Optimal Data Analysis
Vol. 2, Release 2 (November 7, 2013), 159-171

2155-0182/10/$3.00

Ascertaining an Individual Patient's *Symptom Dominance Hierarchy*: Analysis of Raw Longitudinal Data Induces Simpson's Paradox

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

Consider an application with a *single individual* assessed on an attribute on a longitudinal series of measurement (testing) sessions: measurement 1, measurement 2, and so forth, ending at measurement *N*. Each measurement records the value of an attribute: the value recorded at the initial measurement is denoted as A_1, the value recorded at the *i*th measurement is denoted as A_i, and the value of the attribute recorded at the final measurement is denoted as A_N. Examples of such ordered series include the monthly closing price of a corporate stock, a dieter's weight each Friday morning upon waking, or real-time heart rate of an ICU patient recorded minute-by-minute. The present study examined a single-case series for a patient with fibromyalgia, consisting of 297 sequential daily ratings of nine symptoms commonly associated with the disease. Symptoms were compared statistically by UniODA to identify their dominance hierarchy (most to least severe) for this patient. Analysis of raw data suggested that fatigue was the dominant symptom, but analysis of ipsatively standardized data indicated that finding was paradoxically confounded: the primary symptom, in fact, was stiffness. Of 36 comparisons of symptom pairs performed using raw data, 28 (78%) were found to be confounded, ten of which (28%) identified the opposite effect (what raw data analysis suggested occurred was, in fact, the opposite of what occurred). These findings reveal a crucial difference between the meaning of the numbers and labels constituting the response scale as perceived and interpreted by the *investigator*, and the meaning of the numbers and labels as perceived and applied by the *individual* to quantify internal cognitive and emotional experiences. The latter issue is the *purpose* of the scale (if subject perception and response scale did *not* interact, then valid measurements would be impossible to obtain) and *motivation* underlying the research. Present findings show raw score analysis addresses the investigator's perspective, and ipsative *z*-score analysis addresses the patient's.

2155-0182/10/$3.00

Recent research investigating single-case serial designs studied the trajectory of such series over sequential measurements. One study found that paradoxical confounding attributable due to global warming reduced the accuracy of weather prediction[1], and a second study noted significant rises of annual crude mortality rate in toxicity-riddled environments.[2] UniODA-based methods were developed to evaluate both *post hoc* and *a priori* hypotheses within single-case series, and also for identifying the most recent, earliest, and *K*th statistically significant change within an ordered single-case series.[3,4]

This investigation is the first to report an integrated substantive application of UniODA used to compare different attributes in a single-case series. Data are a series of 297 sequential daily ratings of nine symptoms associated with fibromyalgia (FM), that were made by a patient with FM using an 11-point Likert-type response scale (0="not at all bothersome", 10="extremely bothersome".[5,6] The objective of the analysis is to identify the *symptom dominance hierarchy*—that is, to order the nine symptoms from least to most severe—for this patient.

The induction of Simpson's paradox has been characterized for sample-based designs.[7,8] Research demonstrating serial *N*-of-1 UniODA-based analyses was first to report that such confounding also arises in single-case series with attributes having ordered[1,9] or dichotomous[10] response scales. Thus, analysis herein is initially conducted using raw symptom ratings, and then using ipsatively standardized *z*-score ratings.[11]

Distributions and descriptive statistics for the raw rating data for this patient's series of 297 entries are presented in Table 1 and Table 2, respectively.

Table 1: Raw Score Distributions for Nine Symptoms: *Investigator's* Perspective

Rating	Pain	Stiffness	Fatigue	Concentration	Memory	Anxiety	Depression	GI	Sleep
0						74	97	233	1
1		1		2		173	85	44	4
2	26	11		18		32	43	8	12
3	48	25	5	24	23	11	37	3	61
4	74	56	25	36	48	6	22	5	85
5	65	66	50	69	69	1	8	1	54
6	49	71	85	69	80		2	2	40
7	27	42	60	50	51		2	1	22
8	7	18	49	22	22		1		18
9	1	7	23	7	4				
10									

Note: Tabled are frequencies. GI=gastrointestinal.

Eyeball examination of Table 1 suggests the patient rated different symptoms using different ranges on the response scale. The anxiety, depression, and gastrointestinal rating distributions are positively skewed (Table 2) and reminiscent of a Poisson distribution. Symptoms with relatively many higher ratings (indicative of increasingly severe symptoms) include stiffness, fatigue, and memory issues, and sleep and concentration issues to a lesser extent.

