Optimal Data Analysis
Vol. 10 (September 20, 2021), 58-84

2155-0182/10/$3.00

Overview of the Optimal Discriminant Analysis and Novometric Paradigms

Paul R. Yarnold, Ph.D., Optimal Data Analysis, LLC
Nathaniel J. Rhodes, Pharm.D., M.Sc., Chicago College of Pharmacy, Midwestern University

and

Ariel Linden, Dr.P.H.
Linden Consulting Group, LLC

Overviews of optimal discriminant analysis (ODA) and novometric theory are presented. Discussion addresses the role of accuracy in translational and precision forecasting research, and of parsimony in theoretical research; the structure of maximum-accuracy univariable and multivariable models; underlying theoretical axioms; assessing statistical significance; identifying the descendant family of all unique statistically viable models within a sample; paradoxical confounding; precision forecasting architecture; ordered and temporal models; and future research directions. Discussion applies to all empirical research designs involving non-Hilbert space.

Empirical ODA, classification tree analysis (CTA), and novometric "maximum-accuracy" models must satisfy three fundamental rules of science. First, model predictions must be consistent with prior empirical results. Second, the most accurate model, assessed in validity analysis, is *translationally* best. Third, Occam's razor, the simplest (most parsimonious) model is *theoretically* best.

Summarizing Model Accuracy

In maximum-accuracy analysis, accuracy which is achieved by a model in training and validity analysis is summarized vis-à-vis a confusion matrix. Figure 1 presents an example of a confusion matrix hypothetically obtained for an application involving discriminating (i.e., classifying) two class categories: class 0, and class 1. Any integers, for example 3 or 145, may serve as indicators of class category status. We routinely use 0 to indicate control condition, and 1 to indicate experimental condition, or to indicate no *vs*. yes, respectively. We find this uniform coding strategy minimizes the occurrence of coding-confusion errors.

In Figure 1, *rows* of the confusion matrix indicate observations' actual class membership status. As seen, there are a total of 100 actual Class 0 observations, of which 80 were correctly classified as being from Class 0, and 20 were misclassified as being from Class 1 Thus, the *p*ercentage of *a*ccurate *c*lassification, or PAC, for Class 0 observations is 80% (this is also called model *specificity*).

2155-0182/10/$3.00

Figure 1

Summarizing Model Accuracy via the Confusion Matrix

		PREDICTED				
Class		0	1	Row Sum	Actual PAC	Chance PAC
ACTUAL	0	80	20	100	80%	50%
	1	50	50	100	50%	50%
Column Sum		130	70			
Predictive Value		61.54%	71.43%			

In Figure 1 there also are a total of 100 actual Class 1 observations of which 50 were misclassified as being from Class 0, and 50 were correctly classified as being from Class 1. Thus, for Class 1 observations, PAC=50% (this is also called model *sensitivity*).

Model accuracy in classifying members *of each class category* is called model sensitivity for class 0, sensitivity for class 1, sensitivity for class 2, etcetera. The accuracy of a model when making classifications *into class categories* are called model predictive value for class 0, model predictive value for class 1, etcetera. In contrast, epidemiology typically compares two groups (discussed further ahead), so the terms sensitivity and specificity usually suffice.

Imagine that the Class 0 observations in Figure 1 were black marbles, and that the Class 1 observations were red marbles. Also imagine that one blindly selects each marble and guesses whether it is black or red. What percentage of the black marbles are expected to be correctly classified ("guessed") on average by chance? And what percentage of the red marbles are expected to be correctly classified on average by chance? As is true if guessing whether a coin flip will yield "heads" or "tails," the average, or expected accuracy for each color is 50%. In Figure 1, this is referred to as *Chance PAC*—the percentage accuracy in classification which is expected by chance.[1,2] If there were three class categories then the expected Chance PAC for each category is 33.3%, and for ten class categories the expected Chance PAC is 10%.

In Figure 1, *columns* of the confusion matrix indicate the class membership of observations as predicted by the model. As seen, a total of 130 observations were predicted to be members of Class 0, of which 80 actual Class 0 observations were correctly classified, and 50 actual Class 1 observations were misclassified. Thus, the *p*redictive *v*alue, or PV, for Class 0 observations is 61.54% (this is called the model *negative predictive value*). In Figure 1 there also are a total of 70 observations predicted to be members of Class 1, of which 20 actual Class 0 observations were misclassified, and 50 actual Class 1 observations were correctly classified, yielding a PV for the Class 1 observations of 71.73% (this is called the model *positive predictive value*). In the ODA paradigm, researchers report the PV for Class 0, the PV for Class 1, the PV for Class 2, and so forth.

Note that whereas model *sensitivity* for Class 1 is 50% (the level of accuracy expected by chance), the *predictive value* for Class 1 is a relatively robust 71.43% owing to the comparatively smaller number of misclassified Class 0 observations. If there are *no* Class 0 misclassifications, then the predictive value of the model for Class 1 observations will equal 100% since every observation predicted to be from Class 1 will be correctly classified. Similarly, if the number of misclassified Class 1 observations is comparatively small relative to the number of correctly classified Class 0 observations, then the predictive value of the model for classifying Class 0 observations will be robust.

Additional consideration of predictive value is presented ahead in the context of precision forecasting.

2155-0182/10/$3.00

Correcting Model Accuracy for Chance

Obtaining the maximum level of predictive accuracy which is possible in validity analysis, given the available data, maximizes the effectiveness of translational research—which takes scientific discoveries made in the laboratory, clinic, or field and transforms them into new procedures that improve applied results. Maximizing model accuracy in validity analysis is also crucial for optimizing precision forecasting, which seeks reliable pathways to success (and failure) for groups of observations which meet specific characteristics defined by model strata.

In the ODA and novometric paradigms model accuracy is adjusted to *remove the effect of chance*. This is accomplished using a statistic called Effect Strength for Sensitivity or ESS.[3] Described earlier, mean PAC is mean percent accurate classification across C class categories. For an application involving a class variable having C>1 categories,

$$ESS = (Mean\ PAC - C^*)/(100 - C^*), \quad (1)$$

$$C^* = 100/C. \quad (2)$$

For the hypothetical example in Figure 1,

$$Mean\ PAC = (80+50)/2 = 130/2 = 65$$

and

$$C^* = 100/2 = 50,$$

so

$$ESS = (65-50)/50 = 15/50 = 30.$$

Using the ESS statistic, 0 is the classification accuracy expected by chance; 100 is perfect (errorless) classification accuracy; values less than 0 represent classification accuracy worse than expected by chance; and -100 is perfectly *incorrect* classification.

Achieving ESS=100 is the ultimate research objective, and applied examples which achieve accuracy that approaches or achieves perfection have been reported.[4,5] In contrast, it is extremely unlikely that perfectly *incorrect* classification will be achieved. For example, consider flipping an unbiased coin 50 consecutive times, and predicting heads or tails incorrectly on every trial: exact $p = 0.5^{50} = 8.882 \times 10^{-16}$.

Monte Carlo research solving hundreds of millions of ODA models discriminating random data was conducted via an IBM-3090/400 supercomputer. The results of these experiments were used to define a qualitative metric for the strength of a finding based on the quantitative ESS value observed.[3] This research indicated the following definitions of strength of effect, which remain in use today.

$0 \leq ESS < 25$ = Relatively Weak Effect

$25 \leq ESS < 50$ = Moderate Effect

$50 \leq ESS < 75$ = Relatively Strong Effect

$75 \leq ESS < 90$ = Strong Effect

$ESS > 90$ = Very Strong Effect

Additional consideration of model effect strength is presented later in the context of comparing exact discrete 95% confidence intervals for model *vs.* chance.

Ascertaining Statistical Significance

Extensive discussion on evaluation of statistical significance of model accuracy in the ODA paradigm is available elsewhere.[3] Thus, a brief review of this matter is discussed presently.

Thirty years ago, an analytic solution for the theoretical distribution of optima obtained using nondirectional ("two-tailed" or exploratory) univariate maximum-accuracy discriminant analysis was discovered for (im)balanced samples, under the assumption that data are random and continuous.[6] Computing the analytic solution for large samples is laborious, so the

2155-0182/10/$3.00

precision and convergence of estimated exploratory Type I error rates obtained by Monte Carlo simulation was verified.[7]

Two years after discovering the two-tailed distribution, a closed-form analytic solution for directional ("one-tailed" or confirmatory) univariate applications was discovered, for which computation is linear over sample size.[8]

Finally, four years later a formal mathematical existence proof of the theoretical confirmatory distribution was published for balanced applications.[9] This existence proof represents a breakthrough in inferential statistics because it asserts that a *p* value may be exactly computed, or precisely estimated, without requiring any distributional assumptions.

In contrast, validity of *p* values obtained by statistical (e.g., multiple regression analysis, Fisher's discriminant analysis, logistic regression analysis), and machine learning models (e.g., support vector machines, random forests), Cox survival analysis, log-linear model, causal analysis, and time series analysis) require sample data to conform to various assumptions.[10-18]

Correcting ESS for Parsimony

Identifying the most parsimonious model for a sample is important in theoretical research because it helps researchers *quantify what is known*, and *how much additional information is needed* to yield a theoretical model which is 100% accurate.

In the ODA paradigm parsimony is defined by the distance statistic (D), which is computed as:

$$D = 100/(ESS/S) - S \quad (3)$$

where S is the number of model strata (predicted endpoints in the model).[19] D is the number of additional strata having comparable mean ESS (i.e., to existing strata in the model) needed to achieve perfect classification accuracy. For a two-strata ODA model:

- If ESS=100, then D=100/(100/2)-2 = (100/50)-2 = 0, no additional endpoints are needed to achieve a perfect model;
- If ESS=75, then D=100/(75/2)-2 = (100/37.5)-2 = 0.67, so 2/3 of a stratum is needed to achieve a perfect model;
- If ESS=50, then D=100/(50/2)-2 = (100/25)-2 = 2, so 2 more strata are needed to achieve a perfect model;
- If ESS=25, then D=100/(25/2)-2 = (100/12.5)-2 = 6, so 6 more strata are needed to achieve a perfect model;
- If ESS=0, then D=100/(0/2)-2 = (100/0)-2 = ∞. This is the definition of *perfect ignorance* because nothing is known about the system when ESS=0.

Additional consideration of model parsimony is presented later in the context of discussion concerning the descendant family and measurement issues.

Structure of an ODA Model

ODA models, also called univariate ODA (or UniODA models), may be constructed for ordered as well as for categorical attributes (i.e., "independent" measures).

