A constrained boundaries approach for monitoring time to event outcomes

Presented March 17, 2008

Dan Gillen
Department of Statistics
University of California, Irvine

Joint work with
Sean Brummel, Dept of Statistics

UCIrvine
University of California, Irvine

Group sequential testing in survival trials

Group sequential testing in clinical trials

- Analyze the data at periodic intervals
- Test times defined on 'information' scale
- Study is stopped as soon as confidence in favor of a decision is reached
  - P-value
  - Predictive probability of final statistic
  - Stochastic curtailment
  - Bayesian posterior probability
  - Likelihood support intervals
- Multiple testing of data using fixed sample critical values results in an inflated type I error

UCIrvine
University of California, Irvine

Operating characteristics to consider at the design stage

1. Efficiency of designs
   - Power at various alternatives
   - Average sample number (ASN) / stopping probabilities
2. Point estimates of treatment effect corresponding to boundary decisions in favor of
   - Efficacy – Futility – Harm
3. Frequentist/Bayesian/Likelihood inference on the boundaries
4. Conditional futility/reversal of decision corresponding to boundary decisions

UCIrvine
University of California, Irvine

Group sequential testing in survival trials

Operating characteristics condition upon exact timing

- During the conduct of a study statistical information from a sampling unit may be different than originally estimated
  - Variance of measurements
  - Baseline event rates (binary outcomes)
  - Censoring and survival distributions (weighted survival statistics)
- Consequences of monitoring under incorrect estimates of statistical information can include
  - Change in nominal type I error rate from originally planned design
  - Change in power from originally planned design

UCIrvine
University of California, Irvine

Monitoring group sequential trials

Popular methods for flexible implementation of group sequential boundaries

1. Christmas tree approximation for triangular tests: Whitehead and Stratton (1983)
2. Error spending functions: Lan and DeMets (1983); Pampallona, Tsiatis, and Kim (1995)
3. Constrained boundaries in unified design family: Emerson (2000); Burrington & Emerson (2003)

Monitoring group sequential trials

Common features

- Stopping rule specified at design stage parameterizes the boundary for some statistic (boundary scale)
  - Error spending family (Lan & Demets, 1983) $\longrightarrow$ proportion of type I error spent
  - Unified family (Emerson & Kittelson, 1999) $\longrightarrow$ point estimate (MLE)
- At the first interim analysis, parametric form is used to compute the boundary for actual time on study
- At successive analyses, the boundaries are recomputed accounting for the exact boundaries used at previously conducted analyses
- Maximal sample size estimates may be updated to maintain power

UCIrvine
University of California, Irvine

Use of constrained boundaries in flexible implementation of stopping rules

1. At the first analysis, compute stopping boundary (on some scale) from parametric family

2. At successive analyses, use parametric family with constraints (on some scale) for the previously conducted interim analyses

- When the error spending scale is used, this is just the error spending approach of Lan & DeMets (1983) or Pampallona, Tsiatis, & Kim (1995)

Group sequential testing in survival trials

Further considerations when considering survival endpoints

- Common to use the logrank statistic for testing survival differences
  - Locally efficient for proportional hazards alternatives
- In this case, translation between sample size and statistical information is trivial
  - Information is proportional to the number of observed events

UCIrvine
University of California, Irvine

Weighted LR statistics

$G^{\rho,\gamma}$ statistic

- When a non-proportional hazards treatment effect is hypothesized some have suggested the use of weighted logrank statistics
  - Potential for increased power by up-weighting areas of survival where largest (most clinically relevant?) effects are hypothesized to occur
- $G^{\rho,\gamma}$ family of weighted logrank statistics (Fleming & Harrington, 1991)

$$G^{\rho,\gamma} = \left(\frac{M_1 + M_0}{M_1 M_0}\right)^{1/2} \int_0^\infty w(t) \left\{\frac{Y_1(t) Y_0(t)}{Y_1(t) + Y_0(t)}\right\} \left\{\frac{dN_1(t)}{Y_1(t)} - \frac{dN_0(t)}{Y_0(t)}\right\}$$

with

$$w(t) = [\hat{S}(t-)]^\rho [1 - \hat{S}(t-)]^\gamma$$

Weighted LR statistics

$G^{\rho,\gamma}$ statistic

- Can be rewritten as the sum, over all failure times, of the weighted difference in estimated hazards

$$G^{\rho,\gamma} = \left(\frac{M_1 + M_0}{M_1 M_0}\right)^{1/2} \sum_{t \in \mathcal{F}} w^*(t) \left[\hat{\lambda}_1(t) - \hat{\lambda}_0(t)\right]$$

with $\hat{\lambda}_i = dN_i(t)/Y_i(t)$ and

$$w^*(t) = \left\{\frac{Y_1(t)Y_0(t)}{Y_1(t) + Y_0(t)}\right\} [\hat{S}(t-)]^{\rho}[1 - \hat{S}(t-)]^{\gamma}$$

UCIrvine
University of California, Irvine

Information growth for the $G^{\rho,\gamma}$ family

Information growth for the $G^{\rho,\gamma}$ family

- Under the null hypothesis $H_0 : S_0 = S_1$, the variance of the $G^{\rho,\gamma}$ statistic calculated at calendar time τ reduces to

$$\sigma^2 \propto \int_0^{\tau} w^2(t) F_E(\tau - t)[1 - F_C(t)] dS(t)$$

- Let σ_j^2 equal the estimated variance of the $G^{\rho,\gamma}$ statistic applied at interim analysis j. Then the proportion of information at analysis j, relative to the maximal analysis J, is given by

$$\prod_j \equiv \left(\frac{M_{1,j} + M_{0,j}}{M_{1,j} M_{0,j}} \right)^{-1} \sigma_j^2 \Bigg/ \left(\frac{M_{1,J} + M_{0,J}}{M_{1,J} M_{0,J}} \right)^{-1} \sigma_J^2,$$

UCIrvine
University of California, Irvine

Introduction
Group sequential operating characteristics
Constrained boundaries approach to monitoring
Monitoring Survival Trials
Weighted LR statistics
Information growth for weighted LR statistics
Ex: Sensitivity of operating characteristics to the censoring distribution
Implementation of group sequential rules
Proposed algorithm for monitoring
Simulation study
Summary