2155-0182/10/$3.00

Table 2: Raw Score Descriptive Statistics for Nine Symptoms: *Investigator's* Perspective

Symptom	Mean	SD	Median	Skewness	Kurtosis	CV	SEM
Pain	4.60	1.52	5	0.20	-0.53	33	0.088
Stiffness	5.32	1.60	5	-0.01	-0.30	30	0.093
Fatigue	6.38	1.44	6	-0.03	-0.59	23	0.084
Concentration	5.39	1.70	5	-0.25	-0.33	32	0.099
Memory	5.57	1.41	6	0.05	-0.56	25	0.082
Anxiety	1.01	0.86	1	1.42	3.32	86	0.050
Depression	1.49	1.55	1	1.17	1.26	104	0.090
Gastrointestinal	0.39	1.03	0	4.34	24.1	267	0.060
Sleep	5.57	1.62	5	0.28	-0.03	29	0.094

Note: n=297. SD=standard deviation; CV=coefficient of variation; SEM=standard error of the mean.

The results of all possible comparisons[9] conducted between pairs of attributes using Uni-ODA[12] are summarized in Table 3: 100,000 Monte Carlo experiments were used to estimate Type I error (p) in each analysis, and a sequentially-rejective Sidak Bonferroni criterion was used to ensure experimentwise $p<0.05$.[4,12,13]

As seen, 30 of 36 pairwise comparisons met the criterion for experimentwise $p<0.05$: 16.7 times more than are expected by chance. Of these 30 statistically significant effects, 19 (63%) were strong/very strong: strongest effects were obtained for comparisons with anxiety, depression, and GI issues.

A symptom dominance hierarchy (SDH) is constructed by mapping all 36 of the pairwise comparisons.[12] Since the SDH is a line the mapping is like solving a 36-piece puzzle with two "ends" rather than four "corners." It is a sound strategy when solving this linear puzzle to look in the pairwise comparison table (Table 3) for attributes associated with the strongest effects (greatest pairwise differences) all in the same direction (the attribute is always on one side of the inequality). In Table 3 Fatigue ratings are significantly greater than all other symptoms so Fatigue forms the high (most severe) end of the SDH (solution of this puzzle is simplified by marking-out all "explained" entries of the pairwise comparisons table using a printed copy). At this point the SDH is mapped as shown.

```
LOW ----------------------------------|-- HIGH
                                      Fa
```

In the pairwise comparisons table with fatigue-related entries eliminated, GI ratings are significantly lower than the other ratings, so GI forms the low (least severe) end of the SDH, now mapped as shown.

```
LOW --|---------------------------------|-- HIGH
      GI                                Fa
```

In the pairwise comparisons table with fatigue- and GI-related entries eliminated, the anxiety ratings are significantly lower than other ratings: the SDH is now mapped as shown.

```
LOW --|-----|--------------------------|-- HIGH
      GI   An                          Fa
```

Depression is next to follow this pattern: the SDH is now mapped as shown.

```
LOW --|-----|-----|--------------------|-- HIGH
      GI   An  De                      Fa
```

2155-0182/10/$3.00

Table 3: Raw Score Summary of UniODA Comparisons of All Pairs of Patient Symptoms: *Investigator's* Perspective

	Stiffness	Fatigue	Concentration	Memory	Anxiety	Depression	Gastrointestinal	Sleep
Pain	St > Pa 18.5, 19.2	Fa > Pa 44.8, 44.8	Co > Pa 22.9, 24.2	Me > Pa 25.9, 27.8	Pa > An 85.2, 85.2	Pa > De 67.0, 68.6	Pa > Ga 93.3, 93.7	Sl > Pa 23.6, 25.0
Stiffness		Fa > St 26.6, 27.6	St = Co 4.4, 5.3	St = Me 7.4, 9.3	St > An 89.9, 89.9	St > De 75.8, 75.8	St > Ga 92.9, 93.3	St = Sl 6.7, 20.4
Fatigue			Fa > Co 23.2, 24.5	Fa > Me 20.2, 21.7	Fa > An 96.0, 96.0	Fa > De 86.5, 87.4	Fa > Ga 96.0, 96.1	Fa > Sl 28.0, 28.9
Concentration				Co = Me 7.1, 17.7	Co > An 87.2, 87.2	Co > De 73.4, 73.5	Co > Ga 92.6, 92.9	Co = Sl 9.1, 24.7
Memory					Me > An 93.9, 94.3	Me > De 80.5, 80.6	Me > Ga 96.0, 96.1	Me = Sl 7.7, 7.8
Anxiety						De > An 21.9, 27.3	An > Ga 53.5, 53.6	Sl > An 92.3, 92.4
Depression							De > Ga 45.8, 46.4	Sl > De 82.5, 82.8
Gastrointestinal								Sl > Ga 94.3, 94.3