An example of an ODA model involving an ordered attribute (i.e., "dependent variable") is presented in Figure 2. The attribute is the age of an observation in the sample. As seen, model attributes are depicted in a circle in the diagram. The ODA algorithm identifies the age threshold (cutpoint) which most accurately discriminates class 0 *vs*. class 1 observations. Observations having an age less than or equal to the cutpoint (which is identified by the ODA algorithm) are predicted to be members of Class 0 (or Class 1, depending on the maximum-accuracy solution), and observations having an age greater than the cutpoint are predicted to be members of Class 1.

2155-0182/10/$3.00

Figure 2

ODA Model for an Ordered Attribute

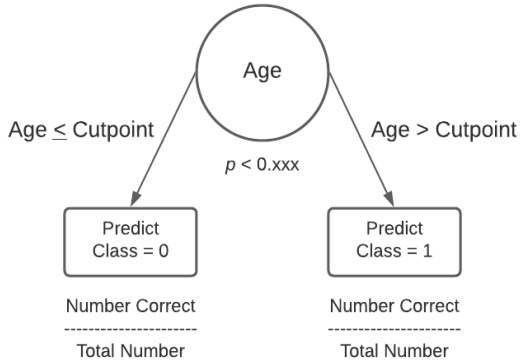
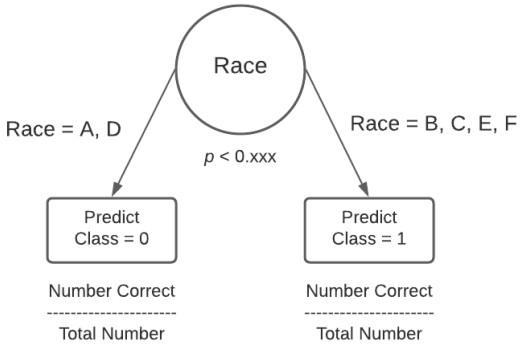


Figure 3

ODA Model for a Categorical Attribute



It should be emphasized that the cutpoint identified by ODA specifically maximizes ESS, and unless multiple cutpoints yielding identical accuracy exist, using any other cutpoint reduces model ESS. Also, it is a common practice of researchers using linear models to modify data, for example combining extreme values into a single grouped value to better fit assumptions underlying the linear model (e.g., when retaining outlier values in the data will cause a loss of linearity). This theoretically unmotivated *measurement practice* can reduce model ESS.[20]

Values used to compute predictive value are indicated beneath each endpoint: the number of correctly classified observations (numerator) and total number of correctly *and* incorrectly classified observations (denominator) are reported for each endpoint. These results are used to populate cells of the confusion matrix, and then to compute ESS and D. The (one- or two-tailed) *p* value for achieved classification accuracy is reported beneath the attribute.

An example of an ODA model involving a categorical attribute is presented in Figure 3. The attribute is the race (one of six categories, A through F) of an observation. As is done for an ordered attribute, model attributes are depicted within a circle. The ODA algorithm identifies the classification strategy *which most accurately discriminates* class 0 *vs*. class 1 observations, as measured by ESS.

In the hypothetical example, class categories A and D are predicted to be from Class 0, and the class categories B, C, E, and F are predicted to be from Class 1.

Optimal (maximum-accuracy) parsing of multicategorical attributes having more than two levels is straightforward using ODA. In contrast, this is usually impossible using linear models in which independent variables (attributes) must be ordered, or binary.[21] When confronted with a multicategorical independent variable, the linear model strategy is the use of dummy-coding to create "reference" variables.[10] In Figure 3, for example, one might select Race A as the reference category and then create five binary contrasts: A *vs*. B, A *vs*. C, A *vs*. D, A *vs*. E, and A *vs*. F. Analytically any category may be used as the reference because isomorphic results accrue regardless of which category is selected as the reference.[22] Obviously, this strategy will not suffice presently, because *combining categories A and D* yields the greatest ESS–a *data-driven conclusion*.

Note that the strategy of creating all possible combinations and permutations of grouped categories fails to *guarantee* the most *accurate* model, since linear models maximize *variance* or *likelihood*—neither of which is isomorphic with ESS.[3] This *intractable shortcoming* is one of the forms of model misspecification intrinsic to *all linear statistical methods*.

2155-0182/10/$3.00

Structure of a CTA Model

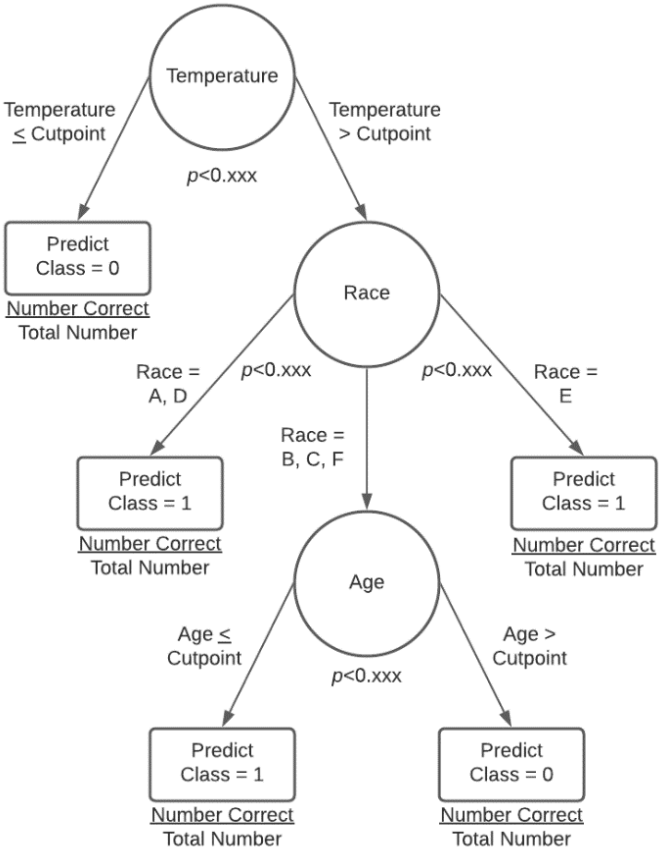
Classification tree analysis (CTA) creates a model to explicitly maximize ESS by chaining ODA models until further growth is impossible, ensuring the desired overall Type I error rate, and pruning the model to achieve maximum possible ESS.[23] Using the hierarchically optimal CTA (HO-CTA) algorithm, the attribute having maximum ESS for the entire sample is selected as the initial (i.e., root) variable for the model. Starting with the left branch emanating from the root variable, ODA is used to identify the attribute returning greatest ESS for the subsample segmented by the branch, given that all attributes in the model have experimentwise $p \leq 0.05$. This process continues until the branch cannot be further grown. This process is then repeated on the right branch emanating from the root variable. When the tree cannot be further grown, it is pruned to explicitly maximize ESS, thereby rendering the HO-CTA model.[24]

Enumerated optimal CTA (EO-CTA) grows a model in a manner similar to HO-CTA but examines many more tree models in search of that having greatest ESS. EO-CTA uses every attribute with $p \leq 0.05$ as a root variable, and for every model it enumerates all attributes having experimentwise $p \leq 0.05$ for nodes to the left and right of the root variable.[25] We favor effects that are stable in one-sample jackknife analysis, to maximize cross-generalizability of CTA models across independent samples and laboratories.[26]

Figure 4 is a hypothetical example of a CTA model (schematic depiction of HO-CTA and EO-CTA models is identical) involving two ordered attributes (temperature, degrees Fahrenheit; age, years), and one categorical attribute (race, six categories). This hypothetical model depicts a parabolic relationship between race and outcome: class 1 is predicted on the left and right flanks, and 0 is predicted in the center of the second level of the model.[27] Obviously, this structure would be difficult (perhaps impossible) to *operationalize*, and virtually impossible

Figure 4

CTA Model with Three Attributes



to *discover* using a linear model, as reference category methodology is unable to identify a parabolic multicategorical effect: an intractable mode of model misspecification error.

ODA and CTA are the statistical tools required to formulate novometric (Latin: *New Measure*) theory—analogous to quantum mechanics for a non-Hilbert space.

Axioms of Novometric Theory

Quantum mechanics explains the nature and behavior of matter and energy on the atomic and subatomic level. This theory holds that the state of a physical system can be described using a Hilbert space—a vector space having an inner product (allowing lengths and angles to be

2155-0182/10/$3.00

defined) that is complete (providing a sufficient number of limits to allow techniques of calculus to be used). The fields of physics, chemistry, and engineering are able to satisfy the requirements of a Hilbert space.

In contrast, attributable to a combination of poorly defined and imprecisely measured constructs, weak *a priori* hypotheses, and analysis of statistically unmotivated samples (which spawn Simpson's Paradox[28]), other sciences are unable to provide a Hilbert space and thus evade analysis by quantum mechanics.[21,29]

Novometric theory was developed as a means of analyzing non-Hilbert space, overcoming weaknesses of traditional statistical methods and explicitly identifying the most accurate models possible for any specific combination of sample, data, and hypothesis. Remarkably, the axioms of consequent novometric statistical theory parallel those of quantum mechanics.[30,31]

Axiom 1: For a random statistical sample S_1 consisting of a class variable, one or more attributes, and a weight: (1) all *p* values reported *in validity analysis* (see Axiom 4) must satisfy the Sidak Bonferroni-type multiple comparisons criterion for experimentwise statistical significance; and (2) for each effect which satisfies (1) the exact discrete 95% confidence intervals for *model* and for *chance* must not overlap.[32,33]

Axiom 2: In applications involving more than one attribute, the subset of attributes in S_1 which yields the globally optimal (GO) model having the lowest D statistic is identified using structural decomposition analysis (SDA), which involves iterative application of ODA in a manner conceptually analogous to principal components analysis, but that explicitly maximizes classification accuracy—instead of variance.[34]

Axiom 3: The GO model for S_1 lies in the descendant family (DF) of models obtained by applying the minimum denominator search algorithm (MDSA) to an initially-unrestricted EO-CTA model configured to predict the class variable using only the attribute(s) selected by SDA. The MDSA algorithm enables discovery of all CTA models in S_1 which originate from an unadulterated (initially unrestricted) model and vary as a function of complexity. The initial statistically motivated member of the DF is that which meets the statistical power/sample-size criterion of Axiom 1. The GO model has the smallest D statistic in the DF.[4,12]

Axiom 4: Validity is assessed using the GO model to classify an independent random S_2 (if possible), and/or for S_1 by cross-generalizability methods such as hold-out, leave-one-out (LOO) one-sample jackknife, K-of-N jackknife, split-half, multi-sample and/or bootstrap methods for static data, and using test-retest, survival (time-to-event), little jiffy, and weighted Markov methods for dynamic (repeated-measures) data.[3] These analyses enable researchers to find limits of cross-generalizability of GO models used to classify independent random samples.

Comparing Exact Discrete 95% Confidence Intervals for Model *vs*. Chance (Axiom 1)

We recently demonstrated use of an R program to obtain novometric exact discrete confidence intervals (CIs).[33] Three simulations reported in that paper are presented below to illustrate the use of this methodology (and support findings of Monte Carlo research discussed earlier) to qualitatively describe ESS in the context of the relative strength of empirical results.

