On the use of prior information for flexibly monitoring group sequential survival trials

Presented October 16, 2009

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

UCIrvine
University of California, Irvine

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 probabilities
- Multiple testing of data using fixed sample critical values results in an inflated type I error

Group sequential testing in survival trials

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 and Bayesian inference on the boundaries
4. Conditional futility/reversal of decision corresponding to boundary decisions

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
  1. Change in nominal type I error rate from originally planned design
  2. Change in power from originally planned design
  3. Change in ASN from originally planned design

UCIrvine
University of California, Irvine

Monitoring group sequential trials

UCIrvine
University of California, Irvine

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

Use of constrained boundaries in flexible implementation of stopping rules

1. At design stage, choose a parametric stopping rule:

   Ex. Unified family (Emerson & Kittelson, 1999) $\rightarrow$ point estimate (MLE)

$$\hat{\theta}_j < a_j = \left(A_a + \prod_j^{P_a} \left(1 - \prod_j \right)^{R_a} \right) G_a \qquad \text{(lower)}$$

$$\hat{\theta}_j > d_j = \left(A_d + \prod_j^{P_d} \left(1 - \prod_j \right)^{R_d} \right) G_d \qquad \text{(upper)}$$

   Notes:

   - Taking $P = 0.5, A = 0, R = 0$ yields a Pocock (1977) design
   - Taking $P = 1, A = 0, R = 0$ yields a O'Brien-Fleming (1979) design

Monitoring group sequential trials

Use of constrained boundaries in flexible implementation of stopping rules

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

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

Notes:

- Maximal sample size estimates may be updated to maintain power
- Works great as long as we know (or have a good estimate of) $\prod_j !$

UCIrvine
University of California, Irvine

Group sequential testing in survival trials

UCIrvine
University of California, Irvine

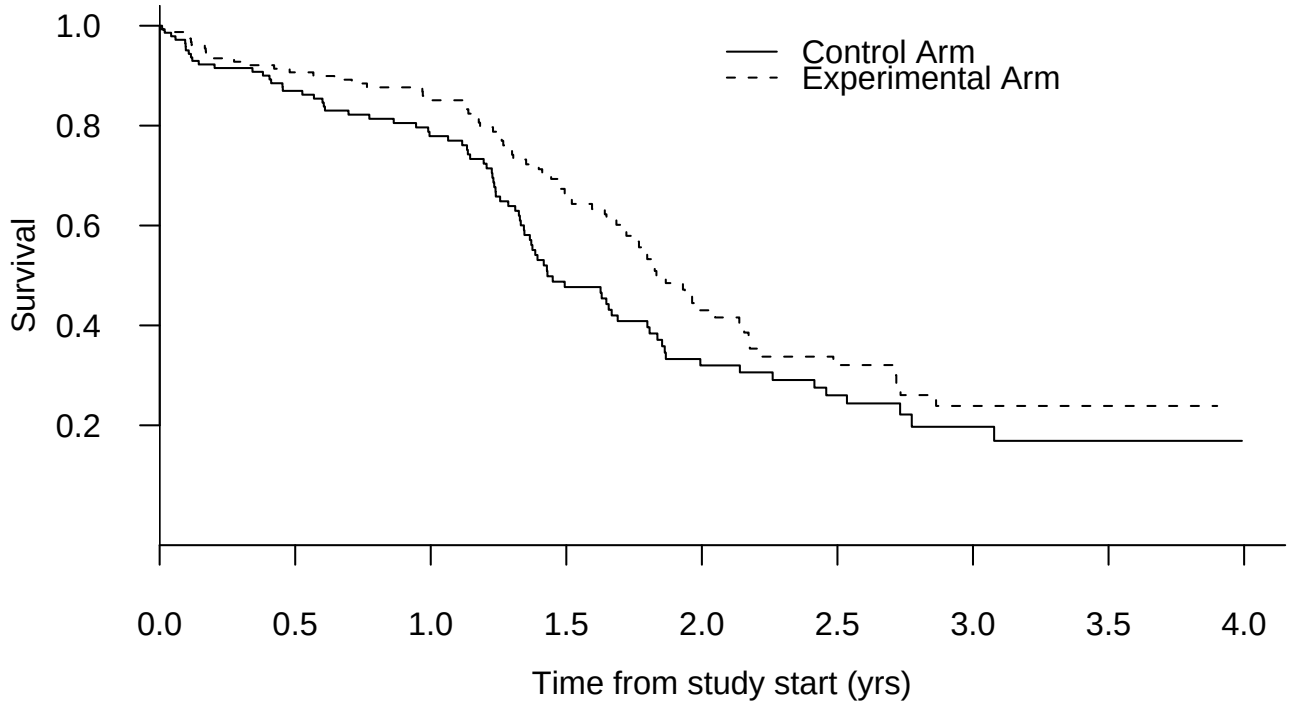
Further considerations when considering survival endpoints