Note: All effects for which an *inequality* is provided have $p<0.05$ at the experimentwise criterion; effects indicated with an *equality* (=) are *not* statistically significant. ESS and ESP values (the left and right entries in second row of every table cell, respectively) shaded in grey indicate weak effects; green indicates moderate effects; red indicates strong effects; and blue indicates very strong effects.[12]

2155-0182/10/$3.00

Pain is next to follow this pattern: the SDH is now mapped as shown.

```
LOW --|-----|-----|-----|----------------|-- HIGH
        GI  An   De   Pa               Fa
```

Remaining symptom ratings represented in unexplained cells of the pairwise comparisons table are significantly greater than pain ratings, but comparisons between ratings of remaining symptoms are not statistically significant. The final SDH according to analysis of the raw score data is mapped as illustrated in Figure 1.

Figure 1: Symptom Dominance Hierarchy for Patient: Raw Score Analysis

LEAST SEVERE	GI	An	De	Pa	{St, Co, Me, Sl}	Fa	MOST SEVERE

From the investigator's perspective an identical rating on the response scale indicates the same level of severity for the patient (and for all patients in sample-based studies), for all rated symptoms. For example, a rating of 4 is taken to indicate the same level of severity if the patient rates stiffness as if the patient rates every other symptom. Seen in this manner fatigue is the patient's dominant, most severe symptom. Significantly less severe than fatigue ratings are a statistically indistinguishable cluster of four symptoms: stiffness, concentration, memory, and sleep issues. Severity ratings for this cluster are significantly greater than pain ratings, which are greater than depression ratings, which are significantly greater than anxiety ratings. GI ratings were least severe.

Ipsative z-scores. The computational formula for a standardized *z*-score is: (patient's score–mean score)/standard deviation (SD). For a *normative z*-score, z_N, mean and SD are based on a sample of observations: conceptually z_N measures the magnitude of an observation's score relative to the population of scores *for all observations*. In contrast, for an ipsative *z*-score, z_I, mean and SD are based only on the data from the observation: conceptually z_I measures the magnitude of any observation's score relative to all scores in the population of scores *for the observation*.[8] Using z_I expresses data on a scale enabling direct comparison of different single-case series, and eliminates inter-series variability attributable to base-rate differences in mean and variance (essentially noise) when series are combined for use in sample-based studies.[8]

Distributions and descriptive statistics for z_I rating data for this patient's series of 297 entries are presented in Table 4 (grey shading is used for z-scores lower than the mean) and Table 5, respectively.

Table 4: Ipsative *z*-Score Distributions for Nine Symptoms: *Patient's* Perspective

Rating	Percentile	Pain	Stiffness	Fatigue	Concentration	Memory	Anxiety	Depression	GI	Sleep
-3.449	99.97									1
-2.704	99.66		1							
-2.579	99.51				2					
-2.341	99.04			5						
-2.211	98.65									4
-2.078	98.12		11							
-1.992	97.68				18					
-1.824	96.59					23				
-1.703	95.57	26								
-1.648	95.03			25						
-1.592	94.43									12

