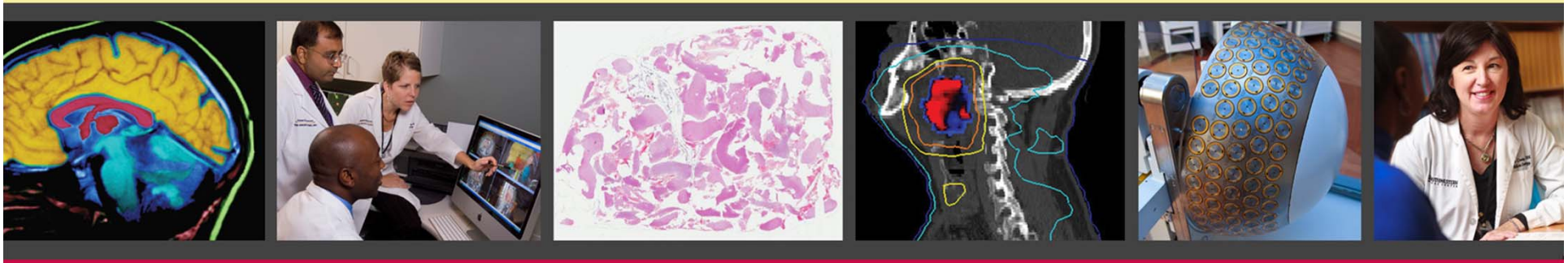
RTOG®
RADIATION THERAPY ONCOLOGY GROUP



Comparison of Futility Monitoring Methods Using RTOG Clinical Trials

Q. Ed Zhang, PhD

ACR® AMERICAN COLLEGE OF RADIOLOGY

Futility Monitoring

- Definition: Monitoring for early determination that trial results will not be in favor of H_1
- Of interest when interim analyses show lack of benefit
- Can be complex decision, require multiple methods (see case study: Dignam et al 1998)
- When monitoring for futility, it is important to
  - Protect against increasing type II error, since early negative trends may be due to chance
  - Distinguish inefficacy (i.e., futility) from harm

Futility Monitoring Methods:

- Conditional power: Lan, et al. 1982
- Asymmetric group sequential methods (GSM):
- DeMets & Ware 1980, 1982, Whitehead, et al. 1983, Pampallona et al. 1994, Chang et al. 1998
- Repeated confidence intervals (RCI): Jennison & Turnbull, 1984, 1989, 2000
- Testing to reject H_1 at stringent alpha (Fleming et al. 1984)
- Bayesian predictive power/probabilities: Spiegelhalter, et al. 1986
- Recent proposal: Linear Inefficacy boundary (LIB20): Freidlin, Korn & Gray, 2010

Metric for Efficacy in Survival Data

- Logrank statistic after d^{th} ordered event is:

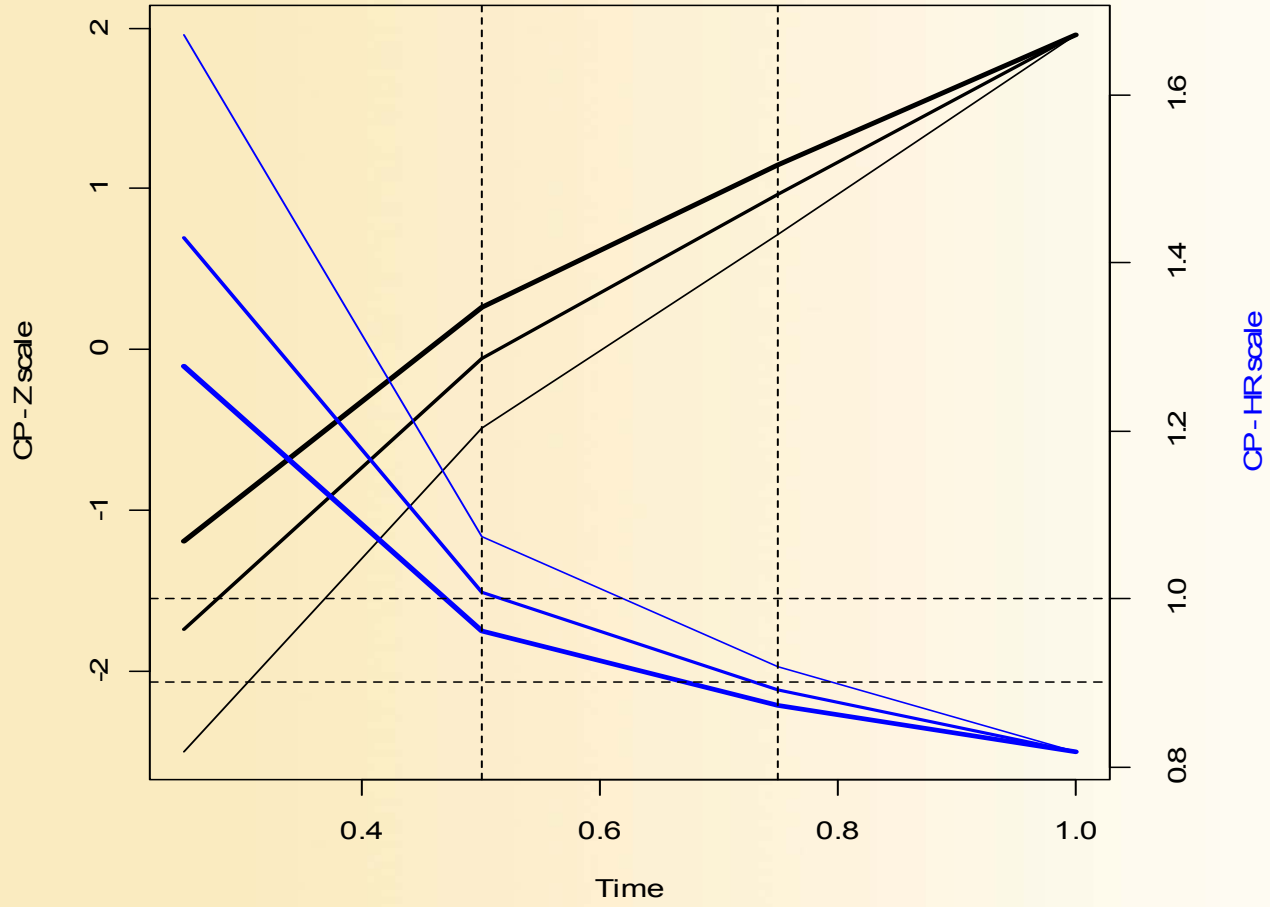
$$Z_d(t) = \sum_{i=1}^{d} (O_i - E_i) / \sqrt{I_d}$$