- 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

Motivating Example: Phase II prostate cancer data

UCIrvine
University of California, Irvine

Simulated from actual trial data



Control	142 (0)	90 (29)	25 (73)	8 (80)	1 (81)
Treatment	158 (0)	99 (21)	31 (61)	9 (72)	1 (72)
Total	300 (0)	189 (50)	56 (134)	17 (152)	2 (153)

Weighted LR statistics

$G^{\rho,\gamma}$ statistic (Fleming & Harrington, 1991)

- 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

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

with

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

UCIrvine
University of California, Irvine

Weighted LR statistics

UCIrvine
University of California, Irvine

$G^{\rho,\gamma}$ 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

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

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

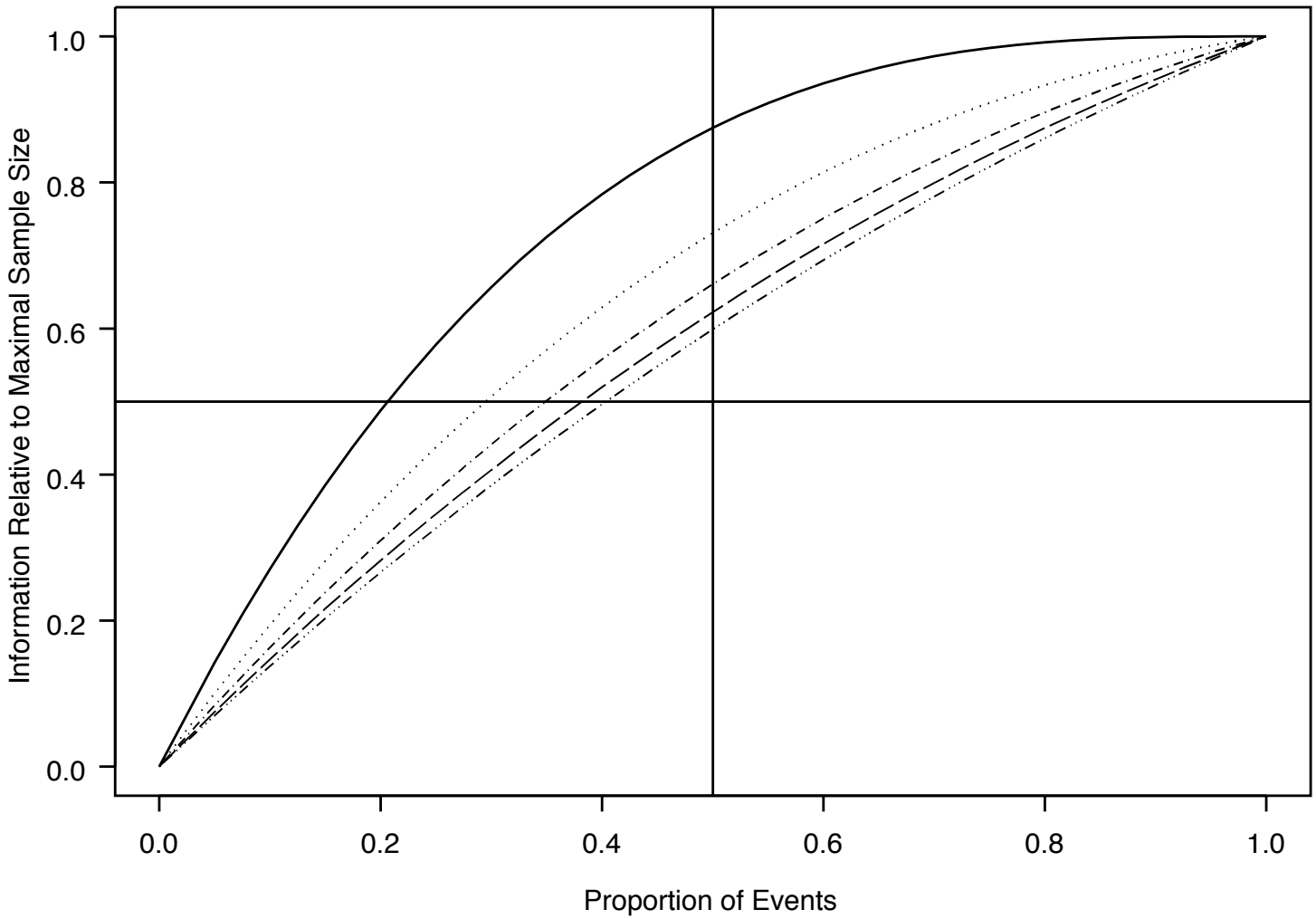
$$\sigma^2 \propto \int_0^\tau w^2(t) F_E(\tau - t)[1 - F_C(t)] dS(t)$$