Information growth for the $G^{\rho,\gamma}$ family

Example: Information Growth for the $G^{1,0}$ and $G^{1,1}$ statistics

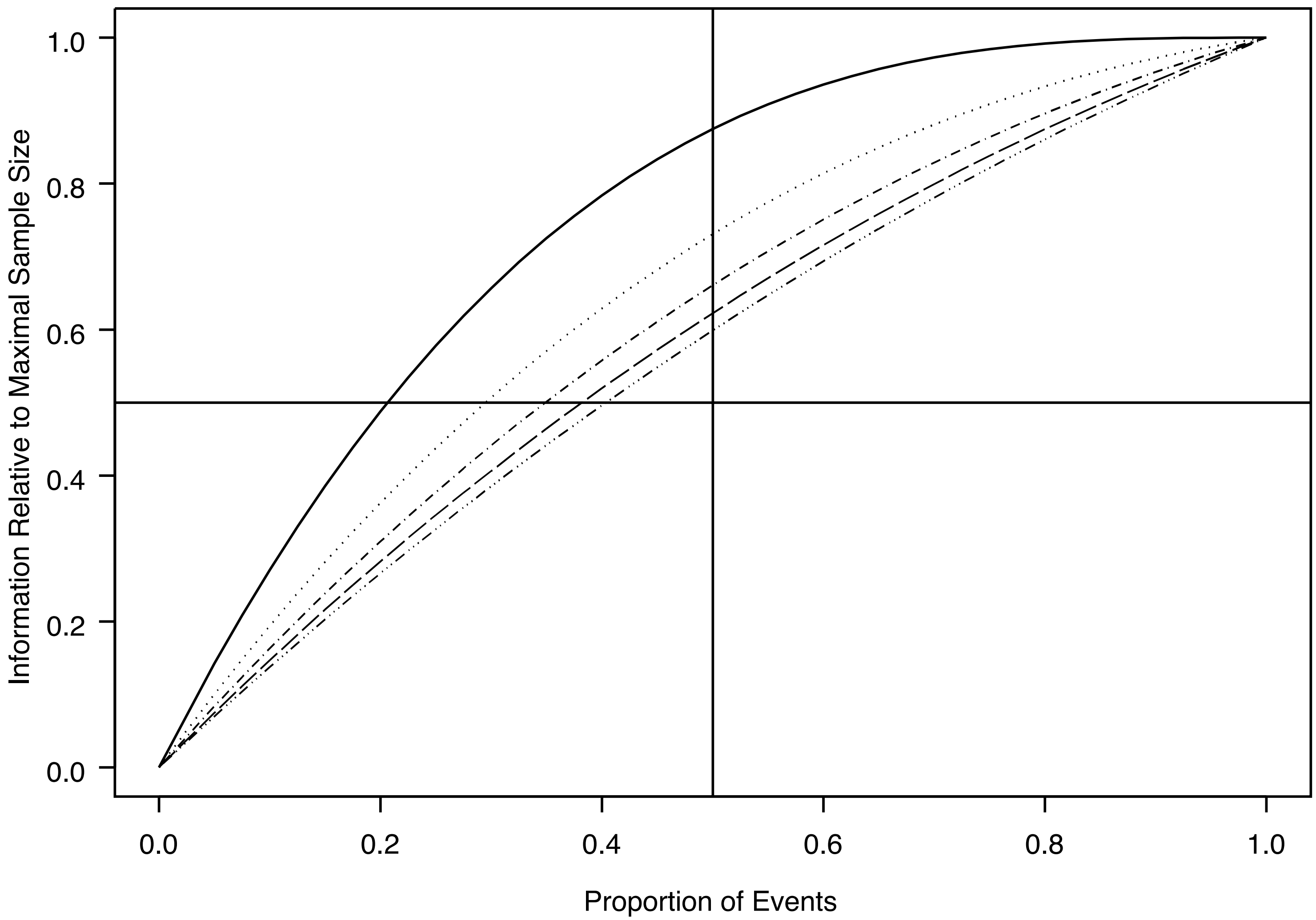
- Consider information growth for the $G^{1,0}$ and $G^{1,1}$ statistics as a function of observed events
- Assume
  - $S_1(t)$ and $S_0(t)$ are Exponential(1)
  - Assume accrual follows a “powered uniform" distribution

$$F_E(t) = \left(\frac{t}{\theta}\right)^r, \text{ with } \theta > 0, r > 0, 0 < t \leq \theta$$

    - Enrollment occurs over interval $(0, \theta)$
    - $r = 1 \Rightarrow$ Unif(0,θ) enrollment
    - $r \rightarrow 0 \Rightarrow$ Instantaneous enrollment at time 0
    - $r \rightarrow \infty \Rightarrow$ Instantaneous enrollment at time θ

