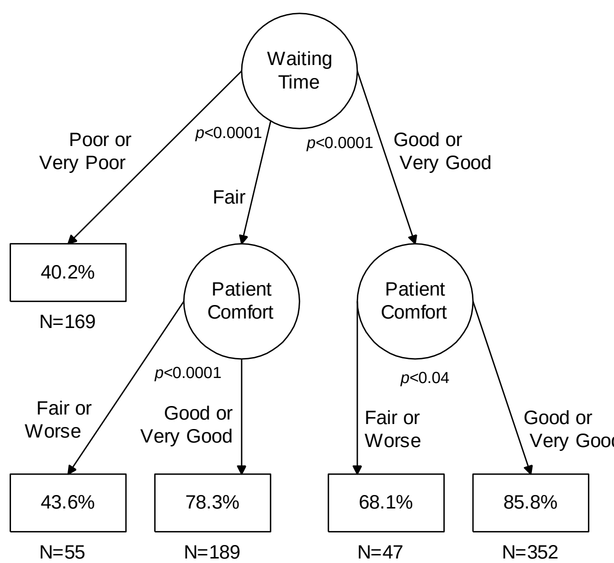
Figure 6: EO-CTA and GO-CTA Models: Physician Took Patient's Problem Seriously



Physician Concern for Patient's Comfort

Figure 7 presents the HO-CTA and EO-CTA models—identical in this application—obtained using satisfaction with waiting time, and patient ratings of satisfaction with physician concern for patient comfort, as the attributes.

Figure 7: HO-CTA and EO-CTA Models: Concern for Patient's Comfort

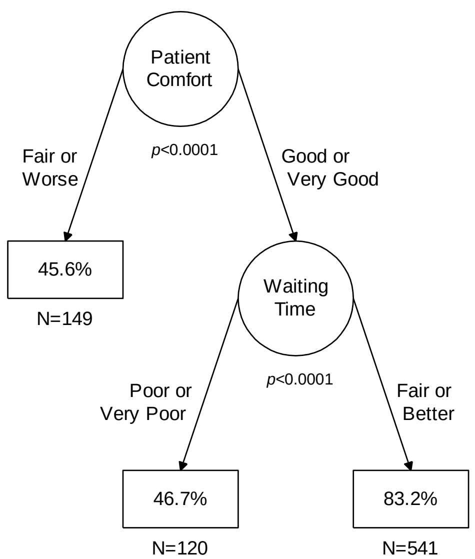


2155-0182/10/$3.00

As seen the model identifies five patient strata: the *D* statistic indicates a mediocre effect, the efficiency of the model (8.03) is weak, some endpoints are redundant, not all Type I error rates are statistically significant at the experimentwise criterion, and the model is difficult to translate into clinical practice.

The GO-CTA model obtained in this application is presented in Figure 8.

Figure 8: GO-CTA Model:
Concern for Patient's Comfort

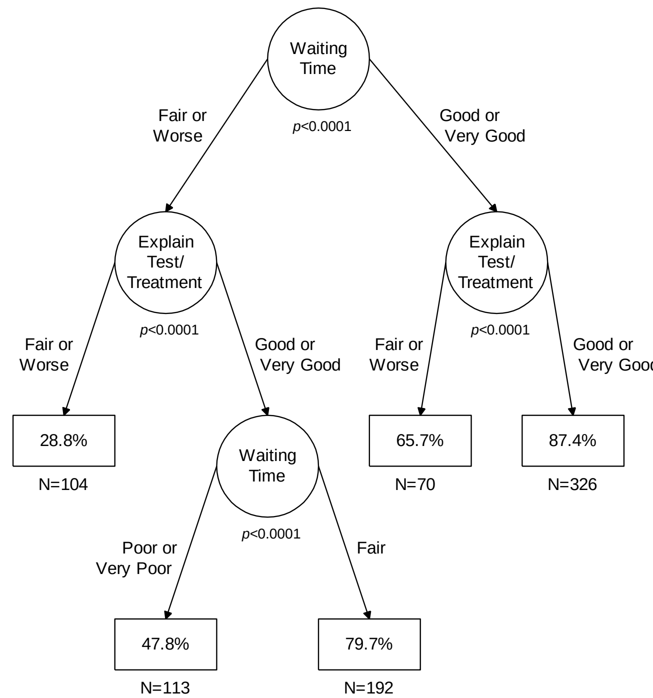


This model (also) has the same structure and optimal thresholds as the GO-CTA model for physician courtesy. The *D* statistic of 4.5 is relatively low, and it would be higher were the two left-most endpoints non-redundant and also greater than the level of classification accuracy that is expected by chance. Comfort, specifically pain management, is a well-known correlate of patient satisfaction.[51]