- 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

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

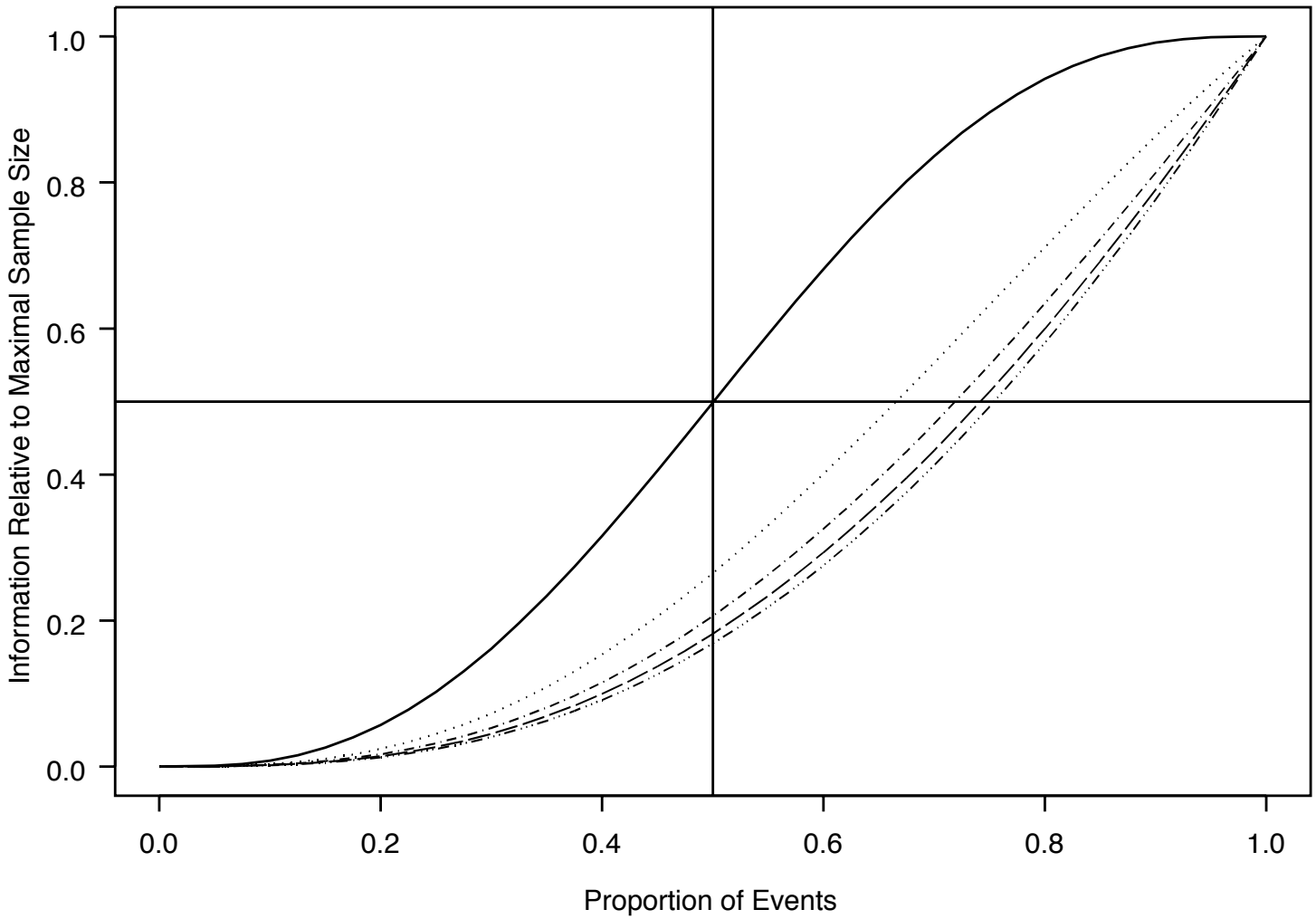
Example: Difference in Information by Accrual for the $G^{1,0}$ Statistic

