

Comparison of Analytical Techniques for Multiple Endpoints in Clinical Trials

Darcy Vavrek, ND, MS¹
Nancy Temkin, PhD^{2,3}

¹Center for Outcomes Studies, Western States Chiropractic College, Portland, OR, USA
²Department of Biostatistics, University of Washington, Seattle, WA, USA
³Department of Neurological Surgery, University of Washington, Seattle, WA, USA

Presenter Disclosures

No relationships to disclose

Darcy Vavrek ND MS, Western States Chiropractic College

Discussion points

- Multiple endpoints: advantages, challenges, and strategies
- Strategies for atypical data
- O'Brien's rank sum composite method
- Reading a power curve at a glance
 - (Type I error and power by eye)
- Evaluating selected analytical techniques
- Recommended analytical techniques for specific situations

Darcy Vavrek ND MS, Western States Chiropractic College

Why multiple endpoints?

- Treat the whole patient
- Measure the whole patient

For example:

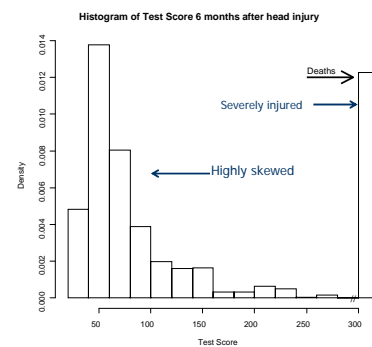
- Decrease pain
- Increase mobility
- Decrease risk of mortality
- Increase cognitive ability

Darcy Vavrek ND MS, Western States Chiropractic College

Challenges in Head Injury trials:

- Subjects die
- Subjects often quite severely injured (cannot be tested)
- Distributions are highly skewed
- Endpoints are highly correlated
- Variables of interest are binary, categorical, or continuous
- And I want to combine these variables!
- It's about assessing the whole patient.

Darcy Vavrek ND MS, Western States Chiropractic College



Strategies for multiple endpoints

- Bonferroni adjustments
- Combine data into a composite measure
 - O'Brien's Rank Sum Test
- Use a global statistic
 - generated by a generalized linear model (GLM)
 - utilizing generalized estimating equations (GEE)

Darcy Vavrek ND MS, Western States Chiropractic College

Strategies for atypical data (non-normality)

- Rank the data:
 - Multiple rank tests (Bonferroni corrected)
 - One rank test (composite rank sum)
 - ANOVA (composite rank sum)
- Dichotomize the data:
 - Multiple chi-square tests (Bonferroni corrected)
 - Global OR (logistic regression GEE)
- Place data on ordinal scale:
 - Global OR (ordinal logistic regression GEE)
- Standardize outcomes:
 - Multiple t-tests (Bonferroni correction)
 - Global change in effect size (simple generalized linear model GEE)

Darcy Vavrek ND MS, Western States Chiropractic College

Composite Rank Sum

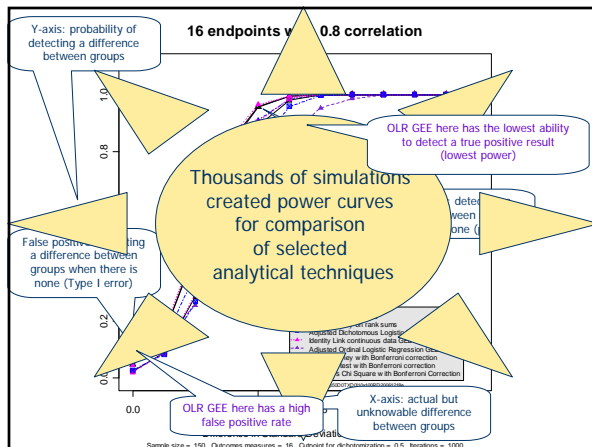
Subject/outcome	Clinical Outcomes			Rank			Sum
	o1	o2	o3	r1	r2	r3	
1	0.7	0.6	0.8	4	2	4	10
2	0.5	0.7	0.4	2	3	2	7
3	0.2	0.9	0.2	1	4	1	6
4	0.6	0.4	0.5	3	1	3	7

