



Disclaimer:

Presentation slides from the Rare Disease Workshop Series are posted by the Kakkis EveryLife Foundation, for educational purposes only. They are for use by drug development professionals and statisticians, and are not to be used to guide the prescribing or use of any of the drugs mentioned in the slides. To obtain information on a particular drug, refer to the drug labeling. Do not reproduce or distribute the slides (full set or any portion of) without the permission of the author.



RARE DISEASE WORKSHOP SERIES

Improving the Clinical Development Process

Statistical Analyses for Multi-Domain Outcomes

James Signorovitch, PhD,
Analysis Group Inc, Boston MA



Why Consider Multi-Domain Outcomes for Rare Diseases?

- Heterogeneity across patients can be an important and unavoidable characteristic of the disease
 - The same underlying disease can lead to multiple domains of impairment
 - Different patients will have baseline impairment in different domains
 - Requiring “substantial” baseline impairment in one domain (or in all domains) may not be possible due to unavoidably small sample sizes



Objectives

- Illustrate the use of multi-domain methods using MPS trial data
- Assess tradeoffs between interpretability and sensitivity vs. single domain analyses

Interpretability	Sensitivity
<ul style="list-style-type: none">-How can we assess whether there is substantial benefit?-How can we interpret the benefit?	<ul style="list-style-type: none">- When is a multi-domain approach more sensitive than a single domain?



Data

- Trials
 - Aldurazyme (N=45)
 - Naglazyme (N=39)
 - Elaprase (N=64)
- Outcomes (Δ baseline to endpoint)
 - 6-minute walk test (6MWT)
 - Forced vital capacity (FVC)
 - Shoulder flexion
 - Health Assessment Questionnaire (HAQ) and Childhood HAQ (CHAQ)



O'Brien Test

- Extends the Wilcoxon rank sum test to multiple domains
 1. Rank changes from baseline separately within each domain
 2. Sum ranks across all domains in the treatment group
 3. Assess significance of the rank sum vs. its known null distribution



Wilcoxon and O'Brien Test Results in Individual Trials

- In these 3 MPS trials, the 3-domain O'Brien test would have been more sensitive than Wilcoxon tests based on any single domain

Test / Outcome	#Trials with a Statistically Significant Treatment Effect
Wilcoxon Tests	
Outcome 1	1* out of 3
Outcome 2	0 out of 3
Outcome 3	1* out of 3
O'Brien test combining all three outcomes	2** out of 3

*Different trials

** Significance in all 3 trials if using absolute vs. %-predicted FVC per protocol



Wei-Johnson Test

- Extends t-test to multiple domains
 - Need to weight each outcome
 - Statistical vs. clinical importance
 - Equal weights, inverse variance weights, optimal weights



T test and Wei-Johnson Test Results in Individual Trials

Test / Outcome	#Trials with a Statistically Significant Treatment Effect (out of 3)
T tests	
Outcome 1	2
Outcome 2	0
Outcome 3	1
Wei-Johnson tests combining all three outcomes	
Equal weights	1
Inverse variance weights	2
Large-sample optimal weights	0



Wilcoxon and O'Brien Test Results in Pooled Trial Data

Test	Mean Rank Sum Difference	SE	P-Value
Wilcoxon tests			
6MWT	23.5	6.6	<0.001
FVC	18.1	6.7	0.007
Shoulder	1.6	6.8	0.117
O'Brien (6MWT, FVC, Shoulder)	17.4	3.9	<0.001