Physician Explanation of Test/Treatment

Figure 9 presents the HO-CTA and EO-CTA models—identical in this application—obtained using satisfaction with waiting time, and patient ratings of satisfaction with physician explanation of the test/treatment, as attributes.

Figure 9: HO-CTA and EO-CTA Models:
Explanation of Test/Treatment



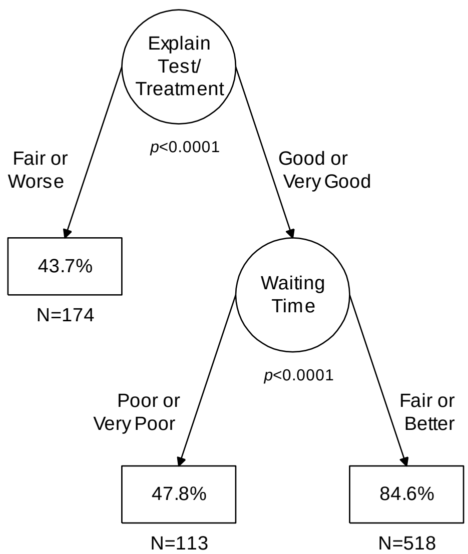
As seen the model uses four attributes (the most complex model reported presently) to identify five patient strata: the quality (*D*=8.67) and efficiency (8.67) of the model are mediocre, two endpoints are redundant, and the model is exceedingly difficult to translate into clinical practice.

The GO-CTA model obtained for this application is presented in Figure 10. Consistent with the GO-CTA model that was identified for physician serious problem-solving orientation, the *D* statistic of 3.9 is low—indicative of a powerful model. Consistent with the GO-CTA model that was identified for physician concern for patient comfort, *D* would have been higher were the two left-most endpoints non-redundant and surpassed the classification accuracy that is expected by chance. Research examining this attribute in the context of patient recommendation of the ED has not been reported, however physician "explanation" has been reported as secondary to physician interpersonal skills in

2155-0182/10/$3.00

predicting hospital recommendations of patients being treated for stroke, diabetes mellitus, Caesarean section, or appendectomy.[52]

Figure 10: GO-CTA Model: Explanation of Test/Treatment



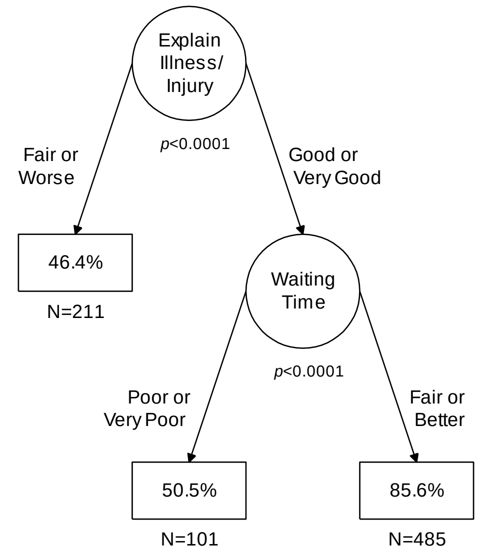
Physician Explanation of Illness/Injury

Finally, Figure 11 presents the identical HO-CTA, EO-CTA, and GO-CTA models that were obtained using satisfaction with waiting time, and patient ratings of satisfaction with physician explanation of illness/injury, as the attributes. Consistent with the GO-CTA model that was identified for physician explanation of illness/injury, the *D* statistic of 3.9 is low—indicative of a powerful model, and would have been higher were the two left-most endpoints non-redundant and if they surpassed the classification accuracy that is expected by chance.