- $Z(t)$ is logrank test at information time $t = d/D$ events

- When $\theta = \log \lambda \to 0$, $Z_d(t) \sim N(\theta\sqrt{I_d}, 1)$

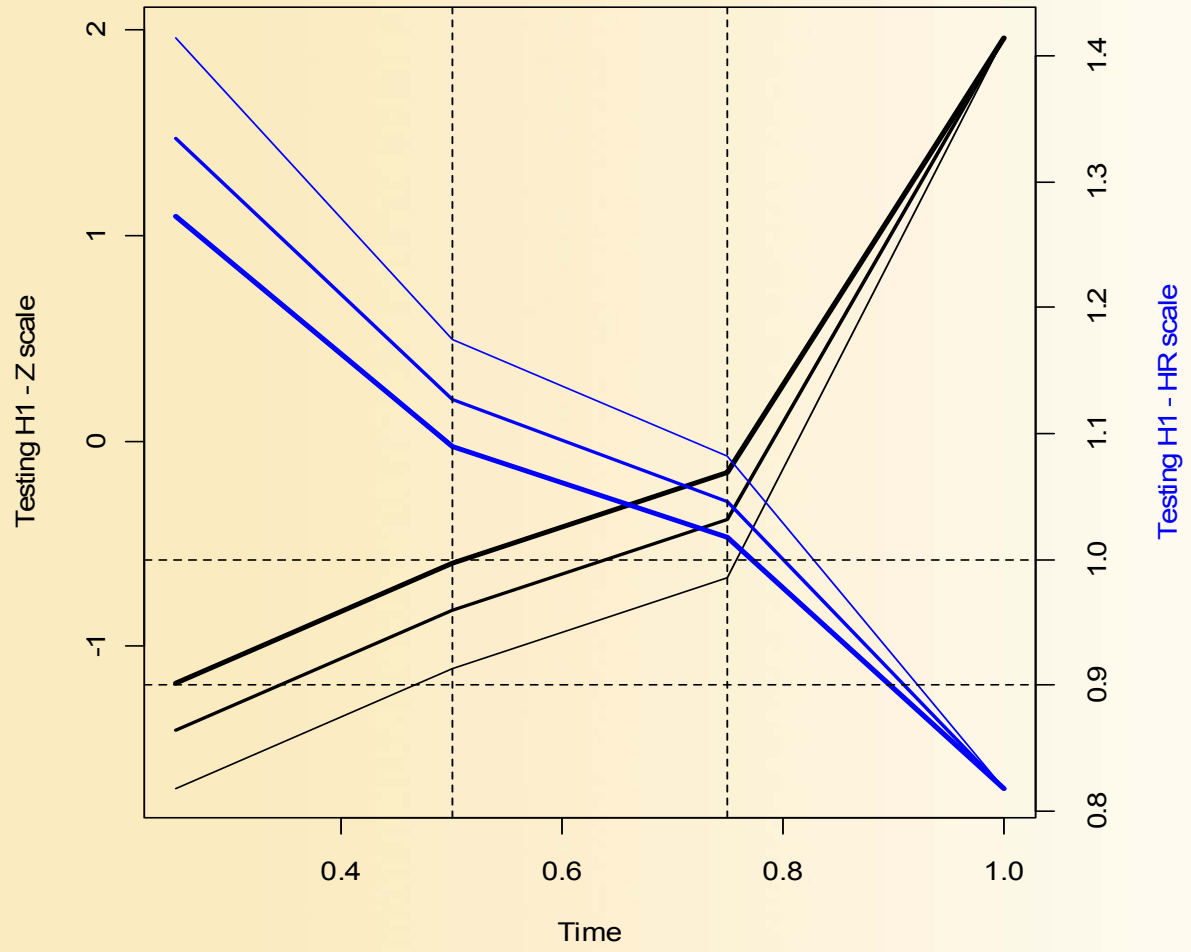
Futility Rules: Conditional Power Boundaries

- $P[Z(t=1) \geq Z_\alpha | Z(t), \theta] = 1 - \Phi\{\frac{|Z_\alpha - Z(t)\sqrt{t} - \theta(1-t)|}{\sqrt{1-t}}\}$



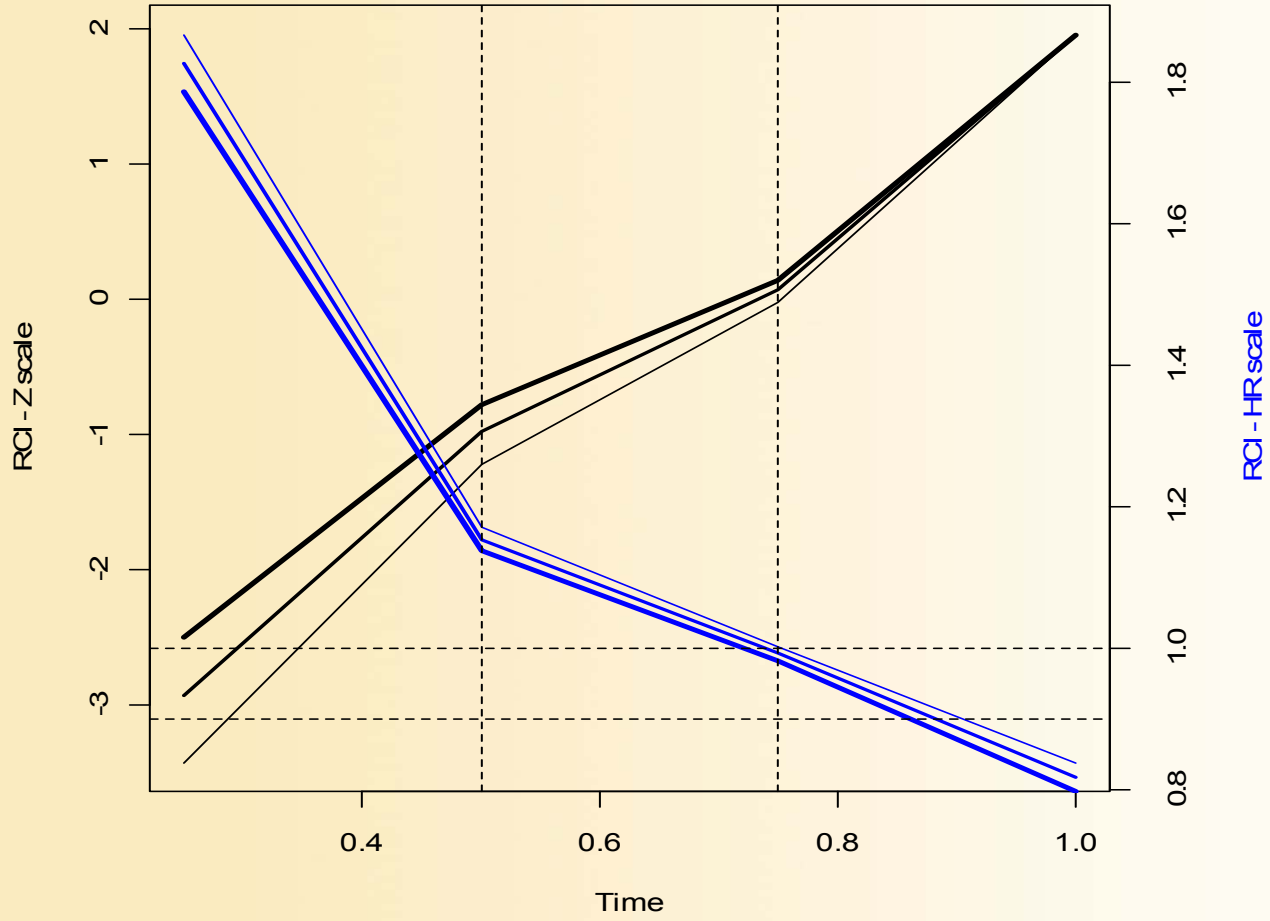
Futility Rules: Testing H_1

- $b = Z_\alpha + \log(\lambda) * \sqrt{d/4}$, α=0.001, 0.0025, 0.005



Futility Rules: Repeated Confidence Intervals

- $Z(t) + C_{OBF} * SE < \theta(t)$



Problems With Current Rules

Freidlin & Korn (2002, 2009) addressed futility stopping in several circumstances:

- Experimental arm is better, smaller effect than H_1 target
- Conditional power is not low
- Time-dependent treatment effect (non-PH)

Some Observations/Recommendations:

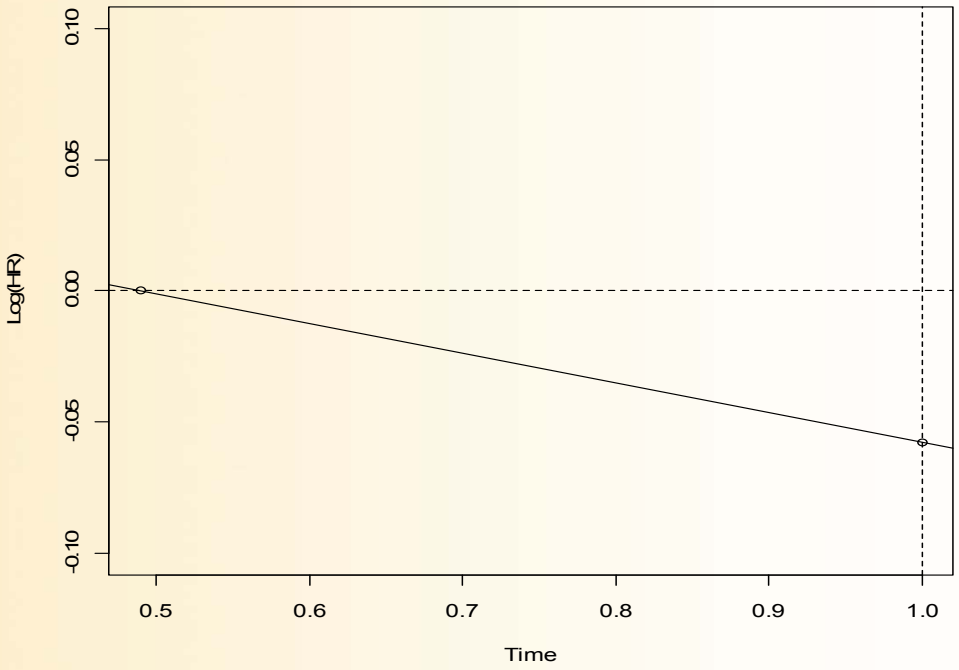
- Aggressive futility monitoring for highly toxic regimens vs. placebo
- Conservative monitoring when treatment is historically effective elsewhere, or may have delayed effect
- Overly aggressive rules stop when there is benefit (Ex. HR = 0.80)

Also:

- RCI and testing against H1 too conservative mid-trial
- Conditional power too aggressive towards end of trial

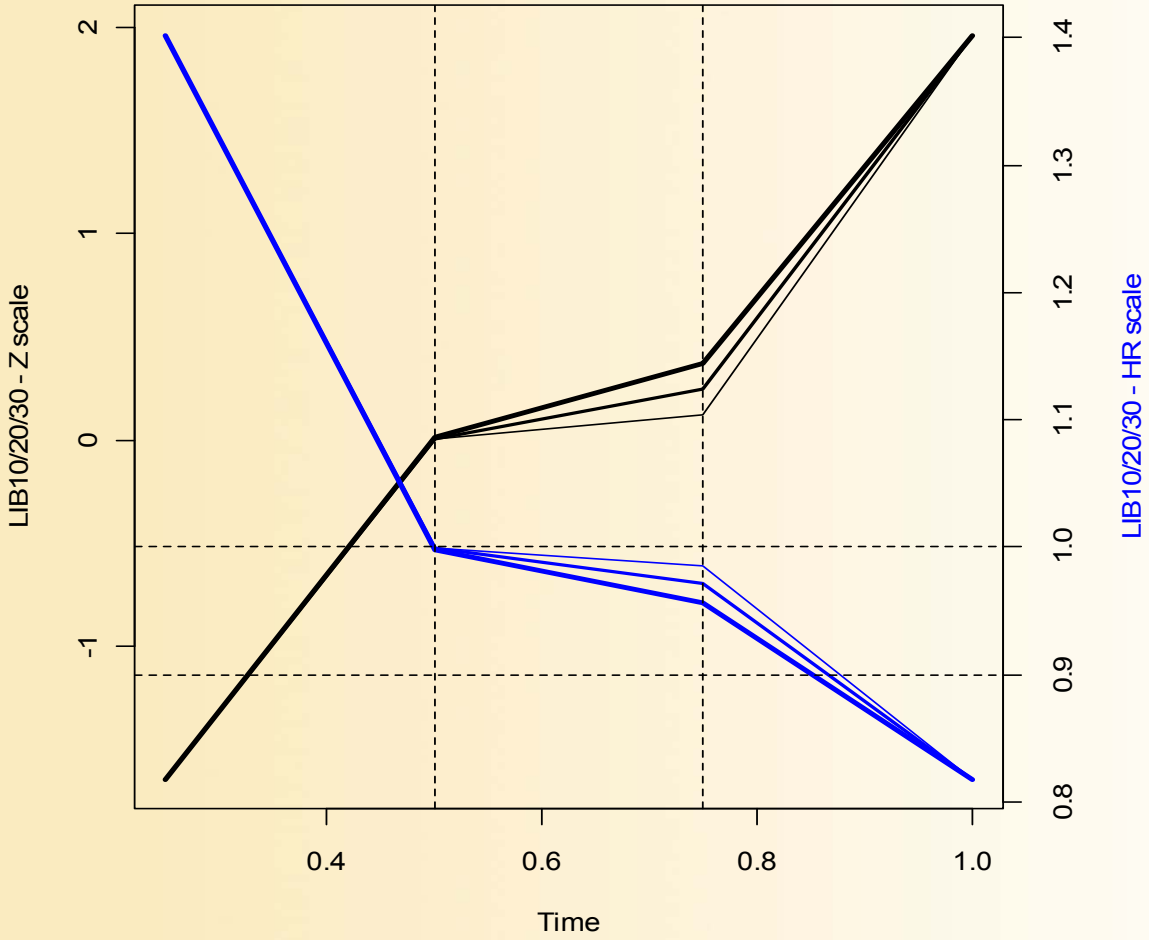
LIB20 - Freidlin, Korn & Gray 2010

- $t_0 = \left(\frac{C_{0.975}}{C_{1-\alpha/2}+C_{1-\beta}}\right)$
- $0.20 * \log\theta_A \frac{(C_{1-\frac{\alpha}{2}}+C_{1-\beta})^2 t - C_{0.975}^2}{(C_{1-\frac{\alpha}{2}}+C_{1-\beta})^2 - C_{0.975}^2}$



- Timely stopping if therapy is harmful or there is no tangible benefit
- Simulation results show that designs with LIB20 will need less patients and shorter duration under H_0 and minor loss of power ($\leq 1\%$) under H_1
- Wieand et al. 1994

LIB10/20/30


LIB10/20/30 - Z scale
2
1
0
-1
0.4
0.6
0.8
1.0
Time
1.4
1.3
1.2
1.1
1.0
0.9
0.8
LIB10/20/30 - HR scale