Example: Difference in Information by Accrual for the $G^{1,0}$ Statistic

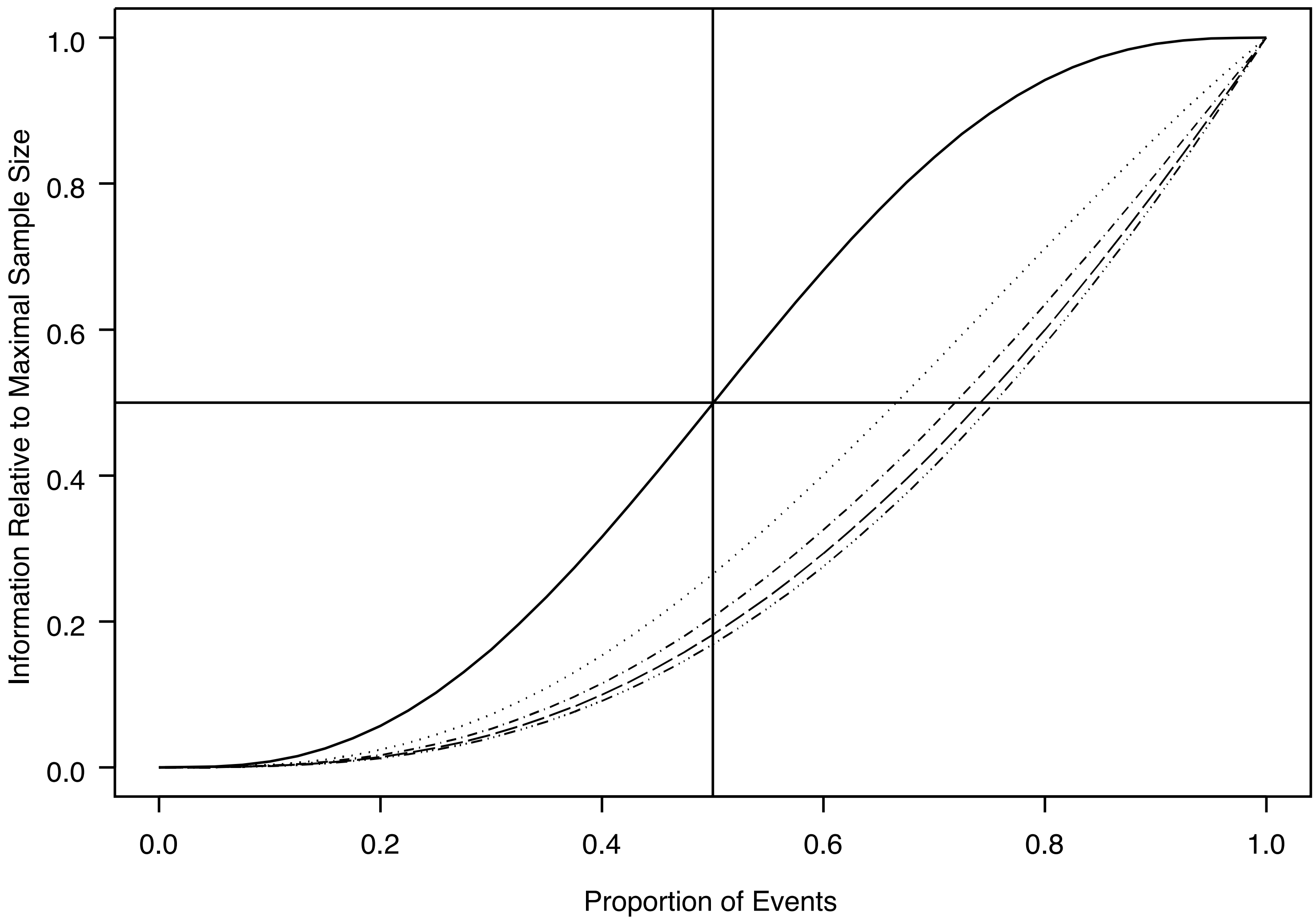
Effect of total censoring: No censoring (solid line) to 66% censoring



UCIrvine
University of California, Irvine

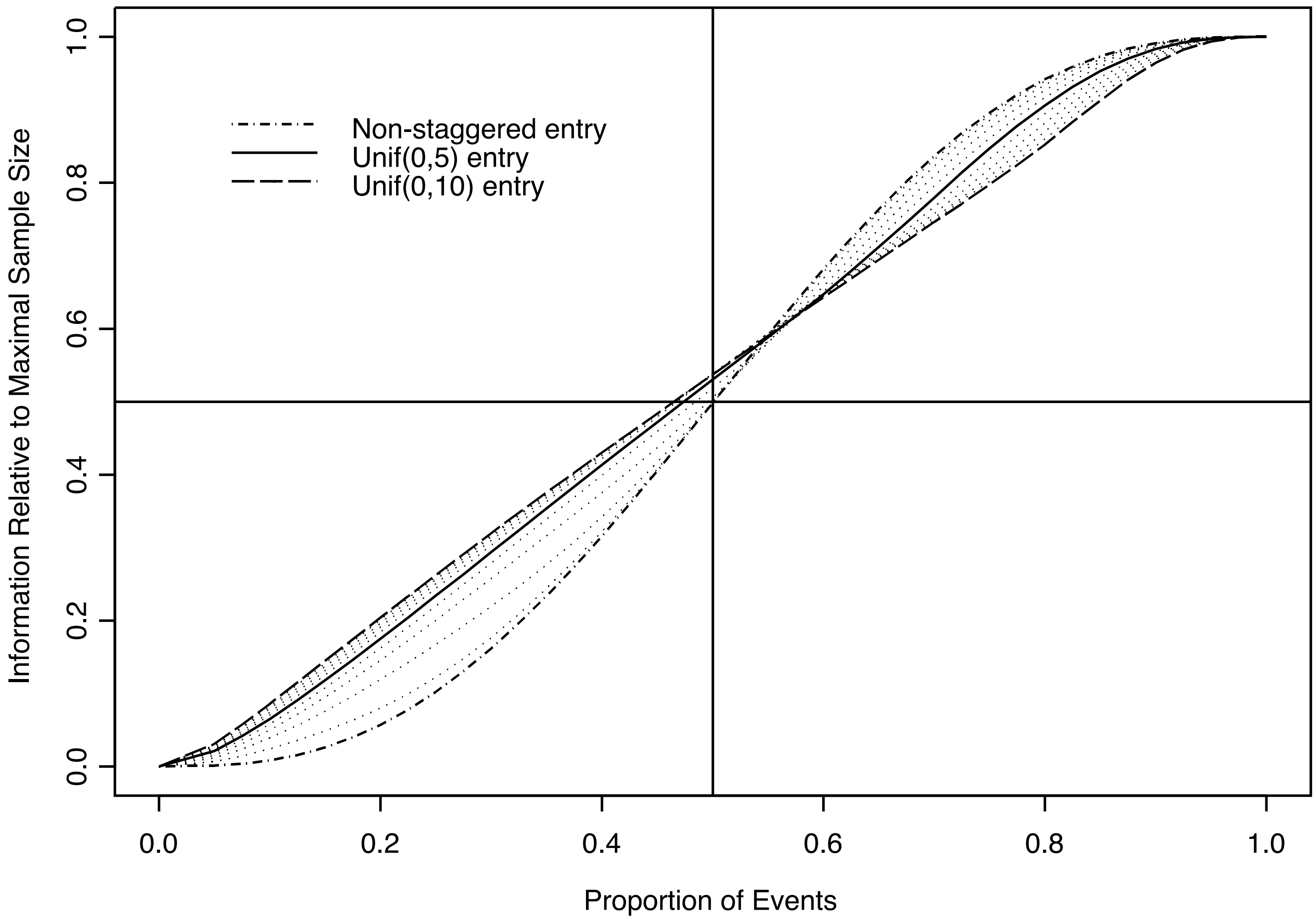
Example: Difference in Information by Accrual for the $G^{1,1}$ Statistic