The first simulation was conducted for a relatively weak model (ESS=24%). The confusion matrix for this example is seen in Table 1.

Table 1: Simulation Confusion Matrix for a Relatively Weak Effect: ESS=24%.

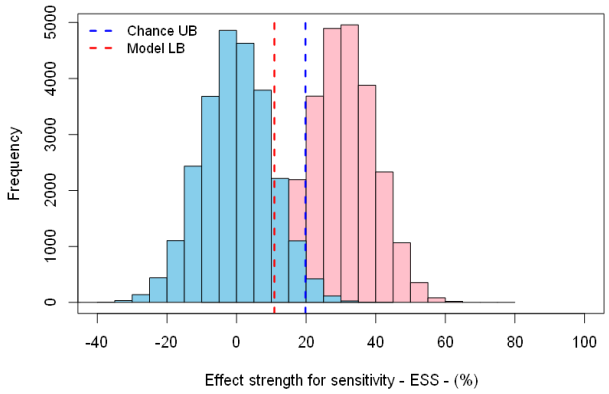
	Predicted Class	
Actual Class	class = 0	class = 1
class = 0	62	38
class = 1	38	62

Figure 5 gives the distribution of model and chance ESS obtained by random sampling with 50% replacement for 25,000 Monte Carlo

2155-0182/10/$3.00

iterations. For chance, the class variable was randomly shuffled before bootstrap resampling.

Figure 5: Bootstrap distribution of Model vs. Chance ESS for a Weak Effect (seed=1234)



The model and chance exact discrete CIs overlapped: dashed red (2.5th percentile of model) and dashed blue (97.5th percentile of chance) lines are incorrectly located relative to their respective distributions (light blue and rose, respectively). For a statistically significant effect the red dashed line should be on the same side as the rose histogram, and the blue dashed line on the same side as the blue histogram.

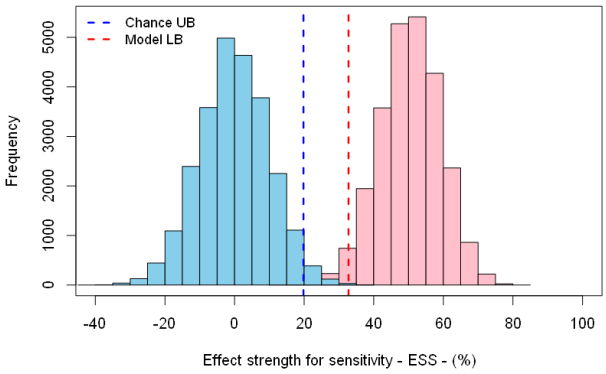
The second simulation was conducted for a moderate strength model (ESS=48%). The confusion matrix for this example is presented in Table 2.

Table 2: Simulation Confusion Matrix for a Moderate Effect: ESS=48%.

	Predicted Class	
Actual Class	class = 0	class = 1
class = 0	74	26
class = 1	26	74

Figure 6 gives the distribution of model and chance ESS obtained by random sampling with 50% replacement for 25,000 Monte Carlo iterations. As seen, for model (2.5% LB) and for chance (97.5% UB), exact, discrete CIs did not significantly overlap, and the red dashed line is on the same side as the rose histogram.

Figure 6: Bootstrap distribution of Model vs. Chance ESS for a Moderate Effect (seed=1234)



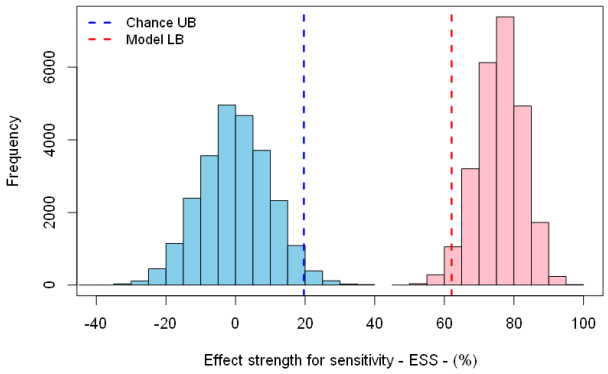
The final simulation was conducted for a relatively strong model (ESS=72%). The confusion matrix for this example is seen in Table 3.

Table 3: Simulation Confusion Matrix for a Relatively Strong Effect: ESS=72%.

	Predicted Class	
Actual Class	class = 0	class = 1
class = 0	86	4
class = 1	14	86

Figure 7 gives the distribution of model and chance ESS obtained by random sampling with 50% replacement for 25,000 Monte Carlo iterations. The exact, discrete CIs for model and for chance did not overlap.

Figure 7: Bootstrap distribution of Model vs. Chance ESS for a Relatively Strong Effect (seed=1234)



2155-0182/10/$3.00

It is important to note that the replicability of empirical findings that rely upon randomization algorithms (e.g., Monte Carlo research) are in part dependent on the seed number that is used to initiate randomization and confirmation using more than one seed is suggested.[35]

The Descendant Family (Axiom 3)

As an example of a descendant family, consider findings obtained discriminating male *vs*. female comparative cancer incidence, for data drawn from the Surveillance, Epidemiology and End Results (SEER) Program.[36] Structural decomposition analysis (SDA) was employed to find the number of statistically viable models within the sample, which vary as a function of statistical power (minimum strata size) and accuracy (ESS).[37] As is seen in Table 4, three solutions (i.e., statistically viable models) exist in this application.

The first solution identified six strata, the smallest of which had *two* observations. The exact discrete 95% CI *for the overall model* lies outside the corresponding CI for chance.[38] Note that in addition to the six strata identified in this solution, the D statistic indicates that 12.1 additional strata with the same mean ESS (a total of 18.1 strata) are needed in order to achieve a perfect model.

The second solution identified had five strata, the smallest of which had 63 observations. The exact discrete 95% CI for the overall model lies outside the corresponding CI for chance.[38] The D statistic indicates a total of 15.1 strata are needed to achieve a perfect model.

The third and final solution identified a 3-strata model. The smallest stratum had a total of 80 observations. The D statistic indicates a total of 9.4 strata are needed to achieve a perfect model. Because this solution meets the minimum power requirement and has the lowest D statistic, by definition the 3-strata model is the globally optimal (GO) solution in this example.

In summary, the 3-strata GO model has greatest statistical power vis-à-vis its minimum strata sample size of 80 observations. Model (and chance) CIs for ESS and D overlay each other across all three models in the descendant family for the sample. Therefore 3-strata GO model has the lowest D statistic, yields comparable accuracy (ESS), and is most parsimonious when compared with the alternative solutions.

Table 4: Parsing Cancer Incidence by Gender: All Sites Combined

Strata	MinD	ESS	Efficiency	D
6	2	33.2	5.54	12.1
		25.4-41.2	4.22-6.87	8.6-17.6
		0.33-7.57	0.06-1.26	73.3-1812
5	63	33.2	6.64	10.1
		25.3-41.2	5.05-8.23	7.1-14.8
		0.33-6.91	0.07-1.38	67.4-1510
3	80	31.9	10.6	6.4
		22.8-40.6	7.60-13.5	4.4-10.2
		0.33-7.57	0.11-2.52	36.6-906

Note: Results for every step of SDA analysis are tabled. Strata is number of CTA model end-points. MinD (minimum denominator) is smallest sample size for any strata. Efficiency is ESS/Strata, a normed index of relative strength of class variable(s) used to identify sample strata. For each model the first line is a point estimate; the second line is a 95% exact discrete bootstrap confidence interval (CI); and the third line is a 95% CI for chance obtained via Monte Carlo analysis (100,000 iterations).

Conceptually it is clear that combining *all cancer categories* into a single measure is an absurd idea because this practice produces a miscellany of confounded solutions—an omnipresent problem in all empirical research called *Simpson's Paradox*.[28] The result of this enigma is that findings obtained for the combined sample differ from findings for subsamples. For example, a hypothetical research question might be: how do males and females compare with respect to penile cancer? Indeed, few pure binary classes exist in non-Hilbert space.[21]

2155-0182/10/$3.00

Perhaps the most commonly reported confounded analysis in medical and social research is comparing an attribute between males and females. Fundamentally, asserting that "males" and "females" are homogeneous groups is true for a few *constants* (e.g., presence of a Y chromosome or of a uterus)—that have no place in statistical comparison of *variables*. Binary measures are used in part due to inherent limitations of popular statistical methods, and to inadequate conceptualization of theoretical phenomena.[39] Discussed later, novometric methods can conduct analysis of *ordered class variables*.

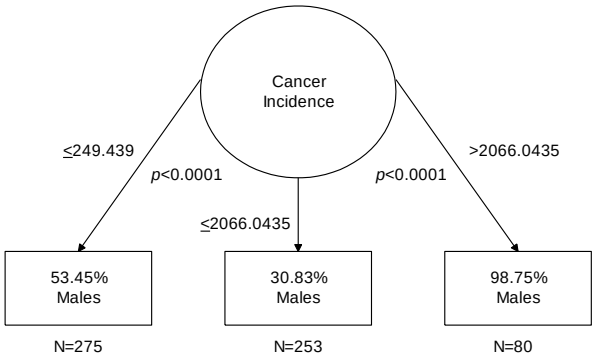
Most applied researchers have little to no expertise in measure theory, so these theoretical concepts may lack evidentiary force. However, the example used to introduce the descendant family makes the issue utterly obvious. Recall that the objective of the example was to compare males *vs*. females with respect to cancer incidence: a comparison of *two* groups. Also recall that three statistically viable solutions emerged in novometric statistical analysis. The three feasible solutions involved *six*, *five*, and *three* groups—*there was no two-group solution*, therefore no statistically motivated solution for "male *vs*. female" existed for the sample.

This point is illustrated in Figure 8. As seen, the 3-strata model identified a stratum composed approximately as half male and half female, in which observations had a low cancer incidence (i.e., less than or equal to 0.25%). Also identified was a stratum indicating moderate cancer incidence (i.e., greater than 0.25% and less than or equal to 2.07%), which was composed approximately as 70% female. The third stratum, composed 99% of males, had the highest cancer incidence of greater than 2.07%.

No legacy parametric or non-parametric statistical method is capable of identifying the set of all statistically motivated solutions which exist for a sample (i.e., the descendant family). Nevertheless, all legacy methods will readily conduct statistical comparison of "whichever groups" are entered into the computer.

Figure 8

Three-Strata CTA Model of Cancer Incidence Rate per 1,000 by Gender, All Cancer Sites



This is one aspect of the issue known as *model misspecification*—which is *not possible* when conducting novometric analysis. Because novometric analysis identifies *every solution* which varies as a function of parsimony and accuracy *that exists* in the sample, model misspecification is *literally impossible*. For example, in this analysis involving use of the (presumed) binary variable "biological sex," no model treating biological sex as a binary measure emerged in the descendant family.