Darcy Vavrek ND MS, Western States Chiropractic College

Results of simulations (thousands of simulations)

- With normal data
 - Best performers (highest power)
 - Composite rank sum
 - Global effect size (simple GLM GEE)
- Large numbers of endpoints
 - Global OR (ordinal logistic regression GEE)
 - Failed with large numbers of highly correlated endpoints
 - Increase in studies falsely declared positive (Type I error)
 - Other methods had low risk of falsely declaring a study positive.
- High correlation
 - Ability to detect a true positive result (power) decreased as correlation increased across all methods
 - Not unexpected

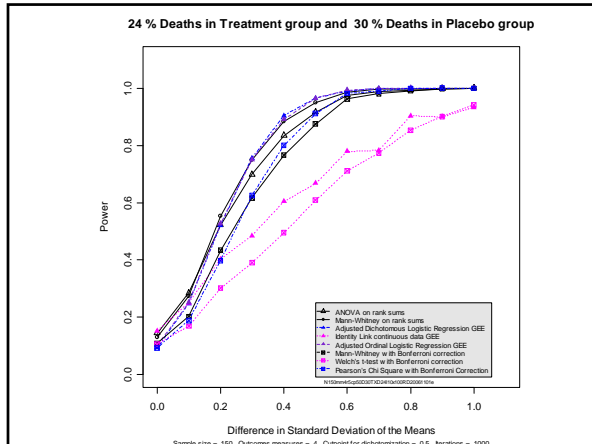
Darcy Vavrek ND MS, Western States Chiropractic College



Results of simulations

- Null treatment effect in some endpoints
 - Greatly decreased performance - most techniques
 - Least affected - **Bonferroni Rank test**
- Deaths with similar mortality in both groups
 - Best - **Global OR (Logistic regression GEE)**
- Deaths with improved mortality in treated group
 - Performed equally well
 - Global OR (Logistic regression GEE)
 - Composite rank sum
 - Notably worse
 - Global effect size (simple GLM GEE)

Darcy Vavrek ND MS, Western States Chiropractic College



Discussion

- O'Brien's composite rank sum with ANOVA performed well with:
 - Large amounts of highly correlated endpoints
 - Large mortality percentages
 - (or those who would be untestable)
- Results favor use of **composite rank sums**
- Rank sum score is difficult to interpret

Darcy Vavrek ND MS, Western States Chiropractic College

Discussion

- Global OR (Logistic regression GEE)
 - Second best technique
 - Performs well despite:
 - Large amounts of highly correlated endpoints
 - Death
 - Interpretable analysis
- Bonferroni
 - Poorest performance in most situations with notable exception of:
 - Null treatment effect in some endpoints
 - This is a caution to investigators about including possibly unaffected endpoints in the model

Darcy Vavrek ND MS, Western States Chiropractic College

Conclusion

- O'Brien's composite rank sum
 - powerful technique in most situations
- Global OR (logistic regression GEE)
 - an effective choice when interpretability of the results is a priority

Darcy Vavrek ND MS, Western States Chiropractic College

Acknowledgements

- Mentors and helpful suggestions
 - Sureyya Dikmen, PhD
 - Gerald van Belle, PhD
 - Mary Lou Thompson, PhD
- Andrew David attentive editing
- Grant Funding: CA0916827, U01 HD42653, and R01 NS19643