Discussion

The consistency of all five of the GO-CTA models that were identified presently is striking—all had identical structure, including optimal threshold values. While this result is to be anticipated if the ratings of physician patient-

Figure 11: HO-CTA, EO-CTA, and GO-CTA Models: Explanation of Illness/Injury



care behaviors are strongly associated, this was not the case: while directional (confirmatory) and non-directional (exploratory) models of inter-rating association were statistically significant at the experimentwise criterion, the ESS statistics ranged between 34.1 and 63.2—that is, between moderate to relatively strong effects.[2,53] Nevertheless, the GO-CTA model obtained by using waiting time and all five of the physician patient-care behaviors as attributes was identical to the GO-CTA model obtained by using only waiting time and rating of whether the physician took the patient's problem seriously (Figure 6). In order to better understand patient recommendation of the ED, superior measures of the current constructs, and/or additional, presently unmeasured attributes are needed.

Another aspect of consistency between the five GO-CTA models is the pattern of the findings. For each model the right-most endpoint (reflecting satisfaction with physician behavior, and absence of dissatisfaction with waiting time) is strongly homogeneous—at least 4 of 5 observations in the endpoint were consistent in reporting being likely to recommend the ED to others. For each model the middle

2155-0182/10/$3.00

endpoint (reflecting satisfaction with physician behavior, and dissatisfaction with waiting time) was least homogeneous—with half of the observations reporting being likely to recommend the ED to others. The left-most endpoint ranged between moderately homogeneous (ratings of physician courtesy, and of serious problem-solving orientation) to heterogeneous (ratings of concern for patient comfort, and explanation of test/treatment and injury/illness). In a theoretically ideal classification model, all endpoints are perfectly homogeneous, and classification accuracy is perfect.[35] Clearly, therefore, the two left-most endpoints identify the patient strata for which the GO-CTA models require additional, presently unmeasured attributes, in order to achieve substantial improvement in their normed accuracy—and thereby minimize *D*.[2,54]

Substantively, the two strongest GO-CTA models identified—based on courtesy and serious problem-solving orientation—are conceptually consistent with theoretical constructs known as *expressive* and *instrumental* predispositions, respectively.[55] Popularized by the study of psychological androgyny—a "personality" typology defined as a behavioral repertoire consisting of many instrumental (concern with getting the job done) and expressive (concern for the well-being of others) capabilities—these behavioral dimensions have also been identified in research in related fields such as management (production- and employee-centered focus, respectively), leadership (initiating structure—or task completion focus, and consideration—or psychological closeness of leader and subordinate, respectively), and conflict resolution (assertiveness—concern with one's own needs, and cooperation—concern with others' needs, respectively), among others.[55] These dimensions may be measured for physicians using brief self-rating instruments.[56] Scores on instruments assessing these dimensions have been found to possess excellent psychometric (i.e., reliability) properties, and some evidence suggests that measures of these constructs are culturally cross-generalizable.[57,58] An androgynous behavioral repertoire has been shown to be related to lower reliance on technology to solve complex and difficult decision-making tasks among both new and also experienced physicians.[59,60] An androgynous predisposition in physicians has also been shown to be related to psychological empathy—reflecting cognitive understanding of (versus sympathy—reflecting an emotional reaction to) the condition of a patient.[61,62] Consistent with results obtained for androgyny, an empathic orientation has been shown to be related to lower utilization of technology when solving complex, difficult decision-making tasks among physicians.[63] Research is warranted that compares the satisfaction and quality-of-care of both patients and their attending physicians, when both members of the dyad have complementary homeostatic preferences for instrumental and expressive aspects of patient-care. In the event that training physicians to detect the preference of the patient for these independent dimensions proves to be difficult and/or unsuccessful, it is possible that an efficient pre-screening of new patients will enable administrators to assign patients and physician dyads that are primed for optimal outcomes with respect to desired and delivered patterns of patient-care.[64]

References

[1]Yarnold PR (2015). Evaluating non-confounded association of an attribute and a class variable using partial UniODA. *Optimal Data Analysis*, *4*, 32-35. URL: http://odajournal.com/2015/05/04/evaluating-non-confounded-association-of-an-attribute-and-a-class-variable-using-partial-unioda/

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

[3]Grimm LG, Yarnold PR (1995). *Reading and understanding multivariate statistics*. Washington DC: APA Books.