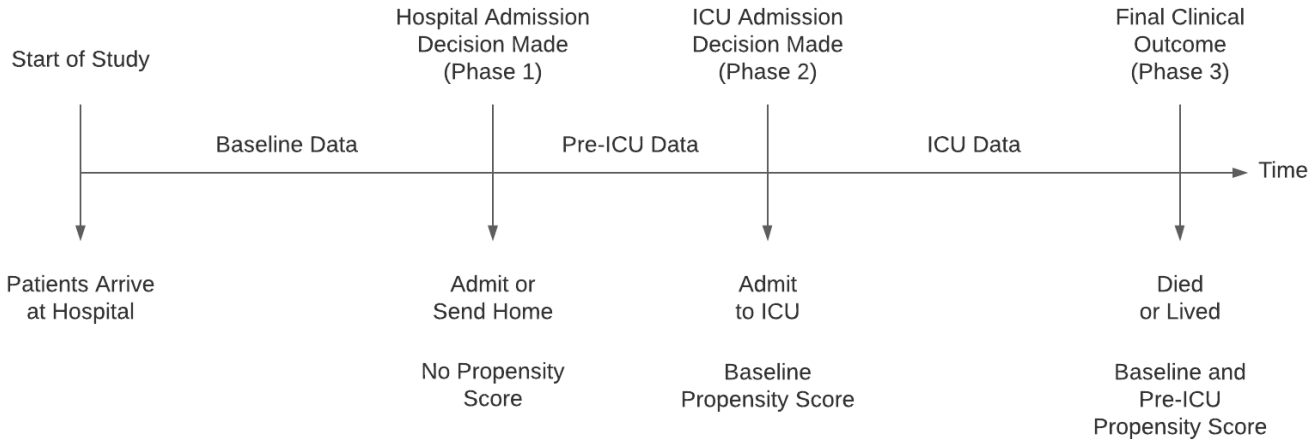
Propensity Scores

Figure 9 illustrates the workflow of a hypothetical three-phase investigation of patient hospital throughput, from intake evaluation for hospital admission (Phase 1), to the decision on admission in the intensive care unit or ICU (Phase 2), to clinical outcome and patient discharge (Phase 3). In this study, patients are either admitted to the ICU or ward, and death at discharge is compared between the two groups. Given that the patients are not randomized to either arm, this study suffers from selection bias (and likely confounding), and thus a naïve comparison of outcomes will likewise be biased. Propensity scoring techniques adjust the groups' baseline characteristics to make them appear "as good as randomized" on observed characteristics.

2155-0182/10/$3.00

Figure 9

Example of a Three-Phase Temporal Study



In Phase 1, upon arrival at the hospital the status and background of every prospective patient is investigated to enable an informed admission decision. During the intake interview a host of "baseline data" information regarding the prospective patient is acquired, including physical characteristics (e.g., weight, height, body-mass index, age, gender, blood pressure, temperature, respiration rate, partial oxygen saturation, medications, etc.), a detailed medical history, and findings of necessary medical tests. The objective of a Phase 1 statistical analysis is to discriminate cases who are admitted *vs*. are not admitted to the hospital for medical care. Novometric analysis in this design employs the minimum denominator selection algorithm (MDSA) to identify the descendant family of statistically viable CTA models.

In Phase 2, the Phase 1 data are retained for the admitted patient cohort: this is the starting point for propensity score development. The analysis begins by conducting MDSA on baseline data to identify the descendant family of propensity-score CTA models and creating subject-level propensity-score-weights to equate patients with respect to inter-patient baseline differences to mitigate potential threats to causal inference.[40,41] Data collected *after* hospital admission (pre-ICU data) are potential attributes. All patients who die before ICU admission are censored. MDSA is then used to model the propensity-score-weighted cases who are *vs*. are not admitted to the ICU in a *static* analysis, or progression of propensity-score-weighted patients admitted to the ICU over time in a *temporal* (optimal survival) analysis.[15] To avoid paradoxical contamination of propensity scores, upstream (baseline) factors must be separated from downstream (post-admission) factors that enter the causal pathway; however, complex interactions between these factors can also be evaluated as candidate propensity score weights.

Phase 3 analysis involves modelling patient survival. For the cohort of *all admitted patients* the clock starts ticking after hospital admission and it continues until death or discharge. For the cohort of patients *admitted to the ICU* the clock starts ticking after ICU admission and continues until death or discharge. In both cases baseline data are used to compute propensity-scores, and post-admission data are employed as attributes in novometric MDSA analysis modelling either static and/or temporal mortality.

2155-0182/10/$3.00

It is important to note that novometric analysis accommodates nominal, multicategorical, and/or ordered class variables and attributes. Also, the analytic course described herein is consistent with the systems engineering vernacular of input, throughput, and output.[42] Finally, the medical care example presented is generalizable to modeling other phenomena, including military confrontation, prospecting, agriculture, fire-fighting, manufacturing, fishing, athletic performance, investing, *etcetera in infinitum*.

Precision Forecasting

Factors which maximize the number of statistically viable models identified by the minimum denominator selection algorithm (MDSA) are a large sample size, and numerous attributes with moderate ESS values that are stable in validity analysis. The model with the strongest ESS has greatest translational value, whereas the globally optimal (GO) model with the lowest D statistic is most theoretically parsimonious. In addition to the highest ESS and lowest D models, novometric analysis identifies the family of all statistically viable models—known as the descendant family (DF)—that together detail all empirically discovered pathways whereby attributes define strata in which sample observations reside. The DF is useful for empirical as well as theoretical *precision forecasting*. Figure 10 is an example of output generated for one model in the DF in analysis from a study of mortality resulting from community-acquired pneumonia (CAP).[43]

The first column (ATTRIBUTE) in the CTA output indicates the code name assigned

Figure 10

CTA Output for a LOO-Stable Enumerated CTA Model Predicting Death from CAP

ATTRIBUTE	NODE	LEV	OBS	p	ESS	LOO p	MODEL
Sodium	1	1	1393	.001	15.67%	.001	<=131.5-->1,39/220,17.73%; >131.5-->0,1082/1173,92.24%
Resp Rate	2	2	217	.000	42.34%	.000	<=27.0-->0,121/131,92.37%*; >27.0-->1,29/86,33.72%*
PSI	3	2	1166	.000	51.25%	.000	<=80.5-->0,740/756,97.88%*; >80.5-->1,75/410,18.29%
Albumin	7	3	337	.000	48.42%	.000	<=2.55-->1,42/92,45.65%*; >2.55-->0,224/245,91.43%
Creatinine	15	4	172	.001	47.43%	.001	<=1.5-->1,10/80,12.50%*; >1.5-->0,91/92,98.91%*

	0	1	
0	952	177	84.32%
1	27	81	75.00%
	97.24%	31.40%	

ESS = 59.32%

D = 3.43

by the programmer for the attribute (Resp=respiration; PSI=pneumonia severity index; see the actual article[43] for final results). The second column (NODE) indicates the node of the CTA model, and the third column (LEV) indicates the level of the node in the CTA model. Node and level information defines CTA model geometry, and the remaining information in the CTA output parameterizes the model (Figure 11).

The fourth column (OBS) indicates the number of observations represented by the indicated node: as seen, there are 1,393 observations in the first ("root") node; 217 observations in the second node; 1,166 observations in the third node, etcetera. The fifth column gives the exact Type I error level (*p* value) for the attribute obtained via Monte Carlo analysis in training analysis. The sixth column indicates the ESS associated with the attribute. Column seven gives the Type I error level (*p* value) obtained via Monte Carlo analysis in leave-one-out (LOO) jackknife cross-generalizability (this example required all attributes in the model to be stable in LOO analysis). Finally, the eighth

2155-0182/10/$3.00

column gives the ODA model for the indicated attribute and node (discussed earlier, CTA models chain successive ODA models).

Six branches in Figure 10, highlighted blue and denoted by asterisks, indicate CTA model endpoints (Figure 12). Branches in Figure 10 highlighted in yellow are profiles accurately predicting the absence of death, and branches highlighted in green are profiles accurately predicting death.

Figure 11: *Node* and *Level* Define the Structure of a CTA Model

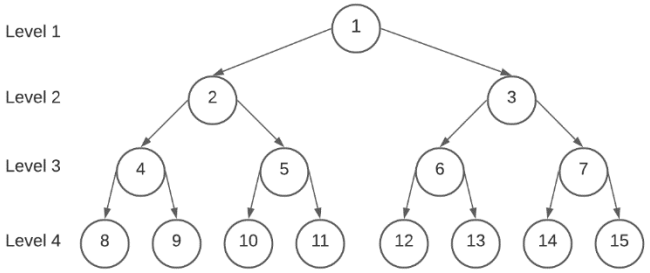
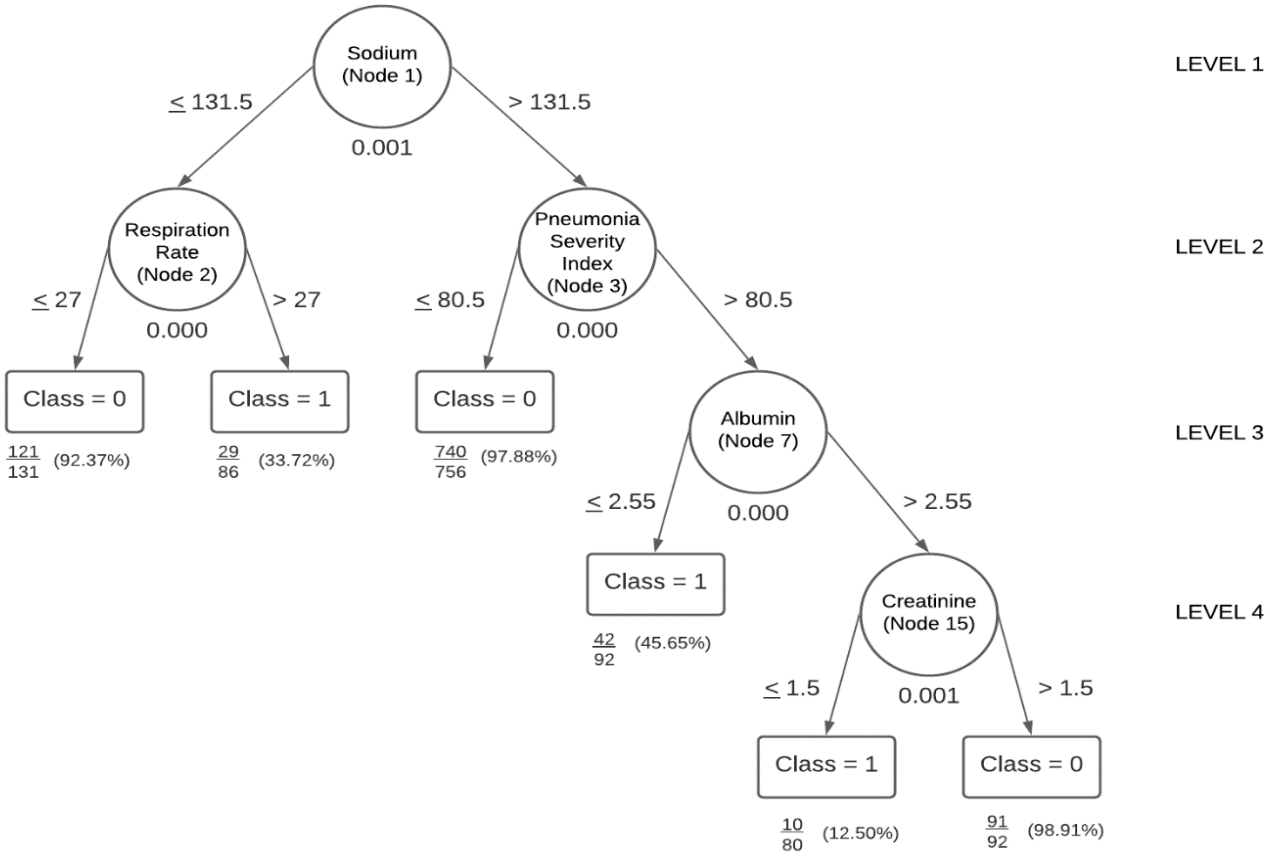