Effect of total censoring: No censoring (solid line) to 66% censoring



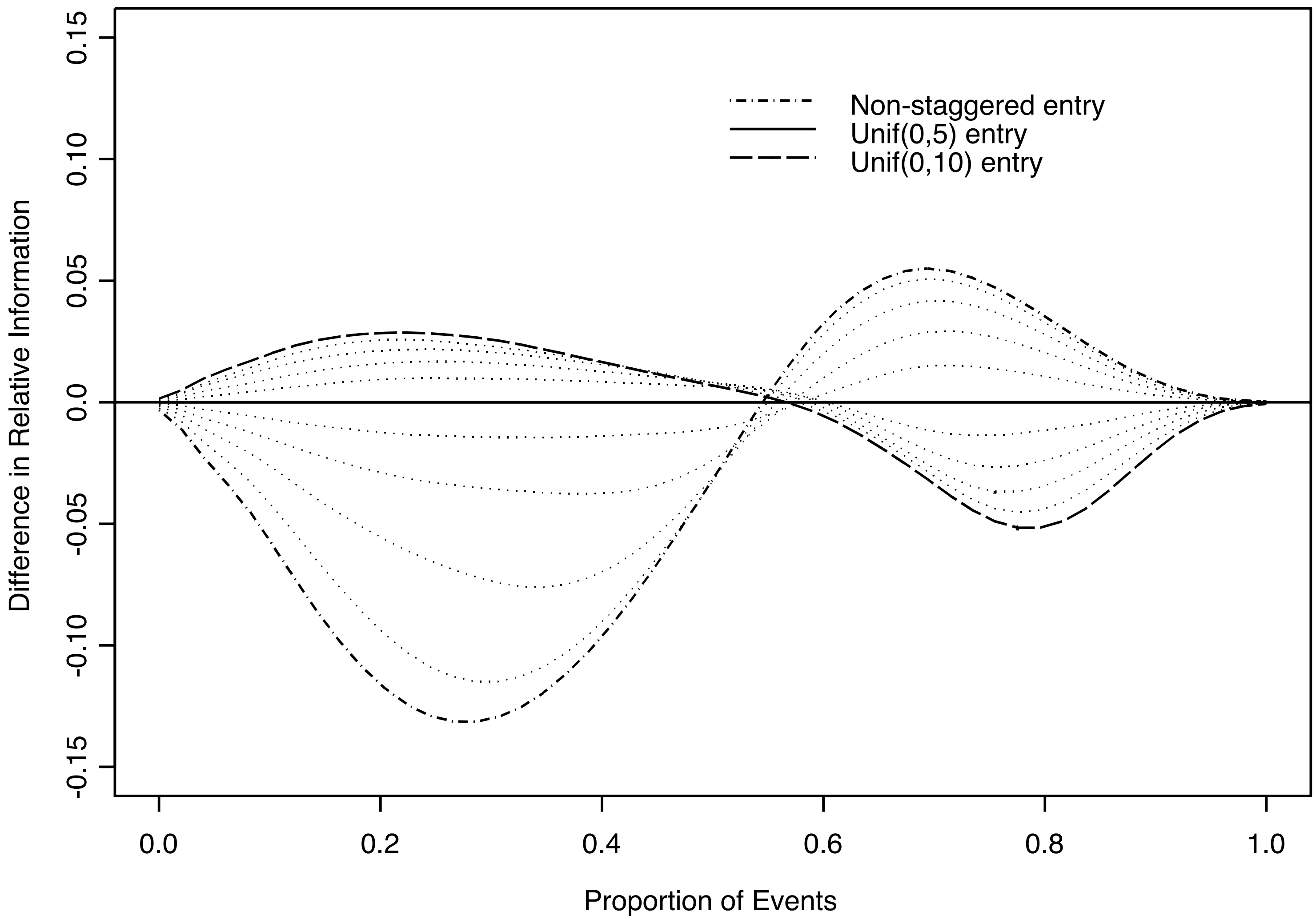
Example: Information Growth for the $G^{1,1}$ Statistic

Uniform accrual with no administrative censoring



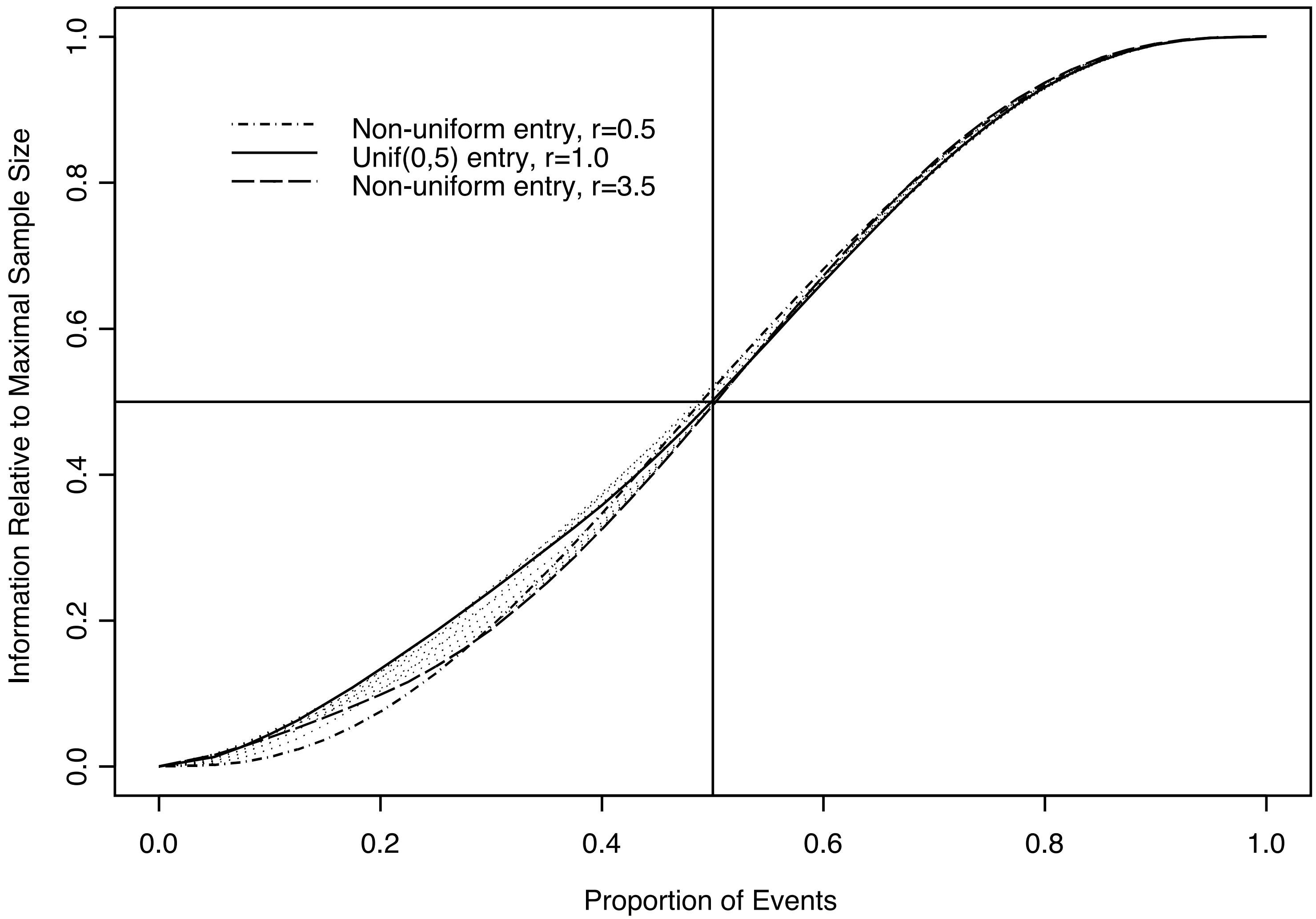
Example: Difference in Information by Accrual for the $G^{1,1}$ Statistic