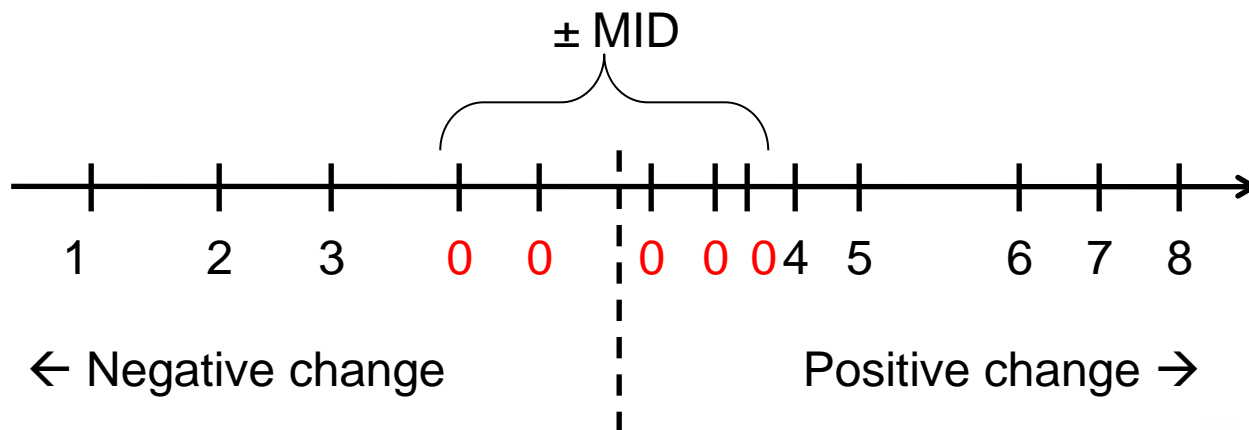
Conclusions

- Treatment improved the average rank across the 3 domains vs. placebo
 1. Is this change clinically meaningful?
 2. How do we interpret this change?



1. Is this change clinically meaningful?

- Partial answer: sensitivity analysis:
 1. Set to 0 all changes with magnitude $<$ the MID and re-rank
 2. Re-run the O'Brien test with the new ranking





Wilcoxon and O'Brien Test Results in Pooled Trial Data: Sensitivity Analysis

- The treatment effect remained after ignoring changes in individual domains that did not exceed the MID
- Difference is not due only to many changes $<$ MID
- Still need to assess whether the treatment difference is meaningful

Test	Mean Rank Sum Difference	SE	P-Value
Wilcoxon tests			
6MWT	20.0	5.8	<0.001
FVC	20.5	5.9	0.007
Shoulder	4.9	5.2	0.341
O'Brien (6MWT, FVC, Shoulder)	15.1	3.4	<0.001



2. How do we interpret this change?

- Responder analyses
 - Multi-domain responder definitions
 - Single domain responder definitions
- Descriptive analyses
 - Mean, median, cumulative distribution for individual domains
- **Assess association between composite change score and other outcomes**



Correlations with CHAQ/HAQ

- The composite score was more strongly correlated with disability than the individual scores

Objective assessment	Pearson Correlation with CHAQ/HAQ*	P-Value
Continuous scores		
6MWT	-0.24	0.032
FVC	-0.19	0.091
Shoulder	-0.23	0.051
O'Brien rank score (6MWT, FVC, Shoulder)	-0.50	<0.001

*Patient-assessments pooled from treatment and placebo arms (n=76); Higher CHAQ/HAQ scores indicate greater impairment .



Conclusions

- The O'Brien test showed increased sensitivity for treatment effects in these three MPS trials vs. separate Wilcoxon tests
- Treatment differences were primarily attributable to differences in 6MWT and FVC
- Treatment differences detected via the O'Brien test were attributable to clinically significant changes in individual domains
- Correlations between the O'Brien composite score and the CHAQ/HAQ suggested that this composite could be more meaningful to the average MPS patient than any single component score



Discussion

- The O'Brien test is one way of weighting outcomes within and across domains and combining them into a single score. Alternative weighting schemes (grounded on supportive data) could increase interpretability and sensitivity
- Sensitivity analysis setting changes $< \text{MID}$ in each domain may be too conservative if $< \text{MID}$ changes in multiple domains can be important in combination
- MID's were not based on MPS patients
- Other settings may be more challenging to interpret (e.g., if no individual score shows a significant treatment effect)



Simulations

- Objective: to assess when combining two domains can improve power vs. studying a single domain