Figure 12: The CTA Model Indicated in Figure 10



All branches indicated in Figure 10 that yield accurate estimates of outcome status may be used in precision forecasting, *regardless of whether they indicate CTA model endpoints*. For example, in the first row of Figure 10, observations scoring 131.5 or less units on sodium are predicted to be from Class=1 (dead): this forms the left branch of the CTA model in Figure 12. In Figure 10, 39 of 220 observations assigned to this branch were correctly classified (17.73% predictive value). And, in the first row of Figure 10, observations scoring greater than 131.5 units

2155-0182/10/$3.00

on sodium are predicted to be from Class=0 (alive): this forms the right branch of the CTA model in Figure 12. In Figure 10, 1,082 of 1,173 observations assigned to the right branch were correctly classified (92.24% predictive value). Thus, every row in a CTA analysis output may be evaluated to obtain a collection of attribute profiles (CTA model branches) to generate a family of precision forecasting profiles. Table 5 gives precision medicine results obtained from the model in Figure 10, ranked by increasing likelihood of death among patients with CAP.

Table 5: Precision Medicine Profiles Derived from the CTA Model in Figure 10

Symptom Profile	p_{death}	
Sodium>131.5, PSI>80.5, Albumin>2.55, Creatinine>1.5	1/92	0.0109
Sodium>131.5, PSI≤80.5	16/756	0.0212
Sodium≤131.5, Respiration Rate≤27.0	10/131	0.0763
Sodium>131.5	91/1173	0.0776
Sodium>131.5, PSI>80.5, Albumin>2.55	21/245	0.0857
Sodium>131.5, PSI>80.5, Albumin>2.55, Creatinine≤1.5	10/80	0.1250
Sodium≤131.5	39/220	0.1773
Sodium>131.5, PSI>80.5	75/410	0.1829
Sodium≤131.5, Respiration Rate>27.0	29/86	0.3372
Sodium>131.5, PSI>80.5, Albumin≤2.55	42/92	0.4565

Note: CAP=Community Acquired Pneumonia. PSI=Pneumonia Severity Illness. Precision medicine profiles are ordered by increasing p_{death}.

Figures 10 and 12 present *one CTA model* in the DF for this application. Table 5 provides the precision medicine profiles derived from this single CTA model. The DF typically consists of a *family* of CTA models which exist for the sample, which vary as a function of both accuracy (ESS) and parsimony (D). Each pathway to mortality consists of a specific set of attributes, cutpoints, and directions (i.e., > or ≤ the cutpoint). Thus, the specific values obtained for a given patient on variable profiles which appear in the DF may be used to obtain multiple estimates of survival likelihood for the patient. The family of relevant constellations of outcomes defined in this way constitute "precision medicine" forecasts in the ODA vernacular. Precision estimates of success *vs*. failure may be made for any outcome, whether categorical or ordered, analyzed vis-à-vis novometric methods.

Precision medicine profiles which differ with respect to the last variable in the profile may be used to compute change in p_{death} associated with modifying the last variable. For example, for the profile {Sodium>131.5, PSI>80.5, Albumin>2,55, Creatinine≤1.5}, p_{death} is 0.1250, whereas for the profile {Sodium>131.5, PSI> 80.5, Albumin>2.55, Creatinine>1.5}, p_{death} is 0.0109. Thus, a creatinine≤1.5 has p_{death} greater by a factor of 0.1250/ 0.0109, or 11.47-fold, *vs*. a creatinine>1.5, for this profile. Similarly, for the profile {Sodium>131.5, PSI>80.5, Albumin≤2.55} p_{death} is 0.4565, and for the profile

2155-0182/10/$3.00

{Sodium>131.5, PSI>80.5, Albumin>2.55} p_{death} is 0.0857. Thus, Albumin≤2.55 has p_{death} greater by a factor of 0.4565/0.0857, or 5.33-fold, *vs*. an Albumin>2.55.

Ordered Class Variables

The partitioning algorithm employed in analysis involving a *value-ordered* (i.e., numerical) class variable creates a set of dummy codes indicating if each observation had a score equal to or less than (0), or greater than (1), for every sequential value in the class variable measurement scale. Each dummy class variable is then used to predict the attribute in a separate CTA analysis, and the model with the smallest D statistic index is selected as the globally optimal (GO) model of the relationship between the class variable and the attribute for the sample.[12]

Data for this example come from the evaluation of a primary care-based medical home pilot program that invited patients to enroll if they had a chronic illness or were predicted to have high costs in the following year. The goal of the pilot was to lower health care costs for program participants by providing intensified primary care that was intended to reduce unnecessary emergency department visits and hospitalizations.[44,45] Ordered variables were measured using integer scales.

In multiple regression analysis (MRA), age was treated as the dependent measure and modeled as a simple main-effects function of all 12 (i.e., 10 ordered, 2 categorical-binary) independent variables. The MRA model (coefficients are reported to two significant digits to the right of the decimal, when possible) was: age=42.55+7.83*treatment condition dummy variable–1.82*female gender dummy variable-3.29*admits+0.51*hospitalization days-1.40*ER visits+0.11*office visits–0.06*other procedures+0.15*laboratory tests+0.27*radiology visits–0.15*home visits+0.11*prescriptions–0.00000064*cost. This MRA model explained 19.30% of the variation in patient age: F(12, 1989)=39.6, p<0.0001.

These findings indicate that a statistically significant, ecologically modest linear relationship exists between age and one or more members of the set of 12 independent variables. The ANOVA source table for individual variables in the MRA model (sum of squares for variable-entered-last method) revealed statistically significant positive associations (with age) of assignment to treatment condition and number of prescriptions (p<0.0001 for both), and negative associations (with age) of female gender (p<0.0003), number of hospital admissions (p<0.0021), and ER visits (p<0.0149).

Finally, considered from the perspective of predictive accuracy, for this MRA model ESS was 0.43 in training analysis—a tiny effect with a 95% confidence interval which overlaps zero (the ESS expected by chance).

For novometric analysis, the age dummy codes were created on the basis of statistical power analysis that indicated a minimum of 32 patients should be classified into each model endpoint to attain 90% power to detect a moderate effect.[46] A total of 16 patients were 18 years old, and 43 were 19 years or younger, so 19 years of age was selected as the minimum age for the partitioning algorithm (PA) to satisfy the minimum power requirement. A total of 31 patients were 63-64 years old, and 95 were ≥62, so 62 years of age was selected as the maximum age for the PA. Since all of the age categories 19 to 62 inclusive were populated by data, the PA created 44 dummy age class variables.

In novometric analysis each age dummy code was treated as a class variable and all other variables were treated as attributes. The LOO-stable two-strata GO model (ESS= 41.52, D= 2.82) is seen in Figure 13. The confusion matrix for the model applied to the data is seen in Table 6. The novometric CTA model accurately predicted 3 of 4 patients 61 years or younger, and 2 of 3 patients older than 61 years. The moderate-strength LOO-stable ESS of 41.5 for the CTA model is 96.5 times greater than the training ESS for the MRA training model.

2155-0182/10/$3.00

Figure 13: Two-Strata, Globally Optimal Model

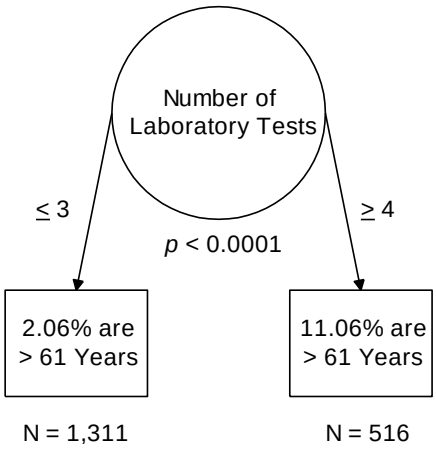


Table 6: GO Model Confusion Matrix

		Predicted Age ≤61	>61	
Actual Age	≤61	1,284	459	73.67%
	>61	27	57	67.86%
		97.94%	11.05%	

Ordered Temporal Class Variables: Interrupted Time Series

The ODA-based machine-learning partitioning algorithm used in analysis involving a *time-ordered* (i.e., numerical) class variable involves creating a set of dummy codes indicating if each observation had a score equal to or less than (0), or greater than (1) each sequential value of the class variable. Every dummy class variable is then used to predict the attribute in a separate ODA analysis, and the model with the smallest D statistic index is selected as the globally optimal (GO) model of the relationship between the class variable and the attribute for the sample.[5,18,47,48]

For example, consider the application of temporal machine-learning used to evaluate if structural breaks exist prior to the initiation of an intervention in a single-group interrupted time series analysis (ITSA), in which the outcome of interest is reported serially over time at the aggregate level (e.g., mortality rates, equity prices, interest rates, harvest productivity, etc.). This design is called an *interrupted time series* since the intervention is expected to "interrupt" the level and/or trend of the outcome variable subsequent to its introduction.[5]

Data for this example come from the evaluation of the 1988 voter-initiated Proposition 99 in California which aimed to reduce smoking rates by raising the cigarette excise tax by 25 cents per pack to fund antismoking campaigns throughout the state.[5] The aggregate outcome variable under study is per capita cigarette sales (in packs), measured annually at the state level from 1970 until 2000 (1989 is the first year of the intervention). The study limits analysis to cigarette sales in the pre-intervention years between 1970 and 1988 to determine if additional interruptions occurred in the time series prior to initiation of the intervention in 1989. Separately for each successive year, ODA finds the model maximizing accuracy achieved in discriminating each pre- *vs*. post-intervention period. For example: if cigarette sales (attribute) ≤ cutpoint predict pseudo-intervention period (class variable) is 1 (post-pseudo-intervention); if cigarette sales > cutpoint predict pseudo-intervention period is 0 (pre-pseudo-intervention).