Uniform accrual with no administrative censoring



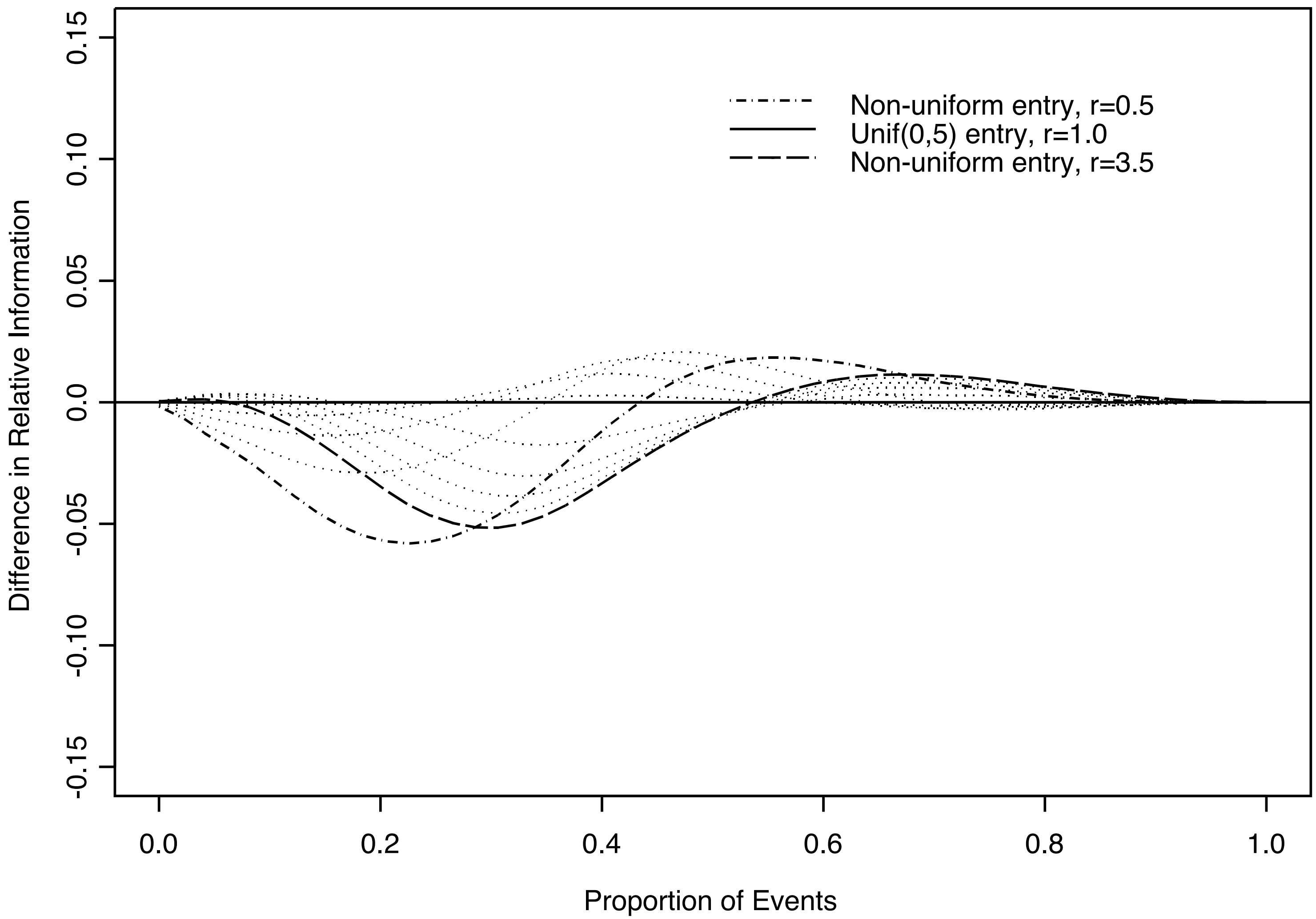
Example: Information Growth for the $G^{1,1}$ Statistic

Nonuniform accrual with no administrative censoring



Example: Difference in Information by Accrual for the $G^{1,1}$ Statistic

Nonuniform accrual with no administrative censoring



Example: Operating characteristics with misspecified accrual distribution

Example: Operating characteristics when testing with the $G^{1,1}$ Statistic

- Design
  - One-sided level .05 test
  - O'Brien-Fleming efficacy bound; Pocock futility bound
  - 4 analyses occurring at proportional information of .25, .50, .75, and 1
  - Power of .90 at alternative HR of .75 $\rightarrow$ 507 max events
- Assumed survival and accrual distributions
  - Pooled survival distributed Exponential(.4)
  - Accrual uniform over 3 years
- Suppose true accrual is uniform over 1 year