2155-0182/10/$3.00

-1.451	92.66			25										
-1.405	92.00						24							
-1.168	87.86									74				
-1.115	86.76								48					
-1.047	85.24	48												
-0.973	83.47													61
-0.964	83.25											97		
-0.955	83.02				50									
-0.825	79.53			56										
-0.818	79.33						36							
-0.406	65.76								69					
-0.391	65.21	74												
-0.374	64.58												233	
-0.354	64.95													85
-0.319	62.51											85		
-0.261	60.30				85									
-0.231	59.13						69							
-0.198	57.85			66										
-0.008	50.32									173				
0.265	39.55	65												54
0.303	38.09								80					
0.326	37.22											43		
0.356	36.09						69							
0.428	33.43			71										
0.432	33.29				60									
0.593	27.66												44	
0.884	18.84													40
0.922	17.83	49												
0.943	17.28						50							
0.970	16.60											37		
1.012	15.58								51					
1.055	14.57			42										
1.125	13.03				49									
1.152	12.47									32				
1.503	6.64													22
1.530	6.30						22							
1.560	5.94												8	
1.577	5.74	27												
1.615	5.32											22		
1.681	4.64			18										
1.721	4.26						22							
1.816	3.47				23									
2.117	1.71						7							
2.121	1.70													18
2.234	1.27	7												
2.260	1.19											8		
2.308	1.05			7										
2.312	1.04									11				
2.430	0.75						4							
2.527	0.58												3	
2.890	0.19	1												
2.904	0.18											2		
3.472	0.03									6				
3.494	0.02												5	
3.549	0.02											2		
4.194	<0.01											1		
4.461	<0.01												1	
4.632	<0.01									1				
5.428	<0.01												2	
8.330	<0.01												1	

Note: Rating is ipsative *z*-score. Percentile is the 1-sided percent of normally-distributed z-scores that exceed the tabled value: that is, the percent of ratings (days) which would be associated with a worse symptom rating.

2155-0182/10/$3.00

Table 5: Ipsative *z*-Score Descriptive Statistics for Nine Symptoms: *Patient's* Perspective

Symptom	Median	Skewness	Kurtosis
Pain	0.265	0.20	-0.53
Stiffness	-0.198	-0.01	-0.30
Fatigue	-0.261	-0.03	-0.59
Concentration	-0.231	-0.25	-0.33
Memory	0.303	0.05	-0.56
Anxiety	-0.008	1.42	3.32
Depression	-0.319	1.17	1.26
Gastrointestinal	-0.374	4.34	24.1
Sleep	-0.354	0.28	-0.03

Note: Mean=CV=0, SD=1, and SEM=0.058, for all symptoms.

Comparison of Tables 1 and 4 reveals that the patient does *not* use the numbers and labels on the response scale in the same manner across scales, as is assumed in the investigator's perspective.

Instead, the way the patient uses the numbers and labels on the response scale to report the personal experience of each symptom differs as a function of the mean and variability of each symptom. That is, the nature of the experience being rated interacts with the patient's interpretation of the numbers and labels of the scale—and as a result the identical numbers and labels indicate different levels of symptom severity, depending on the symptom being rated.

Data were not normally distributed (raw and z_I data produced identical results): for all nine symptoms, $p<0.01$ for Shapiro-Wilk, Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling tests for normality. Therefore the percentile of the standard normal distribution having worse (greater) severity ratings, provided in Table 3 for every z_I rating, is a poor interpretative heuristic. It should be noted that this is generally true in most applied research yet is largely ignored—data are interpreted as though they are normally distributed.

When viewed in terms of *relative negative severity*, the very best (least severe) of all symptoms the patient reported in the series was one single rating of sleep issues, with $z_I = -3.449$ (Table 4). In normally distributed data, a *z*-score as extremely negative as this is better than for 99.97% of the population of ratings on this symptom *for the patient*. The next least severe ipsative symptom ratings occurred for stiffness and concentration.

In contrast, when viewed in terms of *relative positive severity*, and in terms of *absolute overall severity*, the worst (most severe) symptom the patient reported in the series was one single rating of GI issues, with $z_I = 8.330$: in normally distributed data, for a *z*-score as extremely positive as this only 4.4×10^{-14}% of all possible days are worse on this symptom *for the patient*. In addition to eight other extreme positive ipsative GI ratings, the next-most-severe (worst) ipsative ratings occurred for anxiety and depression.

Note that while mean (0) and SD (1) of ipsatively standardized data are equivalent over attributes other patient response distribution moments clearly differ over scales (Table 5).

As was reported for analysis of raw data, results of all possible comparisons conducted on pairs of ipsatively standardized attributes using UniODA are summarized in Table 6 (100,000 Monte Carlo experiments were used to estimate *p* in each analysis, using a sequentially-rejective Sidak Bonferroni procedure to ensure that the experimentwise $p<0.05$.