Table 7 presents annual actual cigarette sales per capita, the ODA derived cutpoint on cigarette sales for predicting belonging to the pre- and post-pseudo-intervention periods, and reliability and accuracy measures (*p* values and ESS) for training and LOO analysis for all pre-intervention years (1970-1988). As seen, *perfect* structural breaks (i.e., ESS in LOO analysis is 100%, and experimentwise $p<0.05$) were identified for 1983 and 1985, so the intended reduction in sales of cigarettes *was already occurring before initiation* of Proposition 99 in California.

2155-0182/10/$3.00

Table 7: Year, Per Capita Sales, ODA Cutpoint, and LOO *p* and ESS Values for Optimal ITSA Models

Year	Per Capita Sales	ODA Cutpoint	*p*<	ESS
1970	123.00	122.45	----	----
1971	121.00	120.60	0.33	47.1
1972	123.50	120.60	0.18	50.0
1973	124.40	120.60	0.34	28.3
1974	126.70	120.60	0.19	37.1
1975	127.10	120.60	0.10	44.9
1976	128.00	120.60	0.04	52.4
1977	126.40	120.60	0.015	60.2
1978	126.10	120.60	0.0045	68.9
1979	121.90	120.60	0.0010*	78.9
1980	120.20	119.40	0.0002*	87.5
1981	118.60	117.00	0.0003*	85.7
1982	115.40	113.10	0.0006*	83.3
1983	110.80	107.80	0.0001*	100.0
1984	104.80	103.80	0.016	68.3
1985	102.80	101.25	0.0010*	100.0
1986	99.70	98.60	0.20	44.1
1987	97.50	93.80	----	----
1988	90.10	----	----	----

Note: ---- indicates no ODA model is possible, and * indicates experimentwise $p<0.05$

Ordered Temporal Class Variables: Time-to-Event ("Survival") Analysis

In Phase-1 research designs involving only baseline data (*viz*. Figure 9), research has shown that compared to time-to-event models derived using Cox regression analysis, models created using CTA are more accurate, parsimonious, statistically robust, and transparent.[49,50]

In Phase-2 and Phase-3 time-to-event research designs involving propensity scores created using baseline data and using subsequent data collected during treatment as potential outcome predictors, CTA-based survival models are most likely to identify a correctly-specified propensity score model, and thereby avoid the biases characteristic of traditional methods. Furthermore, as the true treatment effect is never known in observational data, CTA should be used for estimating outcomes since no statistical assumptions are required, and composite weights may be used.[15,51,52]

Multicategorical Class Variables

Stepwise ODA[53] was used to identify a CTA model to discriminate three class categories using two attributes (Figure 14). The initial step uses ODA to evaluate every possible comparison of three class categories, first treating *x* as the attribute, and then evaluating *y*. The comparison having greatest ESS in LOO cross-generalizability analysis is used as the root node of the stepwise model. Table 8 reports the findings of the initial set of exploratory hypotheses evaluated for attribute *x* and then for attribute *y*.

2155-0182/10/$3.00

Figure 14: Simulated Three-Class Data Configuration in Two-Dimensional Space
(**Blue**=Class 1; **Red**=Class 2; **Black**=Class 3)

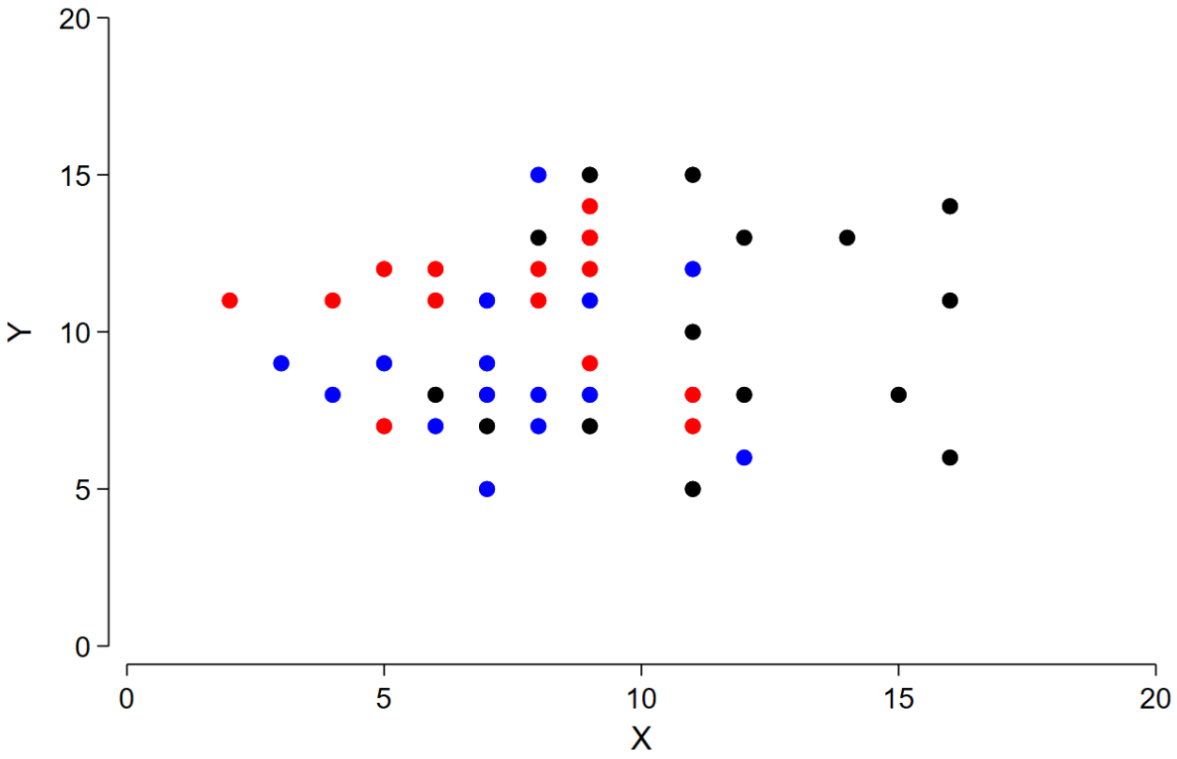


Table 8: Initial Set of Analyses

Attribute	Hypothesis	ESS	*p*<
x	1 ≠ 2 ≠ 3	36.67	0.022[a]
y	1 ≠ 2 ≠ 3	33.33	0.060
x	1 ≠ (2 = 3)	36.67	0.060[b]
y	1 ≠ (2 = 3)	36.67	0.067[b]
x	(1 = 2) ≠ 3	53.33	0.0018
y	(1 = 2) ≠ 3	26.67	0.281
x	(1 = 3) ≠ 2	26.67	0.279
y	(1 = 3) ≠ 2	36.67	0.066

[a]Leave-one-out (LOO) single case jackknife analysis is conducted when training $p<0.05$. Unless otherwise noted, LOO is stable. For this analysis, LOO ESS=20.00, $p<0.185$.
[b]The discriminant threshold for the model for x was 8.5 (stable in LOO), and 9.5 for y (not stable in LOO).

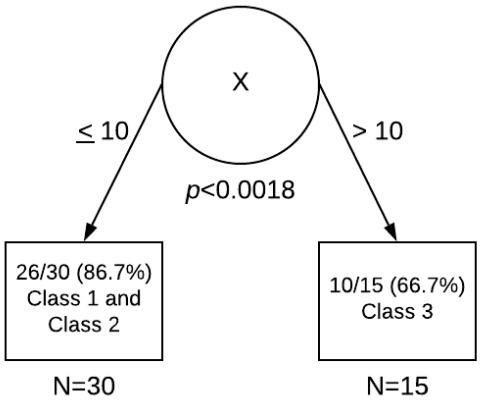
The ODA model contrasting values on x between class 3 *vs*. the combined 1 and 2 classes yielded relatively strong ESS (no model having $p<0.05$ emerged for attribute y). The model was: if x>10 then predict class=3; otherwise predict class=1 or class=2. Classification performance in training and LOO analysis is given in Table 9. The model identified in the first step is seen in Figure 15. Five misclassified observations for the right-hand endpoint renders inadequate statistical power to support further development of this branch—which thus is terminal.

Table 9: Classification Performance of Initial Analysis for Attribute x: (1 = 2) ≠ 3

		Predicted Category		
		1,2	3	*Sensitivity*
Actual	1,2	26	4	86.67
Category	3	5	10	66.67

2155-0182/10/$3.00

Figure 15: Initial Model: First Step



The second step of the analysis entails discriminating combined classes (1 and 2) on the left-hand side of the model. Table 10 summarizes findings of exploratory analysis for x, and then for y.

Table 10: Second Set of Analyses

Attribute	Hypothesis	ESS	*p*<
x	1 ≠ 2	20.00	0.721
y	1 ≠ 2	46.67	0.046

As seen, the model contrasting classes 1 and 2 on attribute y was statistically significant with moderate ESS. The model was: if y≤10 then predict class=1; otherwise predict class=2. Classification performance of this model in both *training and LOO* analysis is given in Table 11.

Table 11: Classification Performance of Second Analysis for Attribute y: 1 ≠ 2

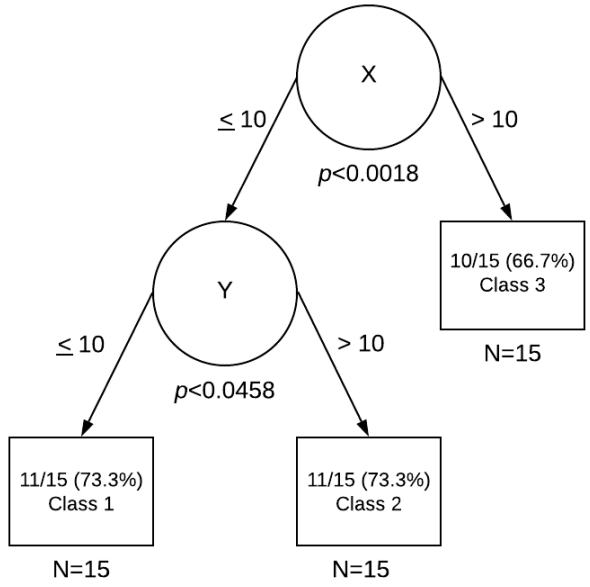
		Predicted Category		
		1	2	*Sensitivity*
Actual Category	1	11	4	73.33
	2	4	11	73.33

Further improvement in discrimination of category 1 *vs*. 2 was infeasible because of inadequate statistical power afforded by the few remaining misclassified observations.