[4]Grimm LG, Yarnold PR (2000.). *Reading and understanding more multivariate statistics*. Washington DC: APA Books.

[5]Searle SR (1987). *Linear models for unbalanced data*, New York: Wiley, 1987.

[6]Yarnold PR, Soltysik RC (1991). Theoretical distributions of optima for univariate discrimination of random data. *Decision Sciences*, *22*, 739-752. DOI: 10.1111/j.1540-5915.1991.tb00362

[7]Soltysik RC, Yarnold PR (1994). Univariable optimal discriminant analysis: One-tailed hypotheses. *Educational and Psychological Measurement*, *54*, 646-653. DOI: 10.1177/0013164494054003007

[8]Carmony L, Yarnold PR, Naeymi-Rad F (1998). One-tailed Type I error rates for balanced two-category UniODA with a random ordered attribute. *Annals of Operations Research*, *74*, 223-238. DOI: 10.1023/A:1018922421450

[9]Yarnold PR (2015). UniODA *vs*. legacy bivariate statistical methodologies. *Optimal Data Analysis*, *4*, 73-80. URL: http://odajournal.com/2015/05/20/unioda-vs-legacy-bivariate-statistical-methodologies/

[10]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://odajournal.com/2013/09/20/maximizing-the-accuracy-of-multiple-regression-models-using-unioda-regression-away-from-the-mean/

[11]Yarnold PR (2013). Maximum-accuracy multiple regression analysis: Influence of registration on overall satisfaction ratings of emergency room patients. *Optimal Data Analysis, 2*, 72-75. URL: http://odajournal.com/2013/10/17/maximum-accuracy-multiple-regression-analysis-influence-of-registration-on-overall-satisfaction-ratings-of-emergency-room-patients/

[12]Yarnold PR (2013). Assessing technician, nurse, and doctor ratings as predictors of overall satisfaction ratings of Emergency Room patients: A maximum-accuracy multiple regression analysis. *Optimal Data Analysis*, 2, 76-85. URL: http://odajournal.com/2013/10/21/assessing-technician-nurse-and-doctor-ratings-as-predictors-of-overall-satisfaction-of-emergency-room-patients-a-maximum-accuracy-multiple-regression-analysis/

[13]Soltysik RC, Yarnold PR (2013). Statistical power of optimal discrimination with one attribute and two classes: One-tailed hypotheses. *Optimal Data Analysis, 2,* 26-30. URL: http://odajournal.com/2013/09/20/statistical-power-of-optimal-discrimination-with-a-normal-attribute-and-two-classes-one-tailed-hypotheses/

[14]Yarnold PR (1996). Discriminating geriatric and non-geriatric patients using functional status information: An example of classification tree analysis via UniODA. *Educational and Psychological Measurement, 56,* 656-667. DOI: 10.1177/0013164496056004007

[15]Yarnold PR, Soltysik RC, Bennett CL (1997). Predicting in-hospital mortality of patients with AIDS-related *Pneumocystis carinii* pneumonia: An example of hierarchically optimal classification tree analysis. *Statistics in Medicine, 16*, 1451-1463. DOI: 10.1002/(SICI)1097-0258(19970715)16:133.0.CO;2-F

[16]Soltysik RC, Yarnold PR (2010). Automated CTA software: Fundamental concepts and control commands. *Optimal Data Analysis, 1*, 144-160. URL: http://odajournal.com/2013/09/19/62/

[17]Yarnold PR (2013). Analyzing categorical attributes having many response categories. *Optimal Data Analysis*, *2*, 172-176. URL: http://odajournal.com/2013/11/08/analyzing-categorical-attributes-having-many-response-options/

[18]Yarnold PR (2013). Univariate and multivariate analysis of categorical attributes with many response categories. *Optimal Data Analysis*, 2, 177-190. URL: http://odajournal.com/2013/11/11/univariate-and-multivariate-analysis-of-categorical-attributes-with-many-response-categories/