As seen all 36 pairwise comparisons met the criterion for experimentwise $p<0.05$: twenty times more than expected by chance. Strongest effects were obtained for comparisons with GI issues, and to a lesser extent, for ratings of sleep and depression issues. Table 7 cross-tabulates the *qualitative* ESS and ESP levels obtained by UniODA models using raw and *z*-score data. Although the qualitative categories are ordered from weakest to strongest effect they are treated as being categorical because they are discussed and conceptualized categorically, not linearly, in the present context.[14]

2155-0182/10/$3.00

Table 6: *z*-Score Summary of UniODA Comparisons of All Pairs of Patient Symptoms: *Patient's* Perspective

	Stiffness	Fatigue	Concentration	Memory	Anxiety	Depression	Gastrointestinal	Sleep
Pain	St > Pa* 18.5, 19.2	Fa > Pa 22.9, 24.2	Co > Pa* 22.9, 24.2	Me > Pa 24.6, 25.5	Pa > An 33.3, 37.4	De > Pa 100, 57.1	Ga > Pa 49.8, 66.6	Sl > Pa* 23.6, 25.0
Stiffness		St > Fa 24.2, 24.7	St > Co 19.9, 21.4	St > Me 20.5, 22.2	St > An 29.6, 34.2	St > De 30.0, 30.1	St > Ga 47.1, 47.6	St > Sl 23.6, 24.0
Fatigue			Co > Fa 28.6, 29.5	Fa > Me* 20.2, 21.7	An > Fa 30.6, 31.8	Fa > De 34.3, 34.3	Fa > Ga 51.5, 51.7	Fa > Sl* 28.0, 28.9
Concentration				Co > Me 23.9, 25.4	Co > An 33.3, 37.1	Co > De 34.3, 34.8	Co > Ga 51.5, 51.7	Co > Sl 28.0, 28.9
Memory					Me > An 36.0, 39.7	De > Me 23.9, 56.8	Ga > Me 47.1, 65.4	Me > Sl 25.9, 27.0
Anxiety						An > De 36.4, 37.1	An > Ga 53.5, 53.6	An > Sl 30.0, 31.2
Depression							De > Ga* 45.8, 46.4	De > Sl 26.3, 57.6
Gastrointestinal								Sl > Ga 52.2, 52.3

Note: All $p<0.05$ at the experimentwise criterion. An asterisk (*) marking an inequality indicates that the direction of the effect, and the ESS and ESP that were obtained using raw and ipsatively standardized data were identical. An inequality indicated in color is an example of paradoxical confounding: red indicates that the opposite effect was obtained for raw data; green shows an effect found presently for which no statistically significant model was obtained using raw data; and blue shows an effect for which strength was overestimated in raw score analysis (more than one type of paradox was noted for some effects, but only the most significant type of paradox—such as finding the opposite effect—is recorded. ESS and ESP values (the left and right entries in second row of every table cell, respectively) shaded in grey are weak effects; green indicates moderate effects; red indicates strong effects; and blue shows very strong effects.[12]

2155-0182/10/$3.00

Table 7: Comparing ESS and ESP Values of UniODA Models Developed Using Raw and Ipsatively Standardized Data

	ESS		ESP	
	Raw	*z*-Score	Raw	*z*-Score
No Effect	6	0	6	0
Weak	6	12	5	9
Moderate	5	19	6	18
Strong	3	4	3	9
Very Strong	16	1	16	0

These data were analyzed using the UniODA Generalizability (Gen) procedure, which identifies the single UniODA model that—when simultaneously applied to multiple samples (or as used presently multiple aspects or dimensions of the findings)—maximizes the *minimum ESS* achieved by the model across classes.[12,15] The resulting Gen UniODA model was: if effect= none or very strong then predict raw data were analyzed; otherwise if effect=weak, moderate or strong then predict *z*-score data were analyzed. The model was statistically significant: by using 100,000 Monte Carlo experiments confidence for experimentwise $p<0.05$ is almost perfect. The model performed strongly in discriminating raw scores and *z*-scores on the basis of the effect strength noted in pairwise comparisons (ESS= 58.3, ESP=67.1)—Table 8 presents a confusion table for the Gen solution. The model was stable in leave-one-out (jackknife) validity analysis, suggesting it may cross-generalize.[12] For ESS the model achieved 61.1 for raw data and 59.7 for z-score data: for ESP these values were 72.0 and 69.5. The findings indicate that compared to models based on *z*-scores, the UniODA models based on raw scores were likely to *miss effects* identified using *z*-scores, and to identify more very strong effects using raw scores. Note that chi-square analysis can't be performed on data in Table 7 considered as a whole or considering ESS and ESP individually, because the expected value is less than five in four cells for ESS, and in two cells for ESP.[16]