Comparing Futility Rules using RTOG Trials

- Reported RTOG phase III trials since 1990 (12 trials, 17 comparisons)
- Multiple disease sites
- Total number of events (if not given) calculated using type I, II error rates in protocol
- Multiple comparisons adjusted
- Logrank tests used for all time to event endpoints
- Upper boundary
- Analysis schedules:
  1. Six/Seven looks: 0.25, (0.40), 0.50, 0.60, 0.70, 0.80, 0.90, 1
  2. Three looks: 0.25, (0.40), 0.50, (0.70), 0.75, 1
  3. One look: 0.5, 1

Calculation of Savings

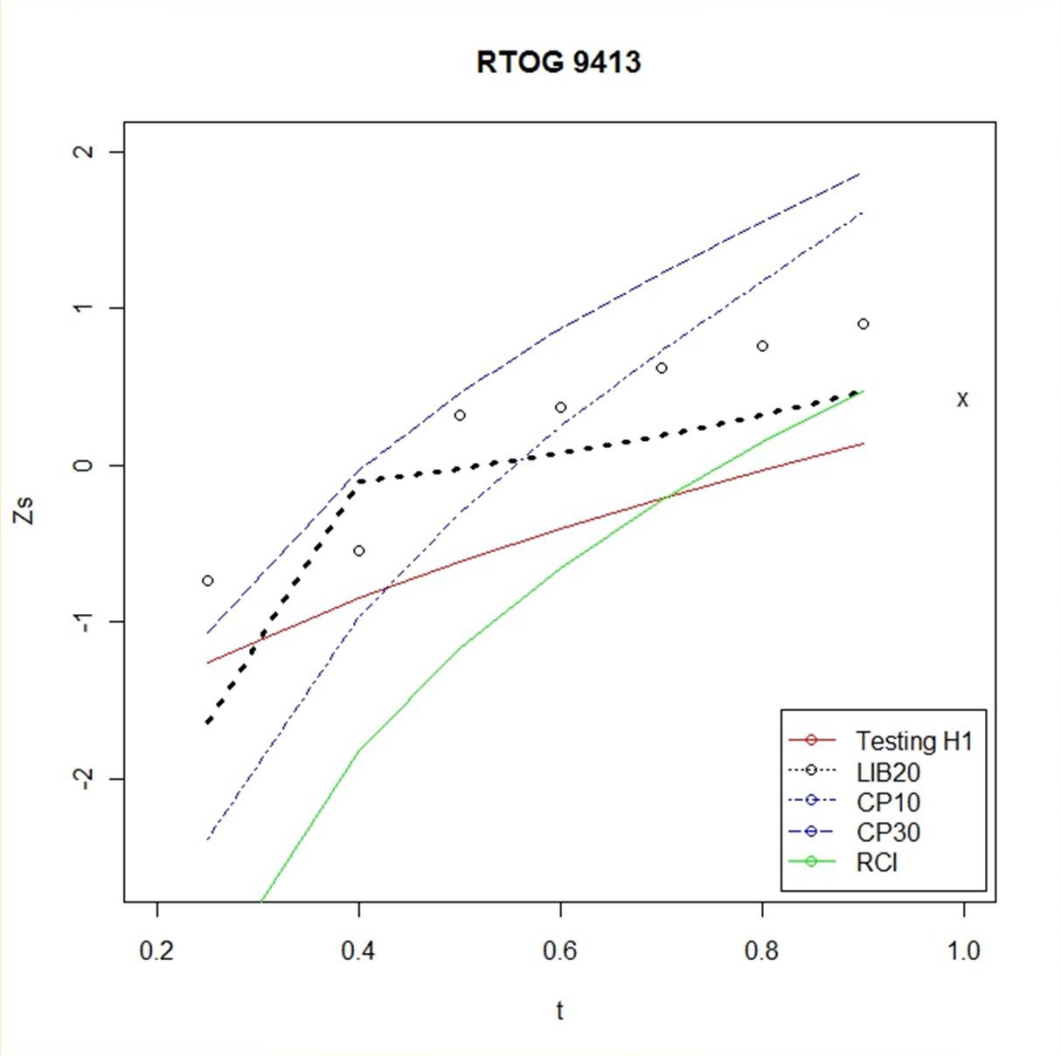
- For each trial and boundary if stop due to futility then
  1. Calculate observed study duration
  2. Calculate number of observed events
  3. Calculate current sample size

- Three groups based on final p values: trials we would not want to stop ($p<0.1$, n=3), neutral ($0.1<p<0.2$, n=3) and stop for sure ($P>0.2$, n=11)

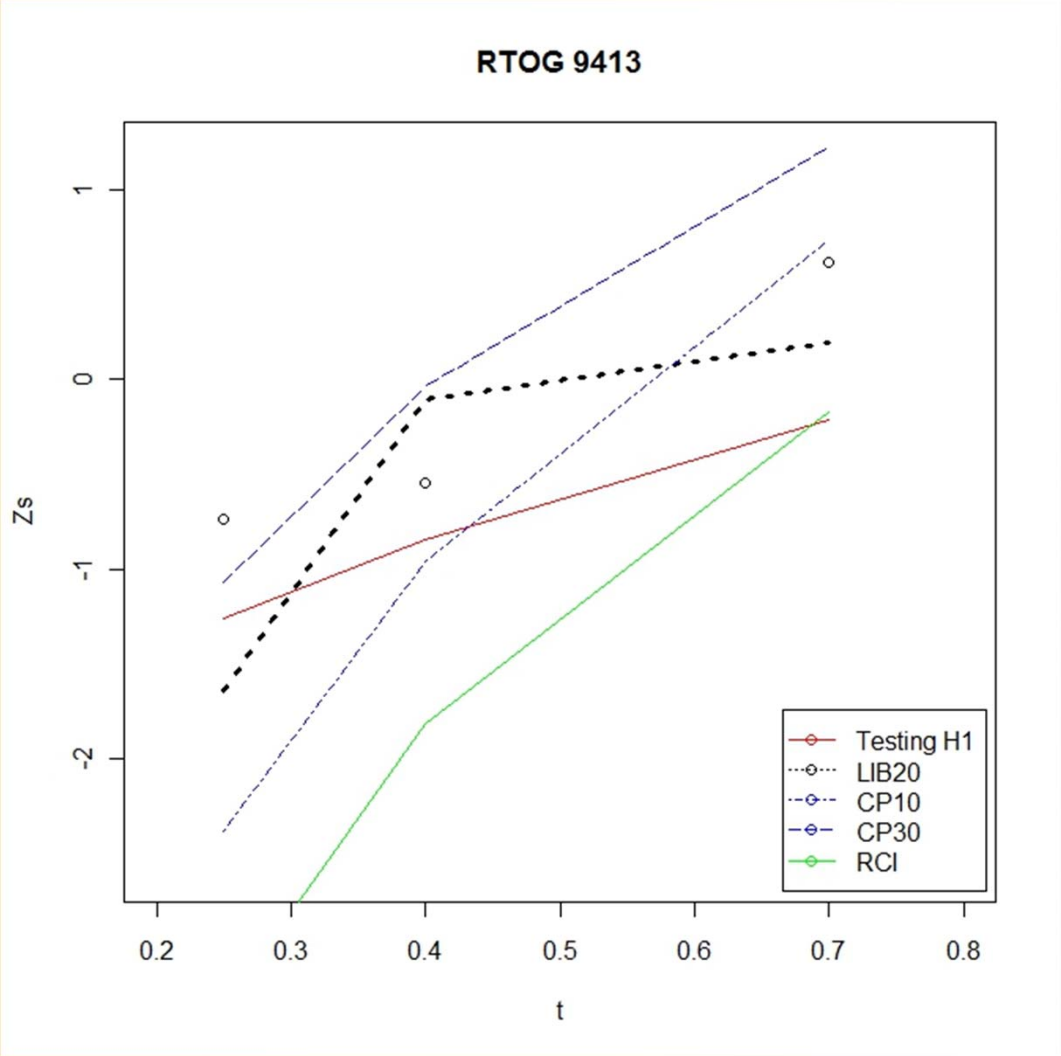
- For all trials in one group given an analysis schedule