2155-0182/10/$3.00

[19]Yarnold PR, Bryant FB (2015). Obtaining a hierarchically optimal CTA model via UniODA software. *Optimal Data Analysis*, *4*, 36-53. URL: http://odajournal.com/2015/05/11/obtaining-a-hierarchically-optimal-cta-model-via-unioda-software/

[20]Yarnold PR, Bryant FB (2015). Obtaining an enumerated CTA model via automated CTA software. *Optimal Data Analysis*, *4*, 54-60. URL: http://odajournal.com/2015/05/14/obtaining-an-enumerated-cta-model-via-automated-cta-software-2/

[21]Yarnold PR (2013). Initial use of hierarchically optimal classification tree analysis in medical research. *Optimal Data Analysis*, 2, 7-18. URL: http://odajournal.com/2013/09/20/initial-use-of-hierarchically-optimal-classification-tree-analysis-in-medical-research/

[22]Bryant FB (2010). The Loyola experience (1993-2009): Optimal data analysis in the Department of Psychology. *Optimal Data Analysis*, *1*, 4-9. URL: http://odajournal.com/2013/09/19/the-loyola-experience-1993-2009-optimal-data-analysis-in-the-department-of-psychology/

[23]Yarnold PR, Soltysik RC (2014). Globally optimal statistical classification models, II: Unrestricted class variable, two or more attributes. *Optimal Data Analysis*, *3*, 78-84. URL: http://odajournal.com/2014/08/25/globally-optimal-statistical-models-ii-unrestricted-class-variable-two-or-more-attributes/

[24]Yarnold PR (2014). What influences patients to recommend an Emergency Department to others? *Optimal Data Analysis*, *3*, 85-88. URL: http://odajournal.com/2014/08/26/what-influences-patients-to-recommend-an-emergency-department-to-others/

[25]Yarnold PR (2014). Increasing the likelihood of an ambivalent patient recommending the Emergency Department to others, *Optimal Data Analysis*, *3*, 89-91. URL: http://odajournal.com/2014/08/27/increasing-the-likelihood-of-an-ambivalent-patient-recommending-the-emergency-department-to-others/

[26]Yarnold PR, Michelson EA, Thompson DA, Adams SL (1998). Predicting patient satisfaction: A study of two emergency departments. *Journal of Behavioral Medicine*, *21*, 545-563. DOI: 10.1023/A:1018796628917

[27]Thompson DA, Yarnold PR (1995). Relating patient satisfaction to waiting time perceptions and expectations: The disconfirmation paradigm. *Academic Emergency Medicine*, *2*, 1057-1062. DOI: 10.1111/j.1553-2712.1995.tb03150.x

[28]Thompson DA, Yarnold PR, Adams SL, Spacone AB (1996). How accurate are patient's waiting time perceptions? *Annals of Emergency Medicine*, *28*, 652-656. DOI: 10.1016/S0196-0644(96)70089-6

[29]Thompson DA, Yarnold PR, Williams DR, Adams SL (1996). The effects of actual waiting time, perceived waiting time, information delivery and expressive quality on patient satisfaction in the emergency department. *Annals of Emergency Medicine*, *28*, 657-665. DOI: 10.1016/S0196-0644(96)70090-2

[30]Yarnold PR (2014). What most dissatisfies Emergency Department patients? *Optimal Data Analysis*, *3*, 92-95. URL: http://odajournal.com/2014/08/28/what-most-dissatisfies-emergency-department-patients/

[31]Yarnold PR (2014). What most satisfies Emergency Department patients? *Optimal Data Analysis*, *3*, 98-101. URL: http://odajournal.com/2014/09/01/what-most-satisfies-emergency-department-patients/

[32]Yarnold PR (2013). Minimum standards for reporting UniODA findings. *Optimal Data Analysis*, *2*, 63-68. URL: http://optimalprediction.com/files/pdf/V2A11.pdf

[33]Yarnold PR (2013). Minimum standards for reporting UniODA findings for class variables with three or more response categories. *Optimal Data Analysis*, *2*, 86-93. URL: http://odajournal.com/2013/10/22/minimum-standards-for-reporting-unioda-findings-for-class-variables-with-three-or-more-response-categories/

2155-0182/10/$3.00

[34]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://odajournal.com/2013/10/29/standards-for-reporting-unioda-findings-expanded-to-include-esp-and-all-possible-aggregated-confusion-tables/