Table 8: Confusion Table for Gen UniODA Model Discriminating ESS and ESP Based on Type of Data Analyzed

		Predicted Data Type		
		Raw	*z*-Score	
Actual Data Type	Raw	44	28	61.1%
	z-Score	1	71	98.6%
		97.8%	71.7%	

More important than the pattern of *qualitative effect strengths* of UniODA models identified using raw or *z*-scores is the *qualitative nature of the effects* identified. Comparison of Tables 3 and 6 reveals that of 36 comparisons of symptom pairs performed using raw data, 28 (78%) were confounded by three different types of paradox when considered from the perspective of ipsative data.[7] As indicated in Table 6, ten (28%) raw-score models identified the opposite effect (what raw data analysis suggested occurred was the opposite of what occurred for ipsative analysis); six (17%) missed statistically significant effects identified using *z*-scores; and twelve (33%) effects identified for raw scores over-estimated parallel effects found using *z*-scores (one moderate effect found for raw data was a weak effect for *z*-scores; one strong effect for raw scores was a moderate effect using *z*-scores; three very strong effects for raw scores were strong effects for *z*-scores; and seven very strong effects identified using raw scores were moderate effects for *z*-scores). Six raw and *z*-score-based models were identical.

Most important is symptom dominance structure underlying pairwise differences found for ipsative data, which is much more complex than structure observed for raw scores. Analysis

2155-0182/10/$3.00

begins as before: ratings of stiffness exceeded all other ratings with $p<0.05$, so at this point the SDH is mapped as shown.

```
LOW ------------------------------------|-- HIGH
                                        St
```

Co ratings exceeded remaining ratings, so the SDH is now mapped as shown.

```
LOW --------------------------------|-----|-- HIGH
                                    Co   St
```

Fa ratings exceed all other ratings except for An ratings, so the SDH is now mapped as shown.

```
LOW ------------------|-----|-----|-----|-- HIGH
                     Fa    An   Co   St
```

At this point the mapping procedure can no longer integrate remaining inequalities in the pairwise comparisons table. A similar situation was encountered using UniODA to sequentially decompose serial structure in stratigraphic samples, also expressed as a set of pairs of inequalities, that were identified in maximum-accuracy Markov analysis.[12] In that analysis the underlying solution involved sequential identification of multiple underlying structures (factors)—as are obtained by principal components analysis.[12,17] Successive models are created using remaining unexplained inequalities until all inequalities are explained. Presently the primary structure identified is illustrated above, and it represents 21 (58%) of 36 inequalities in the pairwise comparisons table.

The search begins for a second structure that explains the 15 remaining unexplained inequalities in the pairwise-comparison table. No linear pattern is evident, so the first row of the table containing five unexplained inequalities is systematically mapped from the left. First we see Me > Pa, and next Pa > An. Integrating the two inequalities gives: Me > Pa > An. The third inequality is De > Pa: the new map is: (Me, De) > Pa > An. Symptoms within parentheses have *not* been disentangled by inequalities addressing their differences: final solution comes when this recursive procedure is completed. The fourth inequality is integrated next: (Me, De, Ga) > Pa > An. The last inequality in the top row is now integrated: (Me, De, Ga, Sl) > Pa > An.

The procedure continues to the fifth row of the pairwise comparisons table, and inequalities involving Me are examined. The first, Me > An, is already represented in the current structure, so this inequality is explained. The second inequality, De > Me, is next mapped within the parentheses (De > Me); the third inequality is added (De, Ga > Me); and the fourth and last inequality in this row of the pairwise comparisons table is mapped (De, Ga > Me > Sl). This substructure is integrated with the second model in progress: (De, Ga > Me > Sl) > Pa > An. Thus far this model explains 9 (60%) of the 15 inequalities remaining after the initial model was identified. The structure within parentheses is not yet resolved.

The procedure continues to the sixth row of the pairwise comparisons table, and inequalities involving An are examined. Since all three of these inequalities disagree with the current mapping, they are set aside.

The procedure continues to the seventh row of the pairwise comparisons table, and inequalities involving De are examined. The first inequality resolves the parentheses in the second model: De > Ga > Me > Sl > Pa > An. The second inequality is already explained in the model. The third inequality disagrees with the model so it is set aside.