Example: Operating characteristics with misspecified accrual distribution

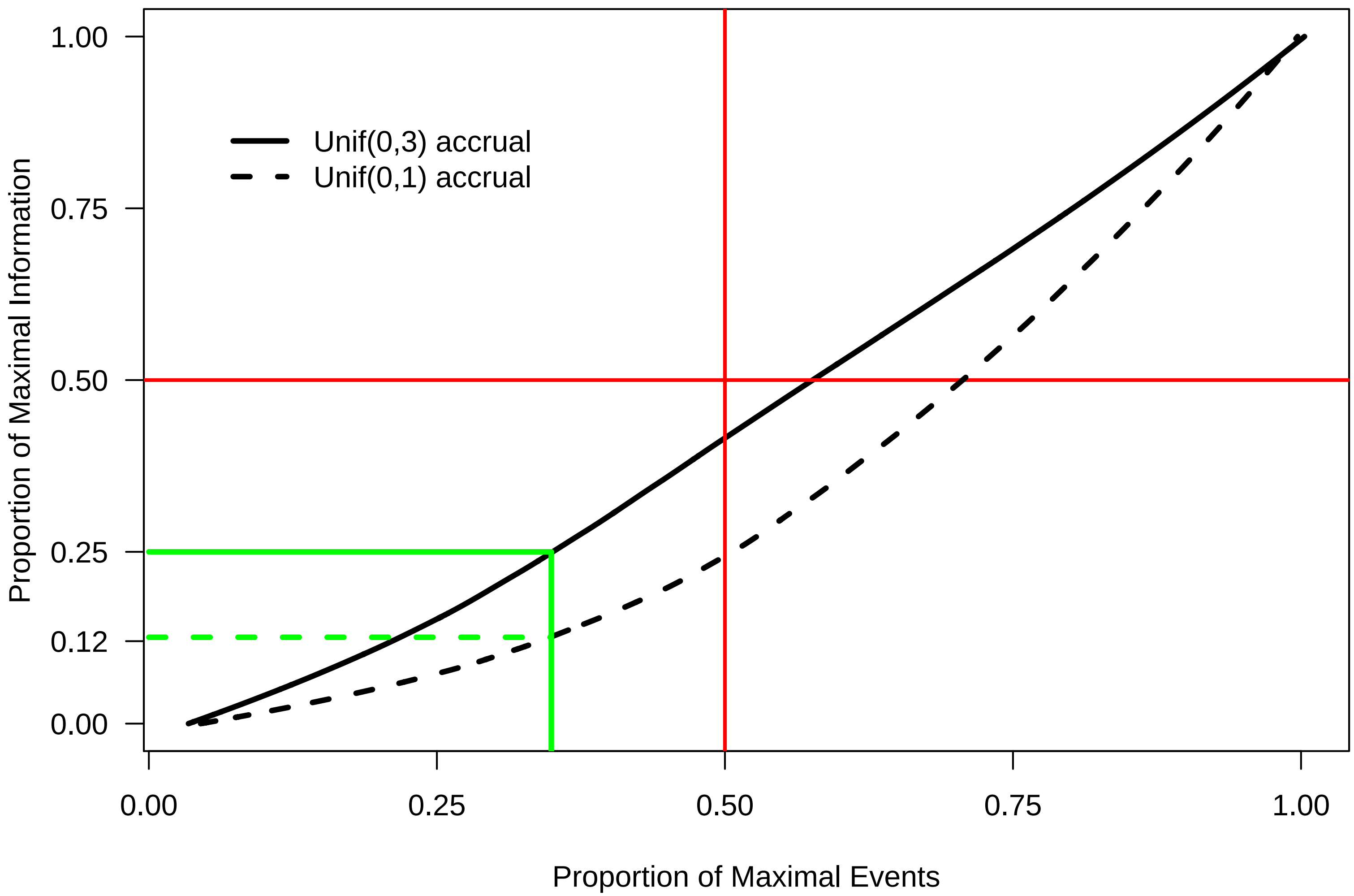
Example: Operating characteristics when testing with the $G^{1,1}$ Statistic

- Stopping boundaries for original design on Z-statistic scale

```
STOPPING BOUNDARIES: Normalized Z-value scale
                              efficacy  futility
    Time 1 (Pi_1= 0.25) -3.2642    0.2094
    Time 2 (Pi_2= 0.50) -2.3082   -0.5534
    Time 3 (Pi_3= 0.75) -1.8846   -1.1387
    Time 4 (Pi_4= 1.00) -1.6321   -1.6321
```

UCIrvine
University of California, Irvine

Example: Operating characteristics with misspecified accrual distribution



Example: Operating characteristics with misspecified accrual distribution

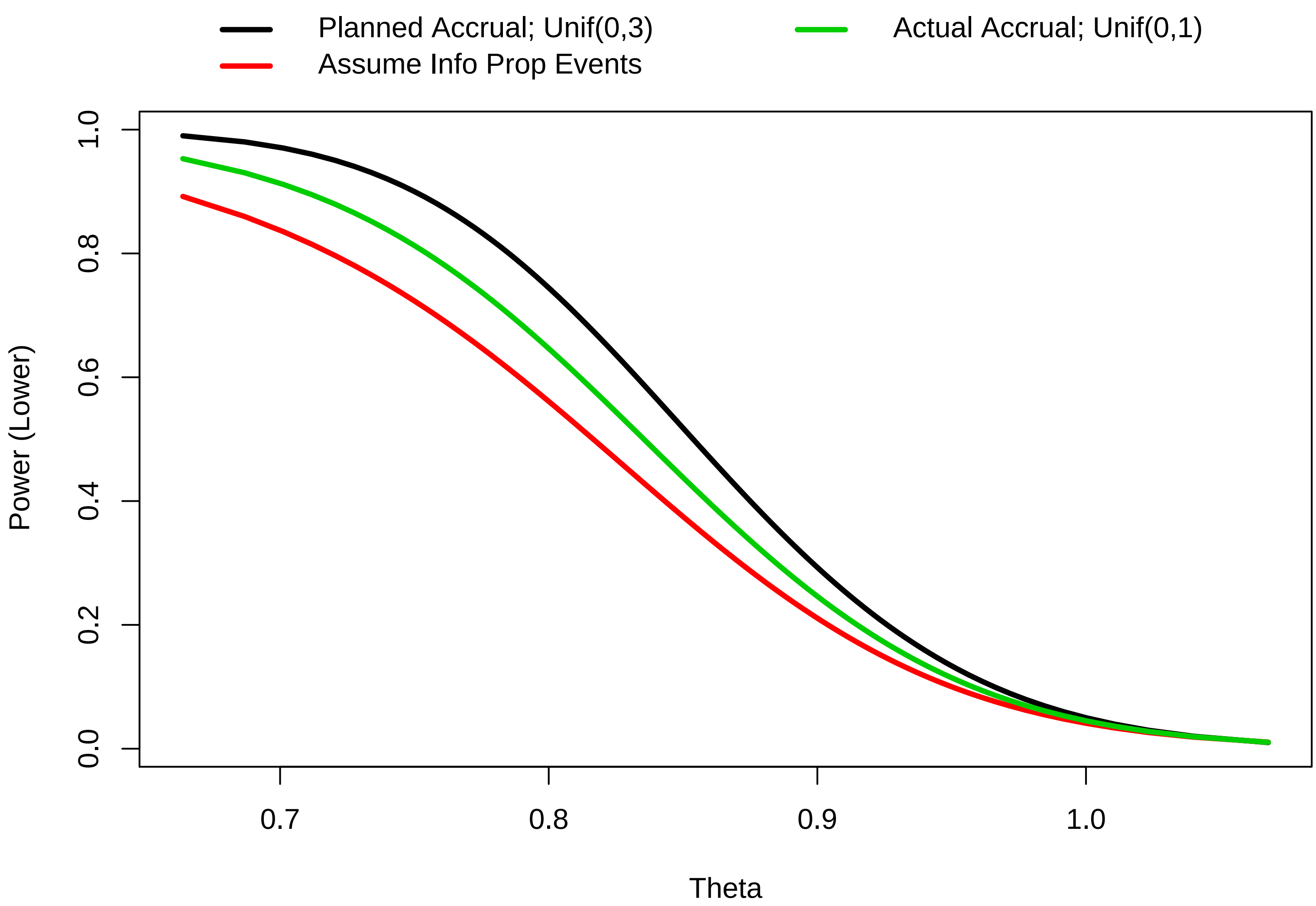
UCIrvine
University of California, Irvine