[35]Yarnold PR, Soltysik RC (2014). Globally optimal statistical classification models, I: Binary class variable, one ordered attribute. *Optimal Data Analysis*, *3*, 55-77. URL: http://odajournal.com/2014/08/18/globally-optimal-statistical-classification-models-i-binary-class-variable-one-ordered-attribute/

[36]Yarnold PR, Soltysik RC (2014). Discrete 95% confidence intervals for ODA model- and chance-based classifications. *Optimal Data Analysis*, *3*, 110-112. URL: http://odajournal.com/2014/10/10/discrete-95-confidence-intervals-for-oda-model-and-chance-based-classifications/

[37]Yarnold PR (2015). Distance from a theoretically ideal statistical classification model defined as the number of additional equivalent effects needed to obtain perfect classification for the sample. *Optimal Data Analysis*, *4*, 81-86. URL: http://odajournal.com/2015/05/20/distance-from-a-theoretically-ideal-statistical-classification-model-defined-as-the-number-of-additional-equivalent-effects-needed-to-obtain-perfect-classification-for-the-sample/

[38]Yarnold PR, Brofft GC (2013). Comparing knot strength with UniODA. *Optimal Data Analysis*, *2*, 54-59. URL: http://odajournal.com/2013/09/20/comparing-knot-strength-using-unioda/

[39]Yarnold PR, Soltysik RC (2010). Maximizing the accuracy of classification trees by optimal pruning. *Optimal Data Analysis*, *1*, 23-29. URL: http://odajournal.com/2013/09/19/maximizing-accuracy-of-classification-trees-by-optimal-pruning/

[40]Yarnold, P.R. (2015). Selecting the minimum denominator in manual and enumerated CTA. *Optimal Data Analysis*, *4*, 14-20. URL: http://odajournal.com/2015/01/15/selecting-the-minimum-denominator-in-manual-and-enumerated-cta/

[41]Yarnold PR (2014). Illustrating how 95% confidence intervals indicate model redundancy. *Optimal Data Analysis*, *3*, 96-97. URL: http://odajournal.com/2014/08/31/illustrating-how-95-confidence-intervals-indicate-model-redundancy/

[42]Yarnold PR, Soltysik RC (2010). Manual *vs*. automated CTA: Optimal preadmission staging for inpatient mortality from *Pneumocystis cariini* pneumonia. *Optimal Data Analysis*, *1*, 50-54. URL: http://odajournal.com/2013/09/19/manual-vs-automated-cta-optimal-preadmission-staging-for-inpatient-mortality-from-pneumocystis-cariini-pneumonia/

[43]Coakley RM, Holmbeck GN, Bryant FB, Yarnold PR (2010). Manual *vs*. automated CTA: Psychosocial adaptation in young adolescents. *Optimal Data Analysis*, *1*, 55-58. URL: http://odajournal.com/2013/09/19/manual-vs-automated-cta-psychosocial-adaptation-in-young-adolescents/

[44]Yarnold PR, Bryant FB, Smith JH (2013). Manual *vs*. Automated CTA: Predicting Freshman Attrition. *Optimal Data Analysis*, *2*, 48-53. URL: http://odajournal.com/2013/09/20/manual-vs-automated-cta-predicting-freshman-attrition/

[45]Levinson W, Lesser CS, Epstein RM (2015). Developing physician communication skills for patient-centered care. *Health Affairs*, *29*, 1310-1318. DOI: 10.1377/hlthaff.2009.0450

[46]Podolsky A, Stern DT, Peccoralo L (2015). The courteous consult: A CONSULT card and training to improve resident consults. *Journal of Graduate Medical Education*, *7*, 113-117. DOI: 10.4300/JGME-D-14-00207.1

[47]Sayah A, Rogers L, Devarajan K, Kingsley-Rocker L, Lobon LF (2014). Minimizing ED waiting times and improving patient flow and experience of care. *Emergency Medicine International*, *1984*, article ID 981472. DOI: 10.1155/2014/981472

2155-0182/10/$3.00

[48]Conrad P, Barker KK (2010). The social construction of illness: Key insights and policy implications. *Journal of Health and Social Behavior*, *51*, S67-S69. DOI: 10.1177/0022146510383495