The last inequality remaining in the last row of the pairwise comparisons table disagrees with the current mapping, and so it is set aside. The second structure is now complete, and it explains 10 (67%) of 15 inequalities that remain unexplained by the first model. Considered together these models explain (21+10)/36=86% of findings in the pairwise comparisons table.

2155-0182/10/$3.00

For the third structure, the first An finding is mapped An > De and the second is added: An > (De, Ga). The third is added next: An > (De, Ga, Sl). In the ninth row of the pairwise comparison table, the inequality involving De and Sl is added to this structure: An > (Ga, (De > Sl)). The last unexplained inequality in the bottom corner of the pairwise confusion table is integrated: An > ((Ga, De) > Sl). The relationship between Ga and De is still unresolved, and no unexplained inequalities exist. However, the high-end of the second symptom dominance hierarchy is De > Ga, which resolves the third structure: An > De > Ga > Sl.

Together the three models, summarized in Table 9, explain the set of 36 inequalities in the pairwise comparisons table. Note the De > Ga comparison that was initially credited to the second model was credited to the third model as it was used in the final step: as seen, the second and third models reflect translocation of the De > Ga inequality.

Table 9: Three Symptom Dominance Hierarchies Identified Using Ipsative *z*-Scores

Model	Structure of Dominance Hierarchy	Pairwise Comparisons Explained (%) Model	Cumulative
1	Stiffness > Concentration > Anxiety > Fatigue	21 (58%)	21 (58%)
2	GI > Memory > Sleep > Pain > Anxiety	9 (25%)	30 (83%)
3	Anxiety > Depression > GI > Sleep	6 (16%)	36 (100%)

Clearly, by the ipsative perspective, not all of the ratings made by the patient on nine symptom scales over 297 sequential days can be explained by a single symptom dominance hierarchy (SDH). By the ipsative perspective three SDHs are needed.

All three structures identified by *z*-scores are reminiscent of phase-shift models produced in Markov models.[12] For example it is plausible that moving from most to least severe in the first model, increased stiffness is followed by greater concentration issues next, which in-turn increase anxiety resulting in greater fatigue. Models may also run in the opposite direction, from least to worst, in the actual series: for example from the third model, increased sleep issues phase into increased GI issues, followed in-turn by phase shift into increased depression, and culminating in significantly greater anxiety.

UniODA model cutpoints for raw and *z*-score data are given in Table 10 for all pairwise comparisons. This information and the pairwise (in)equalities in Tables 3 and 6 are sufficient to deduce all UniODA models. For example for ipsative comparison of Pa and St the cutpoint is -0.29, and direction of effect is St > Pa. Thus the UniODA model is: if *z*-score $\leq$ -0.29 then predict rating=Pa, otherwise predict rating=St.

Finally, Table 11 lists different ipsative z-score cutpoints used in UniODA models, for every raw-score cutpoint used. The precision and sensitivity of the ipsative data cannot be matched by raw scores.

2155-0182/10/$3.00

Table 10: Cutpoints for UniODA Models Based on Raw (Above Diagonal) and *z*-Score (Below Diagonal) Data

	Pain	Stiffness	Fatigue	Concentration	Memory	Anxiety	Depression	GI	Sleep
Pain		4	5	4	4	2	2	1	4
Stiffness	-0.29		5	4	4	2	3	1	3
Fatigue	-0.33	-0.23		5	5	3	3	2	5
Concentration	-0.31	0.39	-0.25		3	2	3	1	3
Memory	0.28	0.37	-0.33	0.33		2	3	2	5
Anxiety	0.13	0.21	-0.13	0.17	0.15		1	0	2
Depression	-1.01	-0.26	-0.29	-0.28	-1.04	-0.16		0	3
GI	-0.38	-0.29	-0.32	-0.30	-0.39	-0.19	-0.35		2
Sleep	-0.37	-0.28	-0.31	-0.29	0.28	-0.18	-0.97	-0.36	

Note: GI=gastrointestinal. Cutpoints for UniODA models based on *raw* data are indicated in black *above* the diagonal; cutpoints for UniODA models based on *ipsatively standardized* data are indicated in red *beneath* the diagonal.