Completing the composite model entails replacing the left-hand endpoint in Figure 15 with a model node representing variable y (Table 10). The final model discriminating all three of the class categories is seen in Figure 16.

Training and LOO classification performance of the final model calculated for three class categories reveals sensitivities of 66.67% for Class 3 (Table 9) and 73.33% for both Class 1 and Class 2 (Table 11). Mean sensitivity over classes is (66.67+73.33+73.33)/3=71.11%, so ESS=[(71.11-33.33)/(100-33.33)]x100=56.67, indicating a relatively strong LOO-stable effect.

Figure 16: Final Model: Second Step



Additional research investigating stepwise methodology using higher-dimensional class variable categories and greater numbers (and metrics) of attributes, and other optimal machine-learning approaches[17] to solving such multicategorical class-variable problems, is clearly warranted.

2155-0182/10/$3.00

Maximum-Accuracy Markov Analysis

If used to model an ordinal outcome for a single outcome variable, a Markov model may be specified using the lowest granularity ordered configuration.[54] For example, to model a serial *symptom rating* made by a single person, on each sequential trial the rating is coded as being lower (less severe symptom rating at time *i*+1 *vs*. time *i*), or unchanged/up (unchanged or more severe symptom rating at time *i*+1 *vs*. time *i*). As seen in Figure 17, "Down" (D) indicates the symptom rating was less severe ("better"), and "Not Down" (ND) indicates the symptom rating was either unchanged or more severe ("worse").

Figure 17: Least-Granular Ordered Markov Configuration for *One* Symptom Rating

	Symptom X	
Symptom X	**Time *i*+1**	
Time *i*	Down	Not Down
Down	D-D	D-ND
Not Down	ND-D	ND-ND

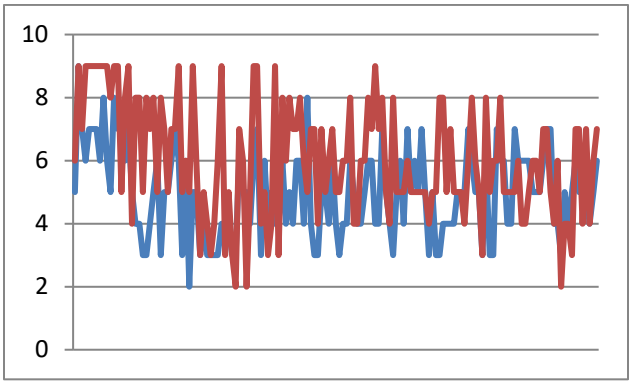
Two possible ordered model configurations are *weighted* and *unweighted*. Unweighted designs contrast relative change in rating over successive ratings—a categorical measure of *down* (worse) *vs*. *not down* (the same or better). Weighted designs weight entries in the Markov table by the absolute value of change in rating: entries with a zero weight are omitted from analysis. We used weighted Markov analysis to model the association of *two* serial ordinal ratings made by a single person (Figure 18).

Figure 18: Least-Granular Ordered Markov Configuration for *Two* Symptom Ratings

	Symptom Y	
Symptom X	**Time *i*+1**	
Time *i*	Down	Not Down
Down	D-D	D-ND
Not Down	ND-D	ND-ND

To illustrate this methodology, data were obtained via the Self-Monitoring And Review Tool (SMART), an interactive internet-based self-monitoring and feedback system that helps individuals discover and monitor links between their own health-related behaviors, management strategies, and symptom levels over time. This involves longitudinal collection: a total of 148 pain (during the day) and sleep status (during the night) ratings made by a single individual using a 10-point Likert-type scale (1=not at all bothersome; 10=extremely bothersome) were used in analysis. A plot of the symptoms over time is presented in Figure 19. No obvious pattern of association is apparent by a visual inspection of these two series.

Figure 19: Patient's Raw Pain (Blue) and Sleep Difficulty (Red) Ratings by Time



Each event in the transition table (Figure 18) was weighted by a weight calculated as the *product* of the absolute value of the difference in sleep rating between the index (t_i+1) *vs*. prior (t_i) day, and the absolute value of the difference in pain rating between the index (t_i+1) *vs*. prior (t_i) day. The performance of the resulting ODA model is indicated in Table 12. For the weighted model ESS=59.02, a relatively strong effect that is statistically significant (p<0.0001) and stable in LOO analysis (p<0.00002). Thus, for this person more severe pain experienced in the day predicts worse sleep in the night, and conversely

2155-0182/10/$3.00

less severe pain in the day predicts better sleep in the night.

Table 12: Transition Table for Change in Sleep Rating as a Function of Weighted (Change in Sleep x Change in Pain) Rating: Raw Data

	Sleep Difficulty	
Pain	**Time *i*+1**	
Time *i*	Down	Not Down
Down	27	6
Not Down	15	30

The weight which was used simply increased the metric precision of each tabular entry—from categorical to a real number. If the weight also involves adjustment via a propensity score to eliminate nuisance variables which impinge on causal inference, it is hypothesized that the weighted ESS would increase, and the weighted D statistic would decrease.

In future research we plan to use this approach to study temporal structure of other wave-like phenomena (e.g., EKG, EEG, polygraph, interest rates, market indices, weather phenomena, traffic patterns), and assess factors which affect and predict the temporal structure.

Maximum-accuracy Markov analysis has also been shown to be instrumental in understanding unidimensional phenomena over time, in a parallel manner. In particular, study of week-to-week changes in the Research Gate score of a single individual has yielded excellent results which suggest this methodology may be useful in modeling and predicting at-sea travels of submarines and fish herds.[55-57]

Finding and *Not* Finding Effects in (Random) Data

Sometimes *finding* effects—especially *strong* effects—is a *bad* outcome, for example when the effects are identified in data that are *known to be* random numbers. Whereas ODA and CTA *do not* identify models in random data, widely used popular machine learning tools for classification (e.g., random forests, boosted regression, support vector machines, multi-layer perceptron neural net, logistic discriminant analysis) *do identify* models in random data.[58-61]

Sometimes *not finding* strong effects is a *bad* outcome, such as when effects are identified in data *assumed to be* randomly generated. For example, whereas linear models *do not find* effects such as interactions that differentiate subjects in a randomized controlled trial (RCT), ODA and CTA *do identify* and *can correct* non-random effects in a RCT. [62]

Initial Success and Future Directions

As is the case in all relatively young scientific paradigms, new discoveries in the ODA and novometric paradigms occur regularly as new statistical hypotheses and data configurations are explored. As theorized by Glanville, every new methodological discovery gives rise to two or more new avenues for new research[63] though the ODA and novometric paradigms greatly surpassed Glanville's lower bound. For example:

- Since discovery of the exact statistical distribution for ODA[6] the maximum-accuracy algorithm has been employed to analyze data conventionally addressed by a plethora of legacy analyses including Bowker's test for symmetry; Bray-Curtis dissimilarity index; CART classification trees; causal mechanism analysis; CHAID classification trees; chi-square; maximum-accuracy classification trees; Cochran's Q-test; cohort tables; confidence intervals; covariate balance; cross-sectional analysis; discriminant analysis; dose-response analysis; Friedman test; generalized interactive models; incomplete multi-factor ANOVA; interval scaling; Kappa; Kendall's coefficient of concordance; Kruscal-Wallis test; little jiffy; logistic regression; logit analysis; log-linear

2155-0182/10/$3.00

model; McNemar's test; Mann-Whitney U test; Markov models; MANOVA; matching algorithms; Matlab classification trees; multicategorical analysis; multigroup ANOVA; multiple-group interrupted time series; multiple regression analysis; one-way ANOVA; pairwise comparisons; parallel-forms reliability; Pearson correlation; polychoric correlation; probit; propensity scores; randomized controlled studies; range test; recursive causal analysis; reliability analysis; ROC analysis; S-PLUS classification trees; sign test; Simpson's paradox; single-group interrupted time series; Spearman rank coefficient; structural decomposition; survival analysis; t-test; tau; test-retest reliability; time-to-event analysis; validation analysis; Wilcoxon rank-sum test; and Yule's Q.

- ODA, CTA, and novometric analysis have been used to analyze a host of applications which no legacy method can address.

- Articles using ODA and novometric analysis have been published in more than 100 indexed journals.

- Articles in the *ODA* eJournal[64] are read by scientists in 193 countries—making it the most widely distributed scientific journal in history.

- Analyses using novometric analysis are capable of suggesting crucial avenues for policy change which promise to improve outcomes research. For example, the NIH requires researchers to address sex as a biologic variable. However, as was shown when all cancers combined were compared between gender using novometric analysis, a naïve comparison that is statistically unmotivated may lead to incorrect conclusions. Thus, novometric analysis used to analyze attributes measuring theoretically salient inter-sex differences offers a pathway to understanding the biological basis of sex-based differences in disease, and to discovering new insights and mediators of these differences.

ODA researchers are conducting tantalizing research in a variety of areas, for example the development of parabolic[27] and sinusoidal models; piecewise novometric analogues to principal components analysis which identify multiple pathways to outcomes within a single sample; developing calculus for non-Hilbert space (derivatives are identified in the descendant family; integrals are computed as ESS and transformed to AUC); and using quantum computing to obtain otherwise intractable optima.