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



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

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



UCIrvine
University of California, Irvine

Example: Operating characteristics with misspecified accrual distribution

UCIrvine
University of California, Irvine

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
- 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

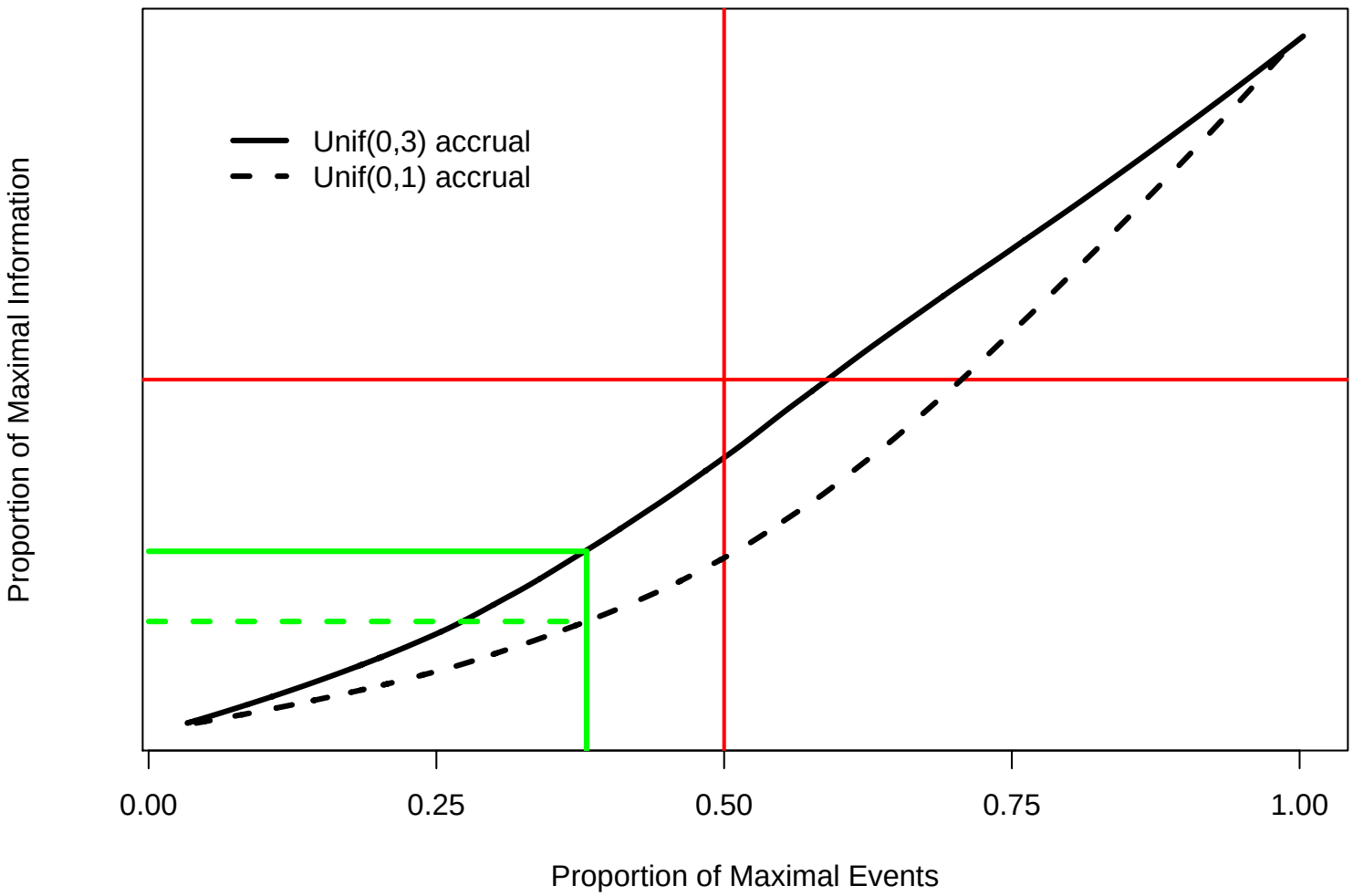
UCIrvine
University of California, Irvine