Simulations

- Two continuous, normally distributed outcomes with equal variance
- Parameters
 - Treatment effect for outcome 1 (expressed relative to the SD)
 - Treatment effect for outcome 2 (expressed relative to the SD)
 - Correlation between outcomes 1 and 2
- 50 subjects, 25 per treatment arm
- 500 simulation runs per parameter combination to determine Type I error rate and statistical power of Wilcoxon and O'Brien tests



Simulations with Correlation $\rho = 0.8$

$\rho = 0.8$		Mean Treatment Effect 2								
		0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
Mean Treatment Effect 1	0.0	-	0.05	0.13	0.22	0.39	0.50	0.59	0.76	0.89
	0.2		0.13	0.22	0.33	0.50	0.62	0.77	0.89	0.95
	0.4			0.37	0.54	0.69	0.78	0.89	0.94	0.97
	0.6				0.69	0.80	0.89	0.96	0.97	1.00
	0.8					0.88	0.94	0.97	1.00	1.00
	1.0						0.98	0.99	1.00	1.00
	1.2							1.00	1.00	1.00
	1.4								1.00	1.00
	1.6									1.00



Simulations with Correlation $\rho = 0.4$

$\rho = 0.4$		Mean Treatment Effect 2								
		0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
Wilcoxon			0.10	0.27	0.51	0.76	0.92	0.98	1.00	1.00
Mean Treatment Effect 1	0.0	-	0.05	0.14	0.24	0.42	0.57	0.67	0.81	0.94
	0.2		0.14	0.27	0.36	0.58	0.70	0.83	0.94	0.97
	0.4			0.43	0.61	0.73	0.85	0.94	0.97	0.99
	0.6				0.78	0.84	0.92	0.98	0.99	1.00
	0.8					0.93	0.97	0.99	1.00	1.00
	1.0						0.99	1.00	1.00	1.00
	1.2							1.00	1.00	1.00
	1.4								1.00	1.00
1.6									1.00	



Simulations with Correlation $\rho = 0.2$

$\rho = 0.2$		Mean Treatment Effect 2								
		0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
Mean Treatment Effect 1	0.0	-	0.10	0.27	0.51	0.76	0.92	0.98	1.00	1.00
	0.2		0.06	0.15	0.27	0.44	0.60	0.70	0.84	0.95
	0.4		0.14	0.29	0.38	0.61	0.75	0.86	0.96	0.98
	0.6			0.43	0.64	0.78	0.89	0.95	0.98	0.99
	0.8				0.81	0.87	0.94	0.99	0.99	1.00
	1.0					0.95	0.98	0.99	1.00	1.00
	1.2						0.99	1.00	1.00	1.00
	1.4							1.00	1.00	1.00
	1.6								1.00	1.00



Simulations with Correlation $\rho = 0$

$\rho = 0.0$		Mean Treatment Effect 2								
		0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
Wilcoxon			0.10	0.27	0.51	0.76	0.92	0.98	1.00	1.00
Mean Treatment Effect 1	0.0	-	0.06	0.16	0.30	0.48	0.63	0.73	0.87	0.97
	0.2		0.15	0.31	0.42	0.66	0.79	0.89	0.98	0.99
	0.4			0.48	0.68	0.82	0.91	0.98	0.99	0.99
	0.6				0.84	0.89	0.96	0.99	1.00	1.00
	0.8					0.96	0.99	1.00	1.00	1.00
	1.0						1.00	1.00	1.00	1.00
	1.2							1.00	1.00	1.00
	1.4								1.00	1.00
1.6									1.00	



Simulations with Correlation $\rho = -0.2$

$\rho = -0.2$		Mean Treatment Effect 2								
		0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
Mean Treatment Effect 1	0.0	-	0.10	0.27	0.51	0.76	0.92	0.98	1.00	1.00
	0.2		0.07	0.16	0.33	0.52	0.71	0.77	0.90	0.98
	0.4		0.17	0.33	0.46	0.70	0.83	0.92	0.99	0.99
	0.6			0.54	0.73	0.85	0.94	0.98	0.99	1.00
	0.8				0.86	0.92	0.96	0.99	1.00	1.00
	1.0					0.98	0.99	1.00	1.00	1.00
	1.2						1.00	1.00	1.00	1.00
	1.4							1.00	1.00	1.00
	1.6								1.00	1.00