Darcy Vavrek ND MS, Western States Chiropractic College

References

1. Temkin NR, Anderson G, Anstrom KJ, Nelson P, Winn HR, Dikmen SS. Valproate vs. phenytoin for preventing early post-traumatic seizures. *Epilepsia*. 1996; 37(Suppl 5): 83.
2. Temkin NR, Dikmen SS, Anderson GD, Wilensky AJ, Holmes MD, Cohen W, et al. Valproate therapy for prevention of posttraumatic seizures: A randomized trial. *J Neurosurg*. 1999 Oct; 91(4): 593-600.
3. Temkin NR, Dikmen SS, Wilensky AJ, Keihm J, Chabal S, Winn HR. A randomized, double-blind study of phenytoin for the prevention of post-traumatic seizures. *N Engl J Med*. 1990 Aug 23; 323(8): 497-502.
4. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics*. 1984 Dec; 40(4): 1079-87.
5. O'Brien PC, Shampo MA. Statistical considerations for performing multiple tests in a single experiment. 5. comparing two therapies with respect to several endpoints. *Mayo Clin Proc*. 1988 Nov; 63(11): 1140-3.
6. Pocock SJ, Geller NL, Tsiatis AA. The analysis of multiple endpoints in clinical trials. *Biometrics*. 1987 Sep; 43(3): 487-98.
7. Tang DI, Geller NL, Pocock SJ. On the design and analysis of randomized clinical trials with multiple endpoints. *Biometrics*. 1993 Mar; 49(1): 23-30.
8. Pocock SJ. Clinical trials with multiple outcomes: A statistical perspective on their design, analysis, and interpretation. *Control Clin Trials*. 1997 12; 18(6): 530-45.
9. Sankoh AJ, Huque MF, Dubey SD. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat Med*. 1997 Nov 30; 16(22): 2529-42.
10. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 71(1): 13.
11. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986 Mar; 42(1): 121-30.
12. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med*. 1995 12/14; 333(24; 24): 1581-8.
13. Tilley BC, Marler J, Geller NL, Lu M, Legler J, Brott T, et al. Use of a global test for multiple outcomes in stroke trials with application to the national institute of neurological disorders and stroke t-PA stroke trial. *Stroke*. 1996 Nov; 27(11): 2136-42.
14. Lu M, Tilley BC, NINDS t-PA Stroke Trial Study Group. Use of odds ratio or relative risk to measure a treatment effect in clinical trials with multiple correlated binary outcomes: Data from the NINDS t-PA stroke trial. *Stat Med*. 2001 Jul 15; 20(13): 1891-901.
15. Stroke Therapy Academic Industry Roundtable II (STAIR-II). Recommendations for clinical trial evaluation of acute stroke therapies. *Stroke*. 2001 Jul; 32(7): 1598-606.
16. Tilley BC, Pillemer SR, Heyse SP, Li S, Clegg DO, Alarcon GS. Global statistical tests for comparing multiple outcomes in rheumatoid arthritis trials. MIRA trial group. *Arthritis Rheum*. 1999 Sep; 42(9): 1879-88.
17. Follmann D. Multivariate tests for multiple endpoints in clinical trials. *Stat Med*. 1995 06/15; 14(11): 1163-75.
18. O'Brien PC, Geller NL. Interpreting tests for efficacy in clinical trials with multiple endpoints. *Control Clin Trials*. 1997 Jun; 18(3): 222-7.
19. Zhang J, Quan H, Ng J, Stepanavage ME. Some statistical methods for multiple endpoints in clinical trials. *Control Clin Trials*. 1997 Jun; 18(3): 204-21.
20. Bregnezer T, Lehmacher W. Directional tests for the analysis of clinical trials with multiple endpoints allowing for incomplete data. *Biometrical Journal*. 1998; 40(8): 911-28.
21. Gray SM, Brookmeyer R. Estimating a treatment effect from multidimensional longitudinal data. *Biometrics*. 1998 Sep; 54(3): 976-88.
22. Chi GYH. Multiple testings: Multiple comparisons and multiple endpoints. *Drug Inf J*. 1998; 32(Suppl): 1347S-62S.
23. Neuhauser M, Volker S, Bretz F. The evaluation of multiple clinical endpoints, with application to asthma. *Drug Inf J*. 1999; 33: 471-7.
24. Wassmer G, Reitmeir P, Kieser M, Lehmacher W. Procedures for testing multiple endpoints in clinical trials: An overview. *Journal of Statistical Planning and Inference*. 1999; 82: 69-81.