Table 11: Comparing UniODA Model Cutpoints for Raw and Ipsatively Standardized Data

Raw Data Cutpoints	Ipsative *z*-Score Cutpoints
0	-0.19, -0.35
1	-0.16, -0.29, -0.30, -0.38
2	0.21, 0.17, 0.15, 0.13, -0.18, -0.29, -0.36, -0.39, -1.01
3	0.33, -0.13, -0.28, -0.29, -0.97, -1.04
4	0.39, 0.28, -0.23, -0.29, -0.31, -0.37
5	-0.23, -0.25, -0.29, -0.31, -0.33

The next analysis required to understand symptom dominance hierarchy structures that underlie ratings made by this patient involves an analysis of temporal expression of symptoms over time. Two methods for accomplishing this using UniODA have been developed, including the serial *N*-of-1 methods reviewed earlier, and the Markov phase-shift approach discussed here. Beyond the domain of this paper, these analyses will be undertaken in subsequent research.

References

[1]Soltysik RC, Yarnold PR (2010). The use of unconfounded climatic data improves atmospheric prediction. *Optimal Data Analysis*, *1*, 67-100.

[2]Yarnold PR (2013). Trajectory of crude mortality rate in North Dakota counties. *Optimal Data Analysis*, *2*, 120-142.

[3]Yarnold PR (2013). Statistically significant increases in crude mortality rate of North Dakota counties occurring after massive environmental usage of toxic chemicals and biocides began there in 1997: An optimal *static* statistical map. *Optimal Data Analysis*, *2*, 98-105.

[4]Yarnold PR (2013). The most recent, earliest, and *K*th significant changes in an ordered series: Traveling backwards in time to assess when annual crude mortality rate most recently began increasing in McLean County, North Dakota. *Optimal Data Analysis*, *2*, 143-147.

2155-0182/10/$3.00

[5]Collinge, W, Soltysik RC, Yarnold PR (2010). An internet-based intervention for fibromyalgia self-management: Initial design and alpha test. *Optimal Data Analysis, 1*, 163-175.

[6]Collinge W, Yarnold PR, Soltysik RC (2013). Fibromyalgia symptom reduction by online behavioral self-monitoring, longitudinal single subject analysis and automated delivery of individualized guidance. *North American Journal of Medical Sciences, 5*, 546-553.

[7]Yarnold PR (1996). Characterizing and circumventing Simpson's paradox for ordered bivariate data. *Educational and Psychological Measurement, 56*, 430-442.

[8]Yarnold PR, Soltysik RC (2013). Ipsative transformations are *essential* in the analysis of serial data. *Optimal Data Analysis*, 2, 94-97.

[9]Yarnold PR (2013). Comparing attributes measured with "identical" Likert-type scales in single-case designs with UniODA. *Optimal Data Analysis*, 2, 148-153.

[10]Yarnold PR (2013). Comparing responses to dichotomous attributes in single-case designs. *Optimal Data Analysis*, 2, 154-156.

[11]Yarnold PR (1992). Statistical analysis for single-case designs. In: FB Bryant, L Heath, E Posavac, J Edwards, E Henderson, Y Suarez-Balcazar, S Tindale (Eds.), *Social Psychological Applications to Social Issues, Volume 2: Methodological Issues in Applied Social Research*. New York, NY: Plenum, pp. 177-197.

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

[13]Yarnold PR, Feinglass J, Martin GJ, McCarthy WJ (1999). Comparing three pre-processing strategies for longitudinal data for individual patients: An example in functional outcomes research. *Evaluation and the Health Professions, 22*, 254-277.

[14]Yarnold PR, Soltysik RC (2010). UniODA vs. chi-square: Ordinal data sometimes feign categorical. *Optimal Data Analysis, 1*, 62-65.

[15]Yarnold PR, Soltysik RC (2010). Gen UniODA vs. log-linear model: Modeling discrimination in organizations. *Optimal Data Analysis, 1*, 59-61.

[16]Yarnold JK (1970). The minimum expectation of χ^2 goodness-of-fit tests and the accuracy of approximations for the null distribution. *Journal of the American Statistical Society, 65*, 864-886.

[17]Bryant FB, Yarnold PR (1995). Principal components analysis and exploratory and confirmatory factor analysis. In: LG Grimm, PR Yarnold,(Eds.)., Reading and understanding multivariate statistics. Washington DC: APA Books, ps. 99-136.

Author Notes

E-mail: Journal@OptimalDataAnalysis.com

Mail: Optimal Data Analysis, LLC
1220 Rosecrans St., #330
San Diego, CA 9210

ODA Blog: http://odajournal.wordpress.com