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
```

Example: Operating characteristics with misspecified accrual distribution



Example: Operating characteristics with misspecified accrual distribution

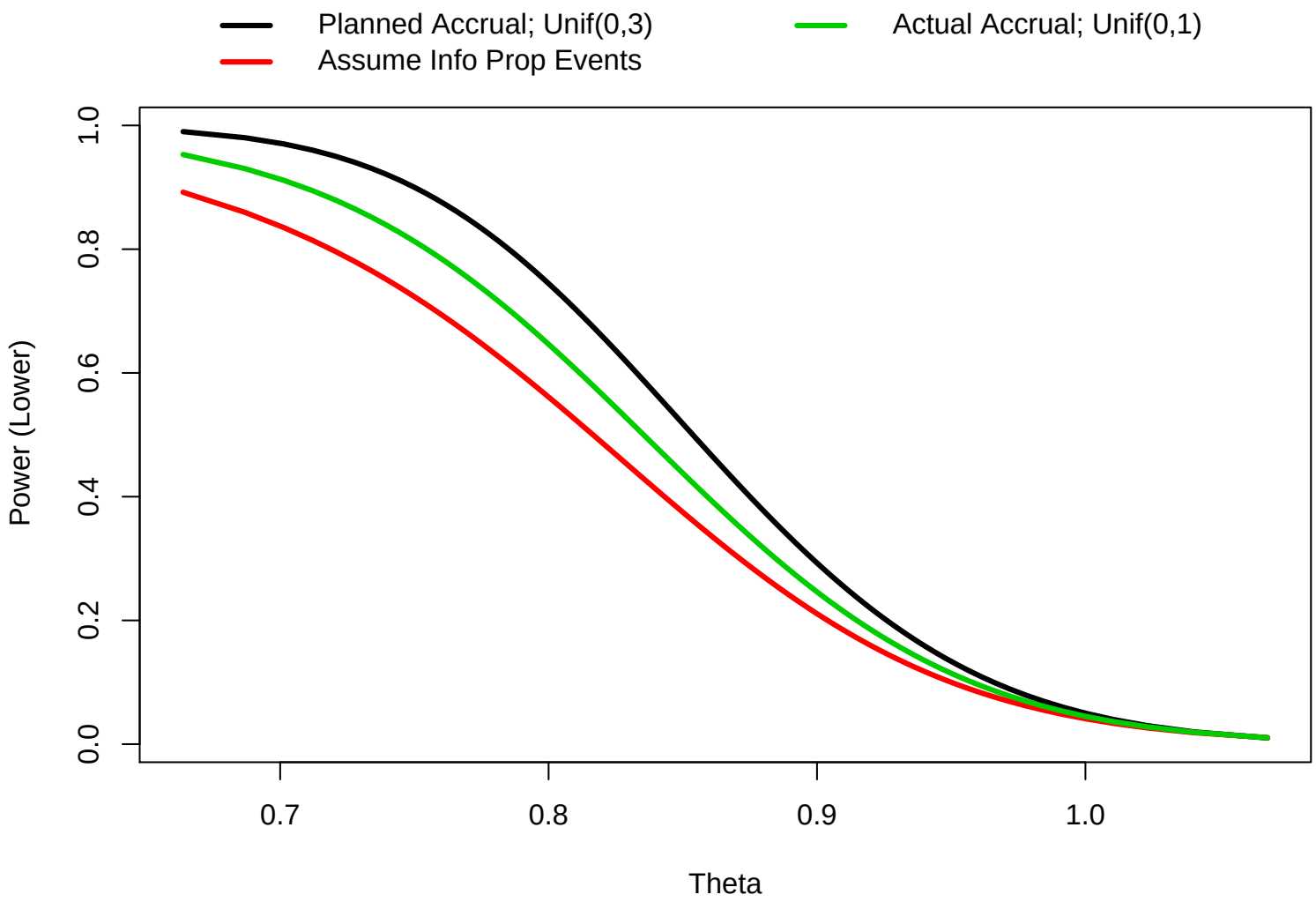
UCIrvine
University of California, Irvine

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



Implementation of group sequential rules

UCIrvine
University of California, Irvine

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
   - seems natural to turn to a Bayesian framework if prior information is available for $S(t)$ and/or $F_E(t)$!

Implementation of group sequential rules

UCIrvine
University of California, Irvine

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 specify prior distributions for $S(t)$ and $F_E(t)$

   1.3 map information increments to proportion of events for desired timing of first analysis

Implementation of group sequential rules

Proposed algorithm: Step 2

2. At each analysis,

   2.1 update estimate of $S(t)$ and $F_E(t)$ with data observed at the time of the analysis

   - Use pooled data so that constraint does not depend on observed treatment effect
   - If no prior data available, could update $S(t)$ and $F_E(t)$ via maximum likelihood

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

UCIrvine
University of California, Irvine

Simulation study

Simulation framework

- Generation of alternatives
  - Simulated from “mixings” of the survival curves from the Phase II pilot data
    1. 0% mixing: obtain S_1 by sampling only from observed survival for control in pilot data
    2. 50% mixing: obtain S_1 by sampling equally from observed survival for control and treatment in pilot data