- Savings=1- $\frac{\sum_{i=1}^{k} Observed}{\sum_{i=1}^{k} Planned}$ for each rule

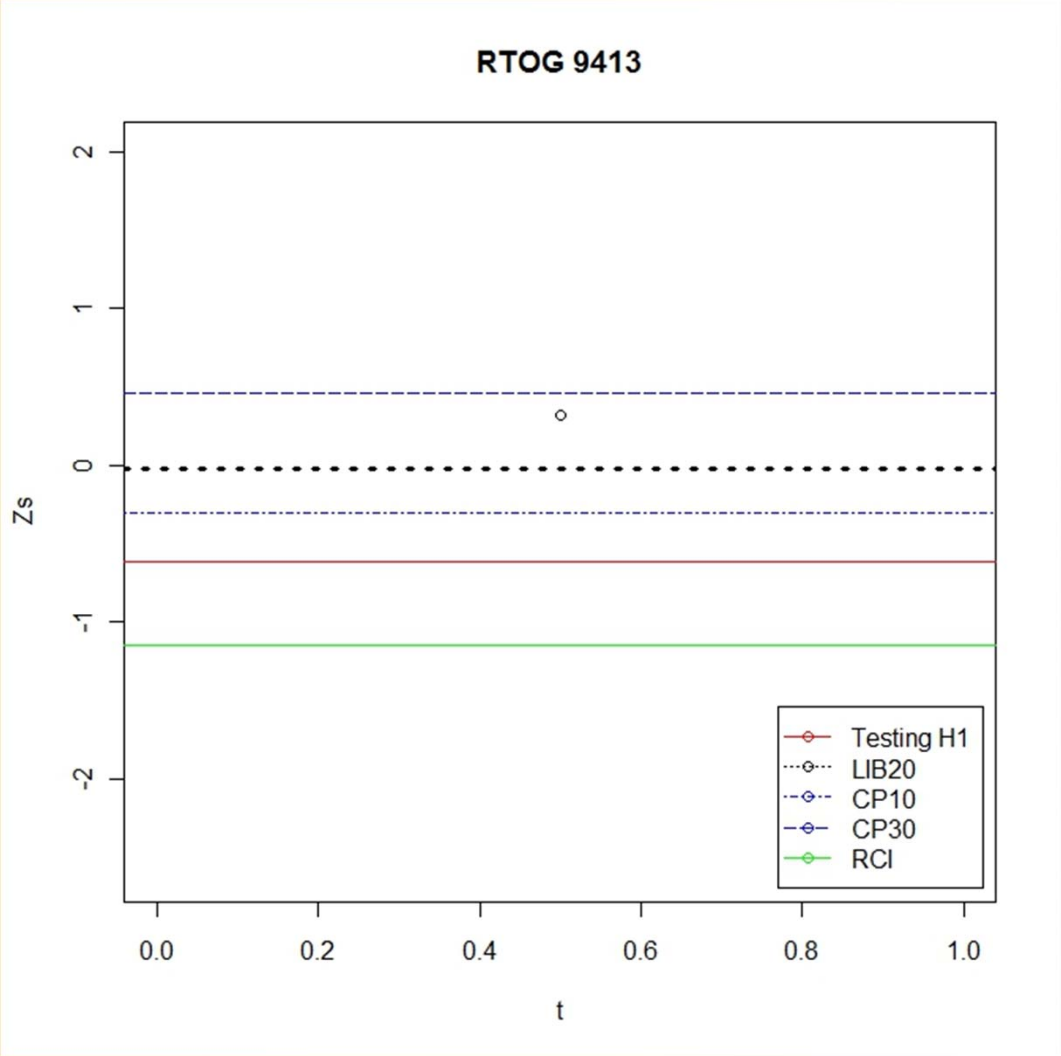
Comparing Futility Rules - Example



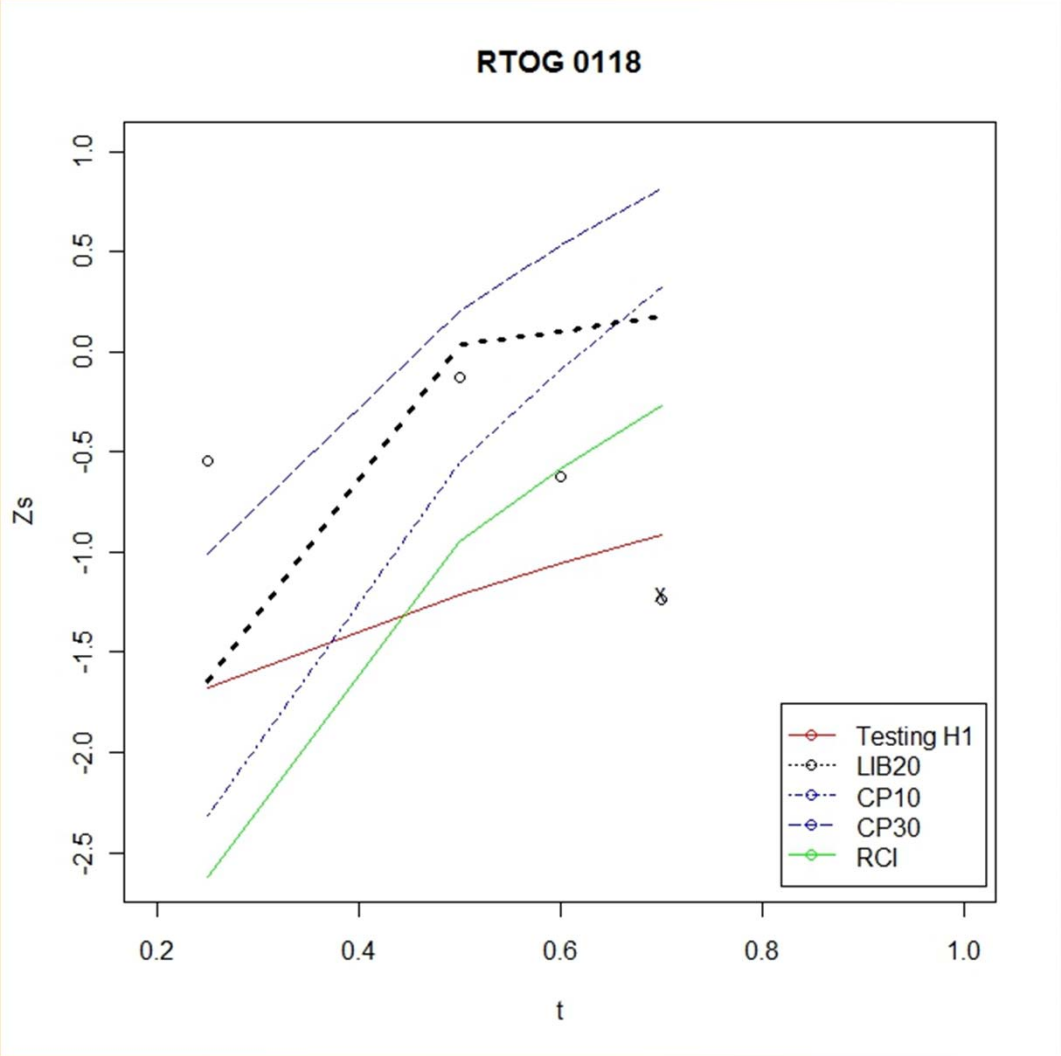
Comparing Futility Rules - Example



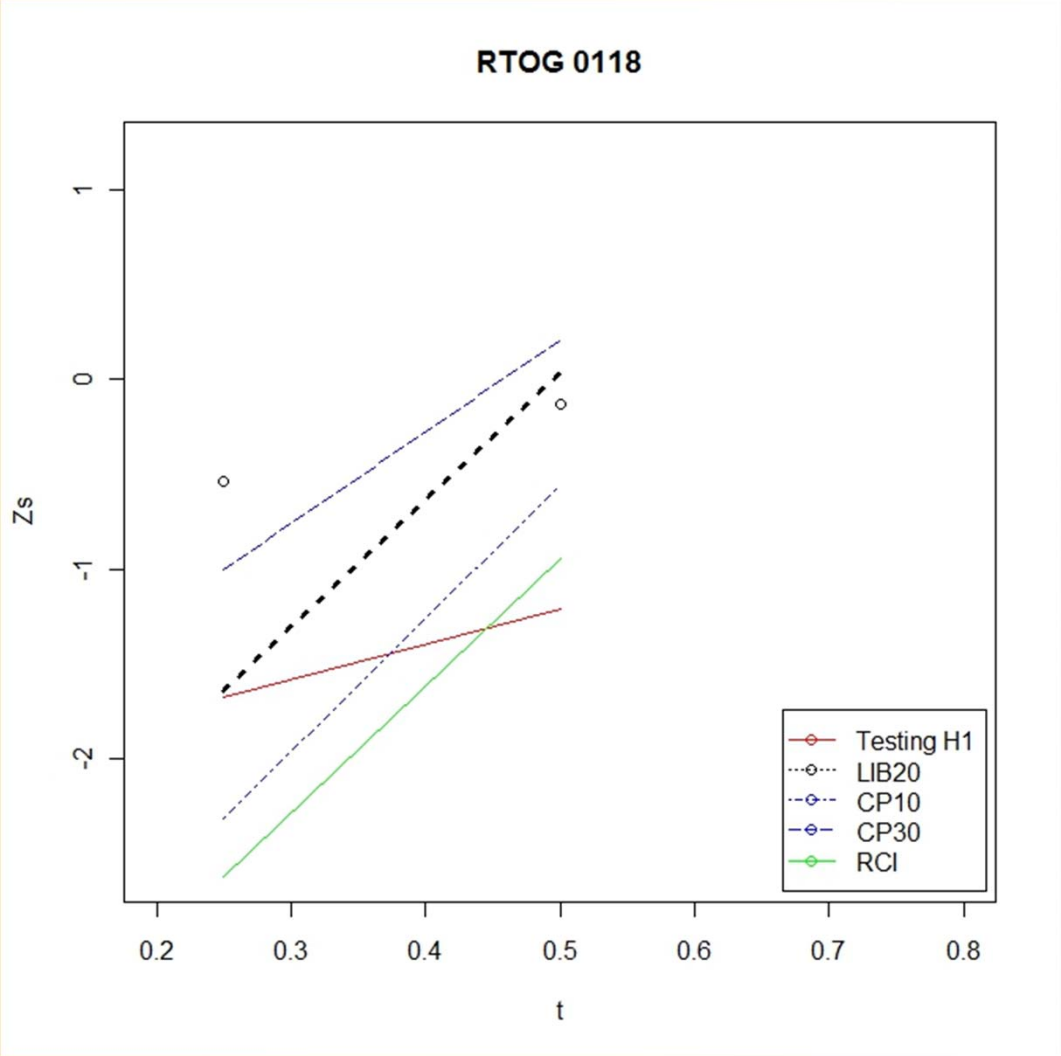
Comparing Futility Rules - Example



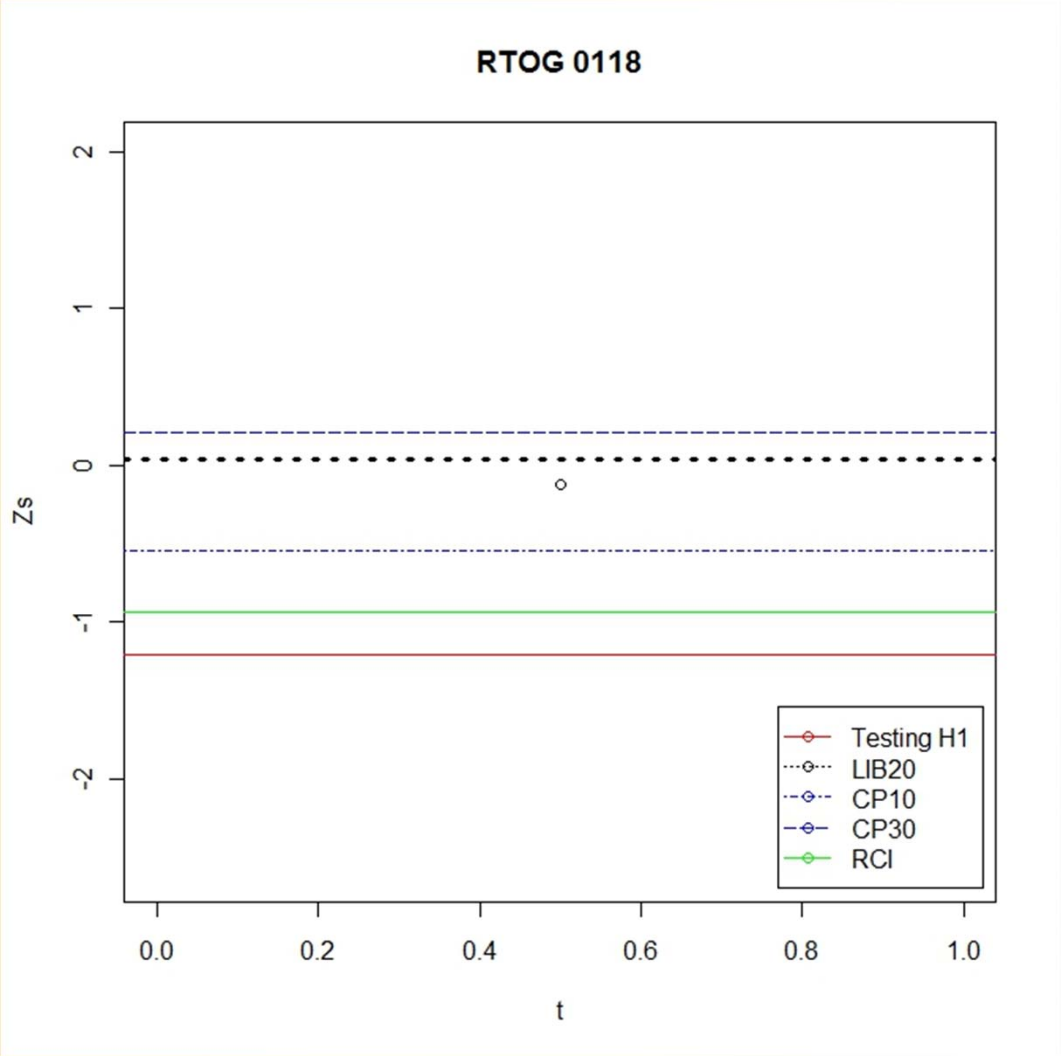
Comparing Futility Rules - Example



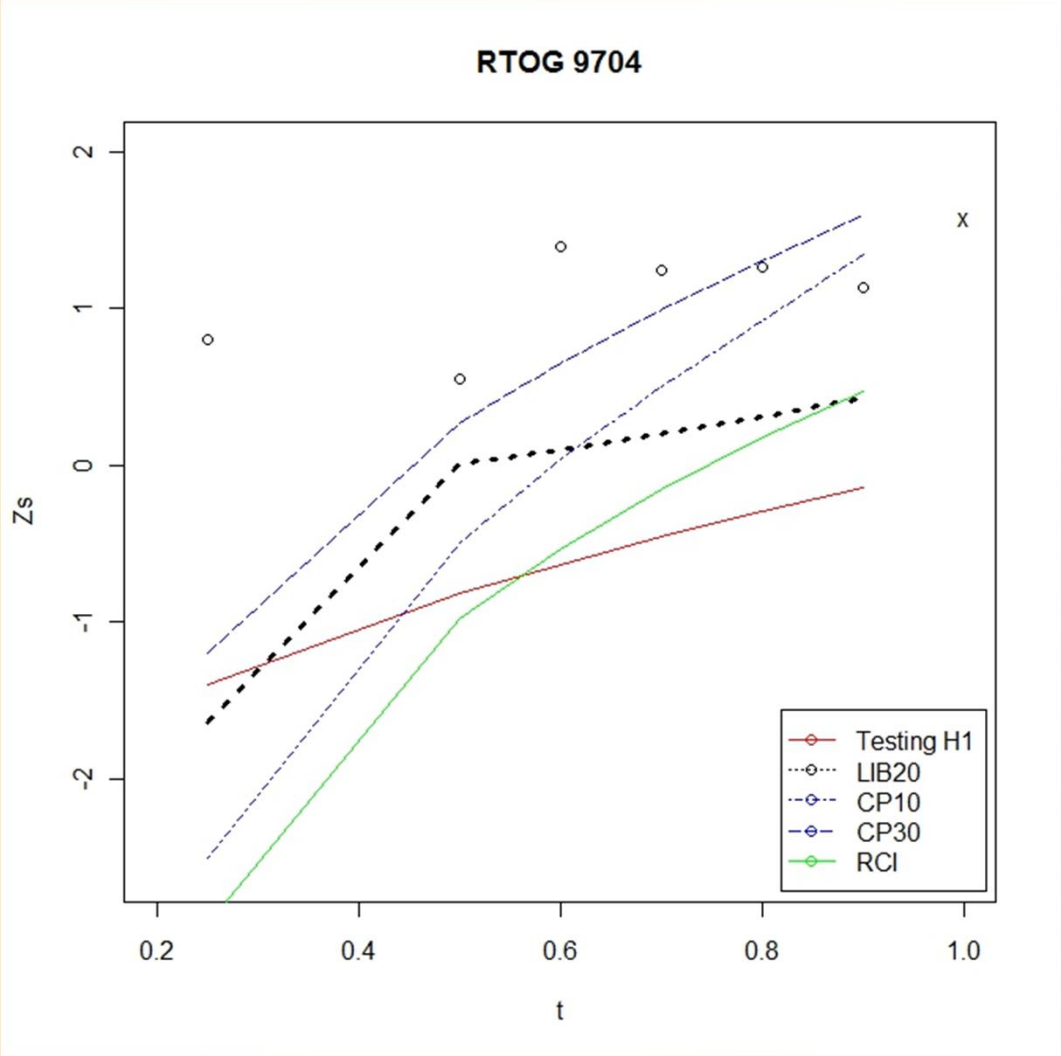
Comparing Futility Rules - Example



Comparing Futility Rules - Example



Comparing Futility Rules - Example



Results

Table 1. Trial time saved, 6/7 looks.

	Testing H1	LIB20	CP10	CP30	RCI
Stop no	0	0	0	0	0
Neutral	0	0.158	0.067	0.309	0
Stop yes	0.052	0.238	0.242	0.324	0.163