Example: Operating characteristics when testing with the $G^{1,1}$ Statistic

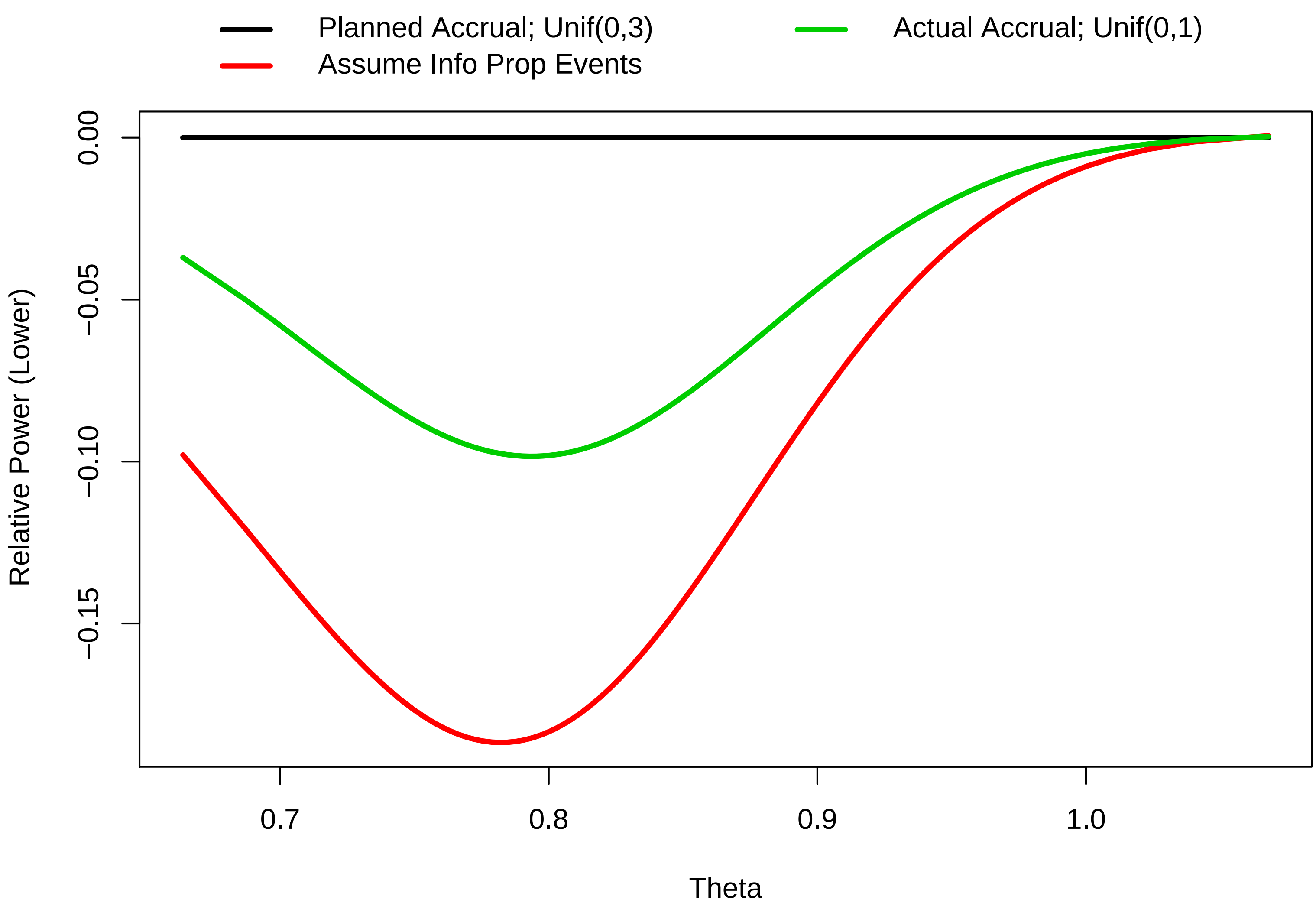
- Stopping boundaries if Unif(0,3) accrual assumed, but true accrual Unif(0,1)

```
STOPPING BOUNDARIES: Normalized Z-value scale
                          efficacy  futility
    Time 1 (Pi_1= 0.12) -3.2642    0.2094
    Time 2 (Pi_2= 0.36) -2.3082   -0.5534
    Time 3 (Pi_3= 0.66) -1.8846   -1.1387
    Time 4 (Pi_4= 1.00) -1.6321   -1.6321
```

Example: Operating characteristics with misspecified accrual distribution



Example: Operating characteristics with misspecified accrual distribution



Implementation of group sequential rules

Goal: Maintain operating characteristics to be as close to design stage as possible