    3. 100% mixing: obtain S_1 by sampling only from observed survival for treatment in pilot data
- 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 design alternative (100% mixing) $\rightarrow$ 507 max events

UCIrvine
University of California, Irvine

UCIrvine
University of California, Irvine

Simulation framework

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

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

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

Simulation study

UCIrvine
University of California, Irvine

Simulation framework

- Comparison of monitoring plans

1. Naive: Assume information proportional to number of events

2. Constrained boundaries using only observed phase III data: Update survival distribution using parametric (Weibull) MLE from phase III data observed data up to the time of the interim analysis

3. Constrained boundaries incorporating prior phase II data: Update survival distribution from posterior with prior derived from the phase II pilot data

UCIrvine
University of California, Irvine

Bayesian estimate of pooled survival at interim analyses

- Likelihood: Assume survival follows a Weibull(μ, α) distribution

$$L(\vec{x}|\mu,\alpha) = \prod_{i=1}^{n} f(x_i|\mu,\alpha)^{\delta_i} S(x_i|\mu,\alpha)^{1-\delta_i}$$

where x_i is the observed at risk time for subject i, δ_i is an event indicator, and

$$f(t|\mu,\alpha) = \mu\alpha t^{\alpha-1} e^{-\mu t^{\alpha}}$$
$$S(t|\mu,\alpha) = e^{-\mu t^{\alpha}}$$

Simulation study

Bayesian estimate of pooled survival at interim analyses

- Prior distribution for μ and α: Independent Gamma priors for μ and α