References

25. D'Agostino RB S. Controlling alpha in a clinical trial: The case for secondary endpoints. *Stat Med.* 2000 Mar 30; 19(6): 763-6.
26. Gong J, Pinheiro JC, DeMets DL. Estimating significance level and power comparisons for testing multiple endpoints in clinical trials. *Control Clin Trials.* 2000 Aug; 21(4): 313-29.
27. Schouten HJ. Combined evidence from multiple outcomes in a clinical trial. *J Clin Epidemiol.* 2000 Nov; 53(11): 1137-44.
28. Senn S, Stevens L, Chaturvedi N. Repeated measures in clinical trials: Simple strategies for analysis using summary measures. *Stat Med.* 2000 Mar 30; 19(6): 861-77.
29. Bender R, Lange S. Adjusting for multiple testing--when and how? *J Clin Epidemiol.* 2001 Apr; 54(4): 343-9.
30. Bloch DA, Lai TL, Tubert-Bitter P. One-sided tests in clinical trials with multiple endpoints. *Biometrics.* 2001 Dec; 57(4): 1039-47.
31. Logan BR, Tamhane AC. Combining global and marginal tests to compare two treatments on multiple endpoints. *Biometrical Journal.* 2001; 43(5): 591-604.
32. Chen JJ, Wang S. Testing for treatment effects on subsets of endpoints. *Biometrical Journal.* 2002 07; 44(5): 541-57.
33. Lu M, Chen J, Lu D, Yi L, Mahmood A, Chopp M. Global test statistics for treatment effect of stroke and traumatic brain injury in rats with administration of bone marrow stromal cells. *J Neurosci Methods.* 2003 Sep 30; 128(1-2): 183-90.
34. Karrison TG, O'Brien PC. A rank-sum-type test for paired data with multiple endpoints. *Journal of Applied Statistics.* 2004 02; 31(2): 229-38.
35. Huang P, Tilley BC, Woolson RF, Lipsitz S. Adjusting O'Brien's test to control type I error for the generalized nonparametric behrens-fisher problem. *Biometrics.* 2005 Jun; 61(2): 532-9.
36. Dueck A, Novotny PJ. Dealing with multiple endpoints. *Curr Probl Cancer.* 2006 11; 30(6): 298-306.
37. Neuhäuser M. How to deal with multiple endpoints in clinical trials. *Fundam Clin Pharmacol.* 2006 12; 20(6): 515-23.
38. Bloch DA, Lai TL, Su Z, Tubert-Bitter P. A combined superiority and non-inferiority approach to multiple endpoints in clinical trials. *Stat Med.* 2007 Mar 15; 26(6): 1193-207.
39. Chuang-Stein C, Stryszak P, Dmitrienko A, Offen W. Challenge of multiple co-primary endpoints: A new approach. *Stat Med.* 2007 03/15; 26(6): 1181-92.
40. Saver JL. Novel end point analytic techniques and interpreting shifts across the entire range of outcome scales in acute stroke trials. *Stroke.* 2007 Nov; 38(11): 3055-62.
41. Senn S, Bretz F. Power and sample size when multiple endpoints are considered. *Pharm Stat.* 2007 Jul-Sep; 6(3): 161-70.
42. Temkin NR, Anderson GD, Winn HR, Ellenbogen RG, Britz GW, Schuster J, et al. Magnesium sulfate for neuroprotection after traumatic brain injury: A randomised controlled trial. *Lancet Neurol.* 2007 Jan; 6(1): 29-38.
43. Mann H, Whitney D. On a test of whether on of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics.* 1947; 18(1): 50.
44. Diggle PJ, Heagerty PJ, Liang KY, Zeger SL. Analysis of longitudinal data. 2nd ed. New York: Oxford University Press; 2002.
45. Connor RJ. Grouping for testing trends in categorical data. *Journal of the American Statistical Association.* 1972 09; 67(339): 601.