1. Need to choose between
   - maintaining maximal statistical information
   - maintaining statistical power
2. In addition, need to update our estimate of the information growth curve at each analysis
   - requires updating our estimate of $S(t)$ and $F_E(t)$ at each analysis

Implementation of group sequential rules

Proposed algorithm: Step 1

1. Specify original design using a parametric design family to satisfy desired operating characteristics
   1.1 specify timing of analyses
   1.2 assume $S(t)$ and $F_E(t)$
   1.3 estimate information growth curve
   1.4 map information increments to proportion of events for desired timing of first analysis

UCIrvine
University of California, Irvine

Proposed algorithm: Step 2

2. At first analysis,

   2.1 estimate $S(t)$ and $F_E(t)$ via parametric model

   - Use pooled data so that constraint does not depend on observed treatment effect
   - Fit weibull and powered uniform distribution, respectively

   2.2 re-estimate information growth curve

   2.3 map information increments to proportion of events for desired timing of future analyses

   2.4 constrain first boundary to exact timing (based upon current best estimate) and re-estimate future boundaries using pre-specified design family

UCIrvine
University of California, Irvine

UCIrvine
University of California, Irvine

Proposed algorithm: Step 3

3. At future analyses,

    3.1 re-estimate $S(t)$ and $F_E(t)$ via parametric model available data up to the analysis

    3.2 re-estimate information growth curve

    3.3 map information increments to proportion of events for desired timing of future analyses

    3.4 constrain previous boundaries to exact timing (based upon current best estimate) and re-estimate future boundaries using pre-specified design family

Simulation framework

- Design
  - One-sided level .05 test
  - O'Brien-Fleming efficacy bound; Pocock futility bound
  - 4 analyses occurring at proportional information of .25, .50, .75, and 1
  - Power of .90 at alternative HR of .75 $\rightarrow$ 507 max events
- Assumed survival and accrual distributions
  - Pooled survival distributed Exponential(.4)
  - Accrual uniform over 3 years

Simulation study

Simulation framework

- Consider Type I error and power under varying accrual distributions
  - Entry time distribution

$$F_E(t) = \left(\frac{t}{\theta}\right)^r, \text{ with } \theta > 0, r > 0, 0 < t \le \theta$$

  - Values of θ considered: 1, 3, 5
  - Values of r considered: .50, .75, 1, 3, 5

- Comparison of monitoring plans
  - Naive: Assume information proportional to number of events
  - Assumed: Assume information curve obtained under Unif(0,3) accrual
  - Constrained boundaries to maintain power and type I error: Update estimated accrual distribution and re-estimate information curve at each analysis

Simulation Study: Accrual over 1 year

Monitoring Strategy	r: 0.5	0.75	1.0	3.0	5.0
HR=1					
Info $\propto$ Events	0.0408	0.0408	0.0408	0.0397	0.0381
Accrual Unif(0,3)	0.0452	0.0452	0.0457	0.0445	0.0439
Constrained Bound	0.0501	0.0501	0.0501	0.0500	0.0500
HR=0.875					
Info $\propto$ Events	0.2858	0.2858	0.2862	0.2725	0.2599
Accrual Unif(0,3)	0.3406	0.3400	0.3449	0.3353	0.3314
Constrained Bound	0.4045	0.4047	0.4041	0.4048	0.4047
HR=0.75					
Info $\propto$ Events	0.7189	0.719	0.7198	0.6917	0.6619
Accrual Unif(0,3)	0.8189	0.818	0.8261	0.8084	0.8006
Constrained Bound	0.9018	0.9021	0.9011	0.9014	0.9012

UCIrvine
University of California, Irvine

Simulation Study: Accrual over 3 years

Monitoring Strategy	r 0.5	0.75	1.0	3.0	5.0
HR=1					
Info ∝ Events	0.0432	0.0461	0.0463	0.0463	0.0438
Accrual Unif(0,3)	0.0479	0.0499	0.0502	0.0489	0.0473
Constrained Bound	0.0504	0.0504	0.0503	0.0498	0.0500
HR=0.875					
Info ∝ Events	0.3273	0.3567	0.3556	0.3424	0.3129
Accrual Unif(0,3)	0.3793	0.4002	0.4048	0.3777	0.3583
Constrained Bound	0.4106	0.4108	0.4079	0.3997	0.4030
HR=0.75					
Info ∝ Events	0.7914	0.8415	0.8412	0.8256	0.7748
Accrual Unif(0,3)	0.8734	0.8998	0.9047	0.8753	0.8486
Constrained Bound	0.9087	0.9094	0.9074	0.8985	0.9008

Simulation Study: Accrual over 5 years

Monitoring Strategy	r: 0.5	0.75	1.0	3.0	5.0
HR=1					
Info $\propto$ Events	0.0460	0.0470	0.0474	0.0479	0.0464
Accrual Unif(0,3)	0.0502	0.0503	0.0508	0.0508	0.0492
Constrained Bound	0.0504	0.0503	0.0502	0.0499	0.0492
HR=0.875					
Info $\propto$ Events	0.3568	0.3694	0.3725	0.3678	0.3433
Accrual Unif(0,3)	0.4064	0.4079	0.4127	0.4057	0.3814
Constrained Bound	0.4131	0.4095	0.4090	0.3924	0.3916
HR=0.75					
Info $\propto$ Events	0.8409	0.8594	0.8644	0.8615	0.8270
Accrual Unif(0,3)	0.9060	0.9075	0.9127	0.9069	0.8799
Constrained Bound	0.9117	0.9088	0.9095	0.8934	0.8903

UCIrvine
University of California, Irvine

Summary

Concluding remarks

- Operating characteristics at the design stage are conditional upon exact timing of analyses (information time)
- With the usual logrank statistic, information is proportional to the observed number of events
- If a weighted logrank statistic is used, information growth depends upon both the entry and failure time distributions
  - Can no longer think of censoring as a nuisance variable
- Implementation of a design requires estimation of the entry and failure time distributions

Summary

Concluding remarks

- To allow for flexibility in actual timing of analyses, constrained boundaries can be employed
- To maintain some (but not all) operating characteristics of the originally planned design, re-estimation of the information growth curve is necessary
- Software available:
  - Constrained boundaries for unified family : SPlus SeqTrial
  - Mapping of statistical information : Me
- Future work:
  - Misspecification of parametric form of entry and survival distributions?
  - Non-stationarity in entry and survival distributions?

UCIrvine
University of California, Irvine