$$\mu \sim \text{Gamma}(\rho_1, \rho_2)$$
$$\alpha \sim \text{Gamma}(\kappa_1, \kappa_2)$$

*Note: (ρ_1, ρ_2) and (κ_1, κ_2) chosen to match mean and variance of μ and α from pilot data

- Posterior distribution for μ and α: Conditional densities of $\mu|\alpha, \vec{x}$ and $\alpha|\mu, \vec{x}$ known up to a constant

$\Rightarrow$ Gibbs sampling (in combination with adaptive rejection sampling) used for sampling from the posterior distribution of $\mu, \alpha|\vec{x}$

Simulation Study: Power with accrual over 1 year

Monitoring Strategy	r = 0.5	0.75	1.0	3.0	5.0
Null Hypothesis					
Info $\propto$ Events	0.0392	0.0396	0.0391	0.0381	0.0377
Const Bound (MLE)	0.0503	0.0503	0.0502	0.0501	0.0501
Const Bound (Bayes)	0.0501	0.0501	0.0501	0.0500	0.0500
Intermediate Hypothesis					
Info $\propto$ Events	0.2772	0.2807	0.2744	0.2623	0.2589
Const Bound (MLE)	0.4081	0.4087	0.4064	0.4053	0.4055
Const Bound (Bayes)	0.4081	0.4088	0.4065	0.4055	0.4055
Design Alternative					
Info $\propto$ Events	0.6967	0.7042	0.6917	0.6661	0.6582
Const Bound (MLE)	0.9049	0.9055	0.9034	0.9020	0.9021
Const Bound (Bayes)	0.9050	0.9056	0.9040	0.9020	0.9022

UCIrvine
University of California, Irvine

Simulation Study: ASN with accrual over 1 year

Monitoring Strategy	r 0.5	0.75	1.0	3.0	5.0
Null Hypothesis					
Const Bound (MLE)	280.1	275.1	276.3	279.8	279.6
Const Bound (Bayes)	264.4	260.2	261.7	262.8	262.2
Intermediate Hypothesis					
Const Bound (MLE)	357.0	353.6	355.9	354.9	354.1
Const Bound (Bayes)	344.9	342.0	344.4	341.9	340.9
Design Alternative					
Const Bound (MLE)	341.2	338.6	341.6	337.6	336.4
Const Bound (Bayes)	327.9	325.9	328.6	324.2	323.0

UCIrvine
University of California, Irvine