All trials	0.035	0.189	0.175	0.274	0.110

Table 2. Trial time saved, 3 looks.

	Testing H1	LIB20	CP10	CP30	RCI
Stop no	0	0	0	0	0
Neutral	0	0.158	0	0.337	0
Stop yes	0.024	0.220	0.113	0.286	0.138
All trials	0.016	0.177	0.076	0.254	0.093

Table 3. Trial time saved, 1 look.

	Testing H1	LIB20	CP10	CP30	RCI
Stop no	0	0	0	0	0
Neutral	0	0	0	0.127	0
Stop yes	0.014	0.146	0.080	0.193	0.014
All trials	0.010	0.099	0.054	0.153	0.010

Results

Table 4. Information time saved, 6/7 looks.

	Testing H1	LIB20	CP10	CP30	RCI
Stop no	0	0	0	0	0
Neutral	0	0.2	0.033	0.333	0
Stop yes	0.038	0.256	0.225	0.347	0.121
All trials	0.024	0.200	0.150	0.282	0.077

Table 5. Information time saved, 3 looks.

	Testing H1	LIB20	CP10	CP30	RCI
Stop no	0	0	0	0	0
Neutral	0	0.2	0	0.367	0
Stop yes	0.028	0.233	0.108	0.282	0.102
All trials	0.018	0.185	0.069	0.246	0.065

Table 6. Information time saved, 1 look.

	Testing H1	LIB20	CP10	CP30	RCI
Stop no	0	0	0	0	0
Neutral	0	0	0	0.167	0
Stop yes	0	0.111	0.047	0.205	0
All trials	0	0.071	0.030	0.161	0

Results

Table 7. Sample size saved, 6/7 looks.

	Testing H1	LIB20	CP10	CP30	RCI
Stop no	0	0	0	0	0
Neutral	0	0.252	0	0.270	0
Stop yes	0.006	0.060	0.018	0.062	0.012
All trials	0.004	0.074	0.013	0.078	0.008

Table 8. Sample size saved, 3 looks.

	Testing H1	LIB20	CP10	CP30	RCI
Stop no	0	0	0	0	0
Neutral	0	0.252	0	0.284	0
Stop yes	0.006	0.053	0.006	0.045	0.009
All trials	0.004	0.069	0.004	0.067	0.006

Table 9. Sample size saved, 1 look.