Simulations with Correlation $\rho = -0.4$

$\rho = -0.4$		Mean Treatment Effect 2								
		0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
Mean Treatment Effect 1	0.0	-	0.10	0.27	0.51	0.76	0.92	0.98	1.00	1.00
	0.2		0.07	0.18	0.36	0.57	0.75	0.82	0.93	0.99
	0.4		0.17	0.35	0.50	0.75	0.87	0.94	0.99	0.99
	0.6			0.58	0.77	0.89	0.97	0.99	1.00	1.00
	0.8				0.90	0.95	0.98	0.99	1.00	1.00
	1.0					0.99	1.00	1.00	1.00	1.00
	1.2						1.00	1.00	1.00	1.00
	1.4							1.00	1.00	1.00
	1.6								1.00	1.00



Simulations with Correlation $\rho = -0.8$

$\rho = -0.8$		Mean Treatment Effect 2								
		0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
Wilcoxon			0.10	0.27	0.51	0.76	0.92	0.98	1.00	1.00
Mean Treatment Effect 1	0.0	-	0.09	0.23	0.47	0.67	0.84	0.93	0.98	1.00
	0.2		0.22	0.44	0.65	0.85	0.95	0.99	1.00	1.00
	0.4			0.70	0.85	0.94	0.99	1.00	1.00	1.00
	0.6				0.95	0.98	1.00	1.00	1.00	1.00
	0.8					1.00	1.00	1.00	1.00	1.00
	1.0						1.00	1.00	1.00	1.00
	1.2							1.00	1.00	1.00
	1.4								1.00	1.00
1.6									1.00	



Conclusions

- Combining two outcomes with positive treatment effects in an O'Brien test does not necessarily increase power vs. considering only the most responsive outcome
- The O'Brien test tends to improve power vs. the Wilcoxon test when the outcomes have similarly sized treatment effects and when they are less positively correlated



MPS Trials Exhibited Settings Favorable to the O'Brien Test

Comparison of Clinical MPS Trial Parameters With Simulation

MPS Trial	Outcome Combination	Estimated MPS Trial Parameter			Statistical Power ²		
		Treatment Effect 1	Treatment Effect 2	Correlation	Wilcoxon 1	Wilcoxon 2	O'Brien
Trial 1	1 & 2	0.89	0.56	0.08	0.76	0.51	0.89
	1 & 3	0.89	0.13	-0.04	0.76	0.10	0.66
	2 & 3	0.56	0.13	0.08	0.51	0.10	0.42
Trial 2	1 & 2	0.64	0.50	0.09	0.51	0.51	0.84
	1 & 3	0.64	0.30	0.17	0.51	0.27	0.64
	2 & 3	0.50	0.30	0.47	0.51	0.27	0.61
Trial 2	1 & 2	0.51	0.40	0.16	0.51	0.27	0.64
	1 & 3	0.51	0.30	0.16	0.51	0.27	0.64
	2 & 3	0.40	0.30	-0.05	0.27	0.27	0.48

Notes:

1. The treatment effects for the clinical MPS trial outcomes are normalized with variance equal to 1.
2. The statistical power of the Wilcoxon test and the O'Brien rank-sum tests are only approximate and correspond to the closest parameter match of the simulation.
3. Clinical MPS trial outcomes for which the Wilcoxon test showed a statistically significant results are highlighted in bold.



Discussion

- The O'Brien test can reduce the chances of missing a multi-domain treatment benefit vs. focusing on a single domain
- Prior information on the associations between outcomes in different domains can be valuable in designing composite scores
- The components of a meaningful and sensitive composite score need not have strong positive correlations