References

[1]Yarnold PR (2018). Visualizing application and summarizing accuracy of ODA models. *Optimal Data Analysis*, *7*, 85-89. https://odajournal.com/2018/11/26/visualizing-application-and-summarizing-accuracy-of-oda-models/

[2]Yarnold PR (2019). Value-added by ODA *vs*. chi-square. *Optimal Data Analysis*, *8*, 11-14. https://odajournal.com/2019/01/19/value-added-by-oda-vs-chi-square/

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

[4]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. https://odajournal.com/2016/11/11/novometric-vs-oda-reliability-analysis-vs-polychoric-correlation-with-relaxed-distributional-

2155-0182/10/$3.00

assumptions-inter-rater-reliability-of-independent-ratings-of-plant-health/

[5]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. doi: 10.1111/jep.12544

[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.x

[7]Yarnold PR, Soltysik RC (2010). Precision and convergence of Monte Carlo Estimation of two-category UniODA two-tailed *p*. *Optimal Data Analysis*, *1*, 43-45. https://odajournal.com/2013/09/19/precision-and-convergence-of-monte-carlo-estimation-of-two-category-unioda-two-tailed-p/

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

[9]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

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

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

[12]Yarnold PR, Linden A (2016). Novometric analysis with ordered class variables: The optimal alternative to linear regression analysis, *Optimal Data Analysis*, *5*, 65-73. https://odajournal.com/2016/09/19/novometric-analysis-with-ordered-class-variables-the-optimal-alternative-to-linear-regression-analysis/

[13]Yarnold PR, Hart LA, Soltysik RC (1994). Optimizing the classification performance of logistic regression and Fisher's discriminant analyses. *Educational and Psychological Measurement*, *54*, 73-85. doi: 10.1177%2F0013164494054001007

[14]Linden A, Yarnold PR (2016). Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice*, *22*, 839-847. doi: 10.1111/jep.12515

[15]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. doi: 10.1111/jep.12859

[16]Yarnold PR (2016). Novometric *vs*. log-linear model: Intergenerational occupational mobility of white American men. *Optimal Data Analysis*, *5*, 218-222. https://odajournal.com/2016/12/10/novometric-vs-log-linear-model-intergenerational-occupational-mobility-of-white-american-men/

[17]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. doi: 10.1111/jep.12610

[18]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 doi: 10.1111/jep.12966

2155-0182/10/$3.00

[19]Yarnold PR, Linden A (2016). Theoretical aspects of the D statistic. *Optimal Data Analysis*, *5*, 171-174. https://odajournal.com/2016/11/04/theoretical-aspects-of-the-d-statistic/

[20]Yarnold PR (2014). "Breaking-up" an ordinal variable can reduce model classification accuracy. *Optimal Data Analysis*, *3*, 19. https://odajournal.com/2014/03/30/breaking-up-an-ordinal-variable-can-reduce-model-classification-accuracy/

[21]Yarnold PR (2018). Minimize usage of binary measurement scales in rigorous classical research. *Optimal Data Analysis*, *7*, 3-9. https://odajournal.com/2018/03/02/minimize-usage-of-binary-measurement-scales-in-rigorous-classical-research/

[22]https://www.theanalysisfactor.com/strategies-dummy-coding/

[23]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%2F0013164496056004007

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

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

[26]Yarnold PR (2016). Determining jackknife ESS for a CTA model with chaotic instability. *Optimal Data Analysis, 5*, 11-14. https://odajournal.com/2016/05/06/determining-jackknife-ess-for-a-cta-model-with-chaotic-instability/

[27]Yarnold PR (2015). An example of nonlinear UniODA. *Optimal Data Analysis*, *4*, 124-128. https://odajournal.com/2015/07/12/an-example-of-nonlinear-unioda/

[28]Yarnold PR (1996). Characterizing and circumventing Simpson's paradox for ordered bivariate data. *Educational and Psychological Measurement, 56*, 430-442. doi: 10.1177%2F0013164496056003005

[29]Yarnold PR (2014). Increasing the validity and reproducibility of scientific findings. *Optimal Data Analysis*, *3*, 107-109. https://odajournal.com/2014/09/30/increasing-the-validity-and-reproducibility-of-scientific-findings/

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

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

[32]Yarnold PR (2018). Comparing exact discrete 95% CIs for model *vs*. chance ESS to evaluate statistical significance. *Optimal Data Analysis*, *7*, 82-84. https://odajournal.com/2018/11/26/comparing-exact-discrete-95-cis-for-model-vs-chance-ess-to-evaluate-statistical-significance/

2155-0182/10/$3.00

[33]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. https://odajournal.com/2020/06/11/generating-novometric-confidence-intervals-in-r-bootstrap-analyses-to-compare-model-and-chance-ess/

[34]Yarnold PR (2015). Modeling intended and awarded college degree via-v-vis UniODA-based structural decomposition. *Optimal Data Analysis*, *4*, 177-178. https://odajournal.com/2015/12/08/modeling-intended-and-awarded-college-degree-vis-a-vis-unioda-based-structural-decomposition/

[35]Rhodes NJ (2021). Assessing reproducibility of novometric bootstrap confidence interval analysis using multiple seed numbers (Invited). *Optimal Data Analysis*, *9*, 190-194. https://odajournal.com/2020/06/23/assessing-reproducibility-of-novometric-bootstrap-confidence-interval-analysis-using-multiple-seed-numbers-invited/

[36]Yarnold PR (2020). What is novometric data analysis? *Optimal Data Analysis*, *9*, 195-206. https://odajournal.com/2020/06/28/what-is-novometric-data-analysis/

[37]Yarnold PR (2016). How many EO-CTA models exist in my sample, and which is the best model? *Optimal Data Analysis*, *5*, 62-64. https://odajournal.com/2016/07/12/how-many-eo-cta-models-exist-in-my-sample-and-which-is-the-best-model/

[38]Yarnold PR (2020). Reformulating the first axiom of novometric theory: assessing minimum sample size in experimental design. *Optimal Data Analysis*, *9*, 7-8. https://odajournal.com/2020/01/14/reformulating-the-first-axiom-of-novometric-theory-assessing-minimum-sample-size-in-experimental-design/

[39]Yarnold PR (2019). More on: "Optimizing suboptimal classification trees: S-PLUS® propensity score model for adjusted comparison of hospitalized *vs*. ambulatory patients with community-acquired pneumonia". *Optimal Data Analysis*, *8*, 56-59. https://odajournal.com/2019/02/12/more-on-optimizing-suboptimal-classification-trees-s-plus-propensity-score-model-for-adjusted-comparison-of-hospitalized-vs-ambulatory-patients-with-community-acquired-pneumonia/

[40]Linden A, Yarnold PR (2017). Using classification tree analysis to generate propensity score weights. *Journal of Evaluation in Clinical Practice*, *23*, 703-712. doi: 10.1111/jep.12744

[41]Yarnold PR, Linden A (2017). Computing propensity score weights for CTA models involving perfectly predicted endpoints. *Optimal Data Analysis*, *6*, 43-46. https://odajournal.com/2017/06/08/1808/

[42]Katz D, Kahn RL (1966). *The social psychology of organizations*. John Wiley and Sons, New York.

[43]Arozullah AM, Parada J, Bennett CL, Deloria-Knoll M, Chmiel JS, Phan L, Yarnold PR (2003). A rapid staging system for predicting mortality from HIV-associated community-acquired pneumonia. *Chest*, *123*, 1151-1160. doi: 10.1378/chest.123.4.1151

[44]Linden A, Adams JL (2012). Combining the regression-discontinuity design and propensity-score based weighting to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, *18*, 317-325. doi: https://doi.org/10.1111/j.1365-2753.2011.01768.x

[45]Linden A (2011). Identifying spin in health management evaluations. *Journal of Evaluation in Clinical Practice*, *17*, 1223-1230. doi: 10.1111/j.1365-2753.2010.01611.x

2155-0182/10/$3.00

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

[47]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. https://odajournal.com/2020/03/07/implementing-oda-from-within-stata-identifying-structural-breaks-in-single-group-interrupted-time-series-designs/

[48]Linden A, Yarnold PR (2018). The Australian Gun Buyback Program and rate of suicide by Firearm. *Optimal Data Analysis, 7*, 28-35. https://odajournal.com/2018/03/25/the-australian-gun-buyback-program-and-rate-of-suicide-by-firearm/

[49]Linden A, Yarnold PR (2017). Modeling time-to-event (survival) data using classification tree analysis. *Journal of Evaluation in Clinical Practice, 23*, 1299-1308. doi: 10.1111/jep.12779

[50]Linden A (2020). Implementing CTA from Within Stata: Modeling Time-to-Event (Survival) Data (Invited). *Optimal Data Analysis, 9*, 79-83. https://odajournal.com/2020/04/05/implementing-cta-from-within-stata-modeling-time%e2%80%90to%e2%80%90event-survival-data-invited/

[51]Yarnold PR, Rhodes, NJ, Linden A (2020). Selecting an appropriate weighting strategy in maximum-accuracy time-to-event (survival) analysis. *Optimal Data Analysis, 9*, 3-6. https://odajournal.com/2020/01/13/selecting-an-appropriate-weighting-strategy-in-maximum-accuracy-time-to-event-survival-analysis/

[52]Rhodes NJ, Jozefczyk C, Moore W, Yarnold PR, Harkabuz K, Maxwell R, Sutton S, Silkaitis C, Qi C, Wunderink R, Zembower T (2021). Characterizing risk factors for *Clostridioides difficile* infection among hospitalized patients with community-acquired pneumonia. *Antimicrobial Agents and Chemotherapy, 65*, 1-13.

[53]Yarnold PR, Linden A (2019). Novometric stepwise CTA analysis discriminating three class categories using two ordered attributes. *Optimal Data Analysis, 8*, 68-71. https://odajournal.com/2019/04/06/novometric-stepwise-cta-analysis-discriminating-three-class-categories-using-two-ordered-attributes/

[54]Yarnold PR (2019). Optimal Markov model relating *two* time-lagged outcomes. *Optimal Data Analysis, 8*, 61-63. https://odajournal.com/2019/02/18/optimal-markov-model-relating-two-time-lagged-outcomes/

[55]Yarnold PR (2020). Modeling an individual's weekly change in RG-score via novometric single-case analysis. *Optimal Data Analysis, 9*, 114-121. https://odajournal.com/2020/04/20/modeling-an-individuals-weekly-change-in-rg-score-via-novometric-single-case-analysis/

[56]Yarnold PR (2020). Optimal weighted Markov analysis: Modeling weekly RG-score. *Optimal Data Analysis, 9*, 122-123. https://odajournal.com/2020/04/23/optimal-weighted-markov-analysis-modeling-weekly-rg-score/

[57]Yarnold PR (2020). Optimal weighted Markov analysis: Predicting serial RG-score. *Optimal Data Analysis, 9*, 124-127. https://odajournal.com/2020/04/25/optimal-weighted-markov-analysis-predicting-serial-rg-score/

[58]Linden A, Yarnold PR (2019). Some machine learning algorithms find relationships between variables when none exist -- CTA doesn't. *Optimal Data Analysis, 8*, 64-67.

2155-0182/10/$3.00

[59]Linden A, Yarnold PR (2019). Effect of sample size on discovery of relationships in random data by classification algorithms. *Optimal Data Analysis*, *8*, 76-80.

[60]Linden A, Yarnold PR (2019). Multi-layer perceptron neural net model identifies effect in random data. *Optimal Data Analysis*, *8*, 94-96.

[61]Linden A, Bryant FB, Yarnold PR (2019). Logistic discriminant analysis and structural equation modeling both identify effects in random data. *Optimal Data Analysis*, *8*, 97-102.

[62]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. doi: 10.1111/jep.12792

[63]Glanville R (1982). Inside every white box there are two black boxes trying to get out. *Behavioral Science*, *27*, 1-11. doi: 10.1002/bs.3830270102

[64]https://ODAJournal.com

Author Notes

No conflicts of interest were reported.