	Testing H1	LIB20	CP10	CP30	RCI
Stop no	0	0	0	0	0
Neutral	0	0	0	0.208	0
Stop yes	0	0.011	0.005	0.040	0
All trials	0	0.008	0.004	0.054	0

Discussion

- More frequent monitoring saves more on study duration, number of events and sample sizes
- LIB20 saves more relative to commonly used rules for all analysis schedules
- 10% Conditional power rules have similar savings for frequent monitoring, but conditional power rules are aggressive at the end
- Testing H1 and RCI methods are more conservative
- LIB20 is a rule with good properties across broad applications
  - Lessens patients' exposure to inactive treatment
  - Improves resource utilization
  - Accelerates dissemination of important clinical information

Discussion – Additional Considerations

- Futility rules affect error rates and power –adjust for in sample size
- Futility rules should be formulated based on trial design needs
- Futility stopping must consider other beneficial aspect of the trial
- Futility stopping also depends on study team guidance and DMC input

Acknowledgement

- Boris Freidlin, PhD, Biometric Research Branch, NCI
- Ed Korn, PhD, Biometric Research Branch, NCI
- Jim Dignam, PhD, University of Chicago, RTOG
- RTOG Statistics Department

References

1. Snappin S., Chen M.G., Jiang Q., Koutsoukos T., (2006), Assessment of futility in clinical trials, *Pharmaceutical statistics*, 5:273-281.

2. DeMets D.L., (2006), Futility approaches to interim monitoring by data monitoring committees. Clinical Trials 3, 522-529.

3. DeMets D.L., Ware J.H. (1980). Group sequential methods for clinical trials with a one-sided hypothesis. *Biometrika* 67, 651–660.

4. DeMets D.L., Ware J.H. (1982). Asymmetric group sequential boundaries for monitoring clinical trials. *Biometrika* 69, 661–663.

5. Pampallona S., Tsiatia A.A. (1994). Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the null hypothesis. *Journal of Statistical Planning and Reference* 42, 19–35.

6. Wang S.K., Tsiatis A.A. (1987). Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics* 43, 193-199.

7. Whitehead J., Stratton I. (1983). Group sequential clinical trials with triangular continuation region. *Biometrics* 39, 227-236.

8. Chang M.N., Hwang I.K. and Shih W.J. (1998). Group sequential designs using both type I and type II error probability spending functions. *Communications in Statistics - Theory and Methods* 27, 1323–1339.

9. Cook T.D., DeMets D.L.,(2007), Introduction to statistical methods for clinical trials. Chapman and Hall/CRC.

10. Freidlin B., Korn E.L., (2002), A comment on futility monitoring. *Controlled Clinical Trials* 23, 355-366.

11. Freidlin B., Korn E.L., Gray R., (2010), A general inefficacy interim monitoring rule for randomized clinical trials. *Clinical Trials* 7, 197-208.

12. Ellenberg SS, Eisenberger MA. An efficient design for phase III studies of combination chemotherapies. Cancer Treat Rep 1985;69:1147–1152.

13. Wieand S, Schroeder G, O'Fallon JR. Stopping when the experimental regimen does not appear to help. Stat Med 1994;13:1453–1458.

References

14. Jennison C, Turnbull BW. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC, Boca Raton, 2000.

15. Freidlin B, Korn EL. Monitoring for lack of benefit: a critical component of a randomized clinical trial. J Clin Oncol 2009; 27: 629–33.

16. Fu, K., T. Pajak, et al. (2000). "A Radiation Therapy Oncology Group (RTOG) Phase III Randomized Study to Compare Hyperfractionation and Two Variants of Accelerated Fractionation to Standard Fractionation Radiotherapy for Head and Neck Squamous Cell Carcinomas: First Report of RTOG 90-03." *Int J Rad Oncol Biol Phys* 48(1): 7-16.

17. Forastiere, A., H. Goepfert, et al. (2003). "Concurrent Chemotherapy and Radiotherapy for Organ Preservation in Advanced Laryngeal Cancer." *N Engl J Med* 349(22): 2091-2098.

18. Cooper, J., T. Pajak, et al. (2004). "Postoperative Concurrent Radiotherapy and Chemotherapy for High-Risk Squamous-Cell Carcinoma of the Head and Neck." *N Engl J Med* 350(19): 1937-1944.

19. Ang, K., J. Harris, et al. (2010). "Human Papilloma Virus and survival of patients with oropharyngeal cancer." *N Engl J Med* 363(1): 24-35.

20. Hanks, G., T. Pajak, et al. (2003). "Phase III Trial of Long-Term Adjuvant Androgen Deprivation After Neoadjuvant Hormonal Cytoreduction and Radiotherapy in Locally Advanced Carcinoma of the Prostate: The Radiation Therapy Oncology Group Protocol 92-02." *J Clin Oncol* 21(21): 3972-3978.

21. Roach, M., M. DeSilvio, et al. (2003). "Phase III Trial Comparing Whole-Pelvic Versus Prostate-Only Radiotherapy and Neoadjuvant Versus Adjuvant Combined Androgen Suppression: Radiation Therapy Oncology Group 94-13." *J Clin Oncol* 21(10): 1904-1911.

22. Albain, K. S., S. Swann, et al. (2009). "Phase III Comparison of Concurrent Chemotherapy Plus Radiotherapy (CT/RT) and CT/RT Followed by Surgical Resection for Stage IIIA (Pn2) Non-Small Cell Lung Cancer (NSCLC): Initial Results From Intergroup Trial 0139 (RTOG 93-09)." The Lancet 374(9687): 379-386.

References

23. Movsas, B., C. Scott, et al. (2005). "Randomized Trial of Amifostine in Locally Advanced Non-Small-Cell Lung Cancer Patients Receiving Chemotherapy and Hyperfractionated Radiation: Radiation Therapy Oncology Group Trial 98-01." *J Clin Oncol* 23(10): 2145-2154.

24. Cairncross, G., B. Berkey, et al. (2006). "Phase III Trial of Chemotherapy Plus Radiotherapy Compared with Radiotherapy Alone for Pure and Mixed Anaplastic Oligodendroglioma: Intergroup Radiation Therapy Oncology Group Trial 9402." *J Clin Oncol* 24(18): 2707-2714.

25. Knisely, J., B. Berkey, et al. (2008). "A Phase III Study of Conventional Radiation Therapy Plus Thalidomide vs. Conventional Radiation Therapy for Multiple Brain Metastases (RTOG 0118)." *Int J Radiat Oncol Biol Phys* 71(1): 79-86.

26. Regine, W., K. Winter, et al. (2008). "Fluorouracil vs Gemcitabine Chemotherapy Before and After Fluorouracil-Based Chemotherapy Following Resection of Pancreatic Adenocarcinoma: A Randomized Controlled Trial." *JAMA* 299(9): 1019-1026.

27. J.J. Dignam et al. (1998) Early stopping of a clinical trial when there is evidence of no treatment benefit: protocol B-14 of the National Surgical Adjuvant Breast and Bowel Project. Controlled Clinical Trials 19:575-588.

28. D.R. Bristol (1989) Designing clinical trials for two sided multiple comparisons with a control. Controlled Clinical Trials 10:142-15

29. Fleming TR, Harrington DP, O'Brien PC. (1984) Designs for group sequential tests. Controlled Clinical Trials 5:348-361