[49]Fitzcharles MA, Yunus MB (2012). The clinical concept of fibromyalgia as a changing paradigm in the past 20 years. *Pain Research and Treatment*, *2012*, article ID 184835. DOI: 10.1155/2012/184835

[50]Grahame R (2001). Time to take hyper-mobility seriously (in adults and children). *Rheumatology*, *40*, 485-487. DOI: 10.1093/rheumatology/40.5.485

[51]Welch SJ (2013). Twenty years of patient satisfaction research applied to the Emergency Department: A qualitative review. *American Journal of Medical Quality*, *25*, 64-72. DOI: 10.1177/1062860609352536

[52]Cheng SH, Yang MC, Chiang TL (2003). Patient satisfaction with and recommendation of a hospital: Effects of interpersonal and technical aspects of hospital care. *International Journal for Quality in Health Care*, *15*, 345-355. DOI: 10.1093/intqhc/mzg045

[53]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://odajournal.com/2014/04/14/how-to-assess-the-inter-method-parallel-forms-reliability-of-ratings-made-on-ordinal-scales-emergency-severity-index-version-3-and-canadian-triage-acuity-scale/

[54]Yarnold PR (2014). Triage algorithm for chest radiography for community-acquired pneumonia of Emergency Department patients: Missing data cripples research. *Optimal Data Analysis*, *3*, 102-106. URL: http://odajournal.com/2014/09/02/triage-algorithm-for-chest-radiography-for-community-acquired-pneumonia-of-emergency-department-patients-missing-data-cripples-research/

[55]Yarnold PR (1984). Note on the multidisciplinary scope of psychological androgyny theory. *Psychological Reports*, *55*, 936-938. DOI: 10.2466/pr0.1984.54.3.936

[56]Yarnold PR (1993). A brief measure of psychological androgyny for use in predicting physicians' decision making. *Academic Medicine*, *68*, 312. DOI: 10.1097/00001888-199304000-00027

[57]Yarnold PR (1994). Comparing the split-half reliability of androgyny and sex-typing measures. *Australian Journal of Psychology*, *46*, 164-169. DOI: 10.1080/00049539408259491

[58]Yarnold PR, Bryant FB, Litsas F (1989). Type A behavior and psychological androgyny among Greek college students. *European Journal of Personality*, *3*, 249-268. DOI: 10.1002/per.2410030403

[59]Yarnold PR, Nightingale SD, Curry RH, Martin GJ (1991). Psychological androgyny and preference in loss-framed gambles of medical students: Possible implications for resource utilization. *Medical Decision Making*, *11*, 176-179. DOI: 10.1177/0272989X9101100306

[60]Yarnold PR, Nightingale SD, Curry RH, Martin GJ (1990). Psychological androgyny and preference for intubation in a hypothetical case of end-stage lung disease. *Medical Decision Making*, *10*, 215-222. DOI: 10.1177/0272989X9001000309

[61]Yarnold PR, Martin GJ, Soltysik RC, Nightingale SD (1993). Androgyny predicts empathy for trainees in medicine. *Perceptual and Motor Skills*, *77*, 576-578. DOI: 10.2466/pms.1993.77.2.576

[62]Yarnold PR, Bryant FB, Nightingale SD, Martin GJ (1996). Assessing physician empathy using the Interpersonal Reactivity Index: A measurement model and cross-sectional analysis. *Psychology, Health, and Medicine*, *1*, 207-221. DOI: 10.1080/13548509608400019

2155-0182/10/$3.00

[63]Nightingale SD, Yarnold PR, Greenberg MS (1991). Sympathy, empathy, and physician resource utilization. *Journal of General Internal Medicine*, *6*, 420-423. DOI: 10.1007/BF02598163

[64]Yarnold PR (1990). Androgyny and sex-typing as continuous independent factors, and a glimpse of the future. *Multivariate Behavioral Research*, *25*, 407-419. DOI: 10.1207/s15327906mbr2503_10

Author Notes

This study involved secondary data analysis of published de-identified data and was exempt from Institutional Review Board review. The author declared no conflict of interest.

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