Summary

UCIrvine
University of California, Irvine

Concluding remarks

- From regulatory perspective, frequentist operating characteristics of clinical trial designs must be considered
- When monitoring trial results in a group sequential framework, operating characteristics depend on information growth patterns that can be non trivial
- Late phase clinical trials are an area where we often have prior information on the distribution of the endpoint of interest (particularly in the control population)
  - Many unwilling to accept the use of prior distributions when quantifying measures of treatment effect

Summary

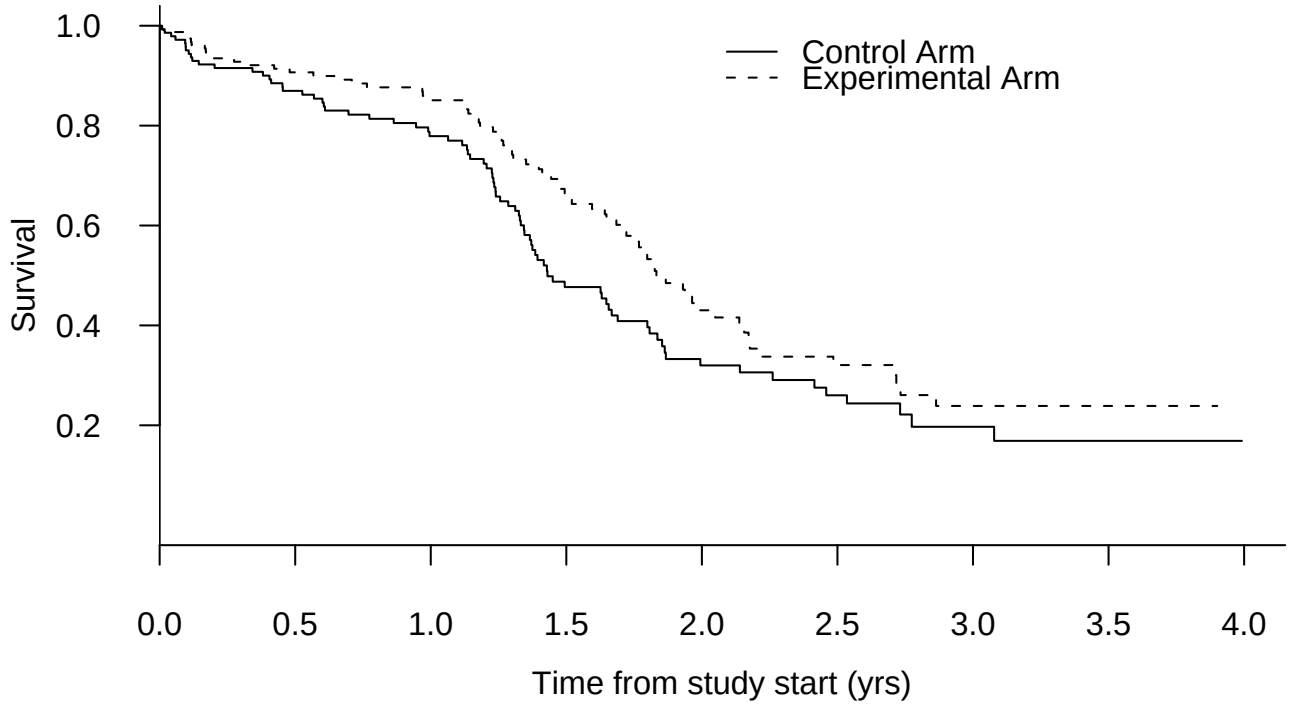
Concluding remarks

- However, in some cases the Bayesian framework may be used to improve the implementation of frequentist designs
  - Binomial proportions
  - Weighted survival statistics
  - Longitudinal measures (eg. comparison of slopes over time)
- Software for monitoring survival is available:
  - Constrained boundaries for unified family : SPlus SeqTrial
  - Mapping of statistical information : Me
- Ongoing work:
  - Misspecification of parametric form of entry and survival distributions?
  - Non-stationarity in entry and survival distributions?
  - Monitoring of longitudinal endpoints...

UCIrvine
University of California, Irvine

Motivating Example: Phase II prostate cancer data

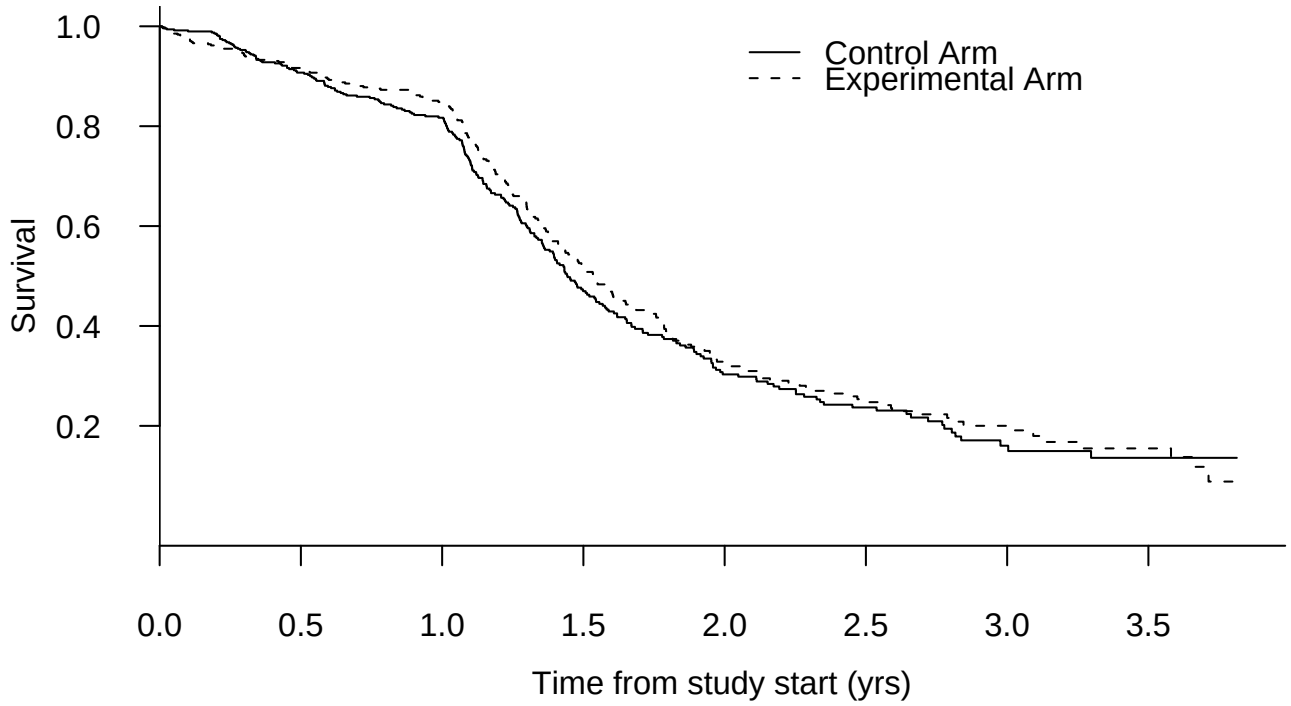
Simulated from actual trial data



Control	142 (0)	90 (29)	25 (73)	8 (80)	1 (81)
Treatment	158 (0)	99 (21)	31 (61)	9 (72)	1 (72)
Total	300 (0)	189 (50)	56 (134)	17 (152)	2 (153)

Alternatives used for simulation study

One simulated dataset under 0% mixing (Null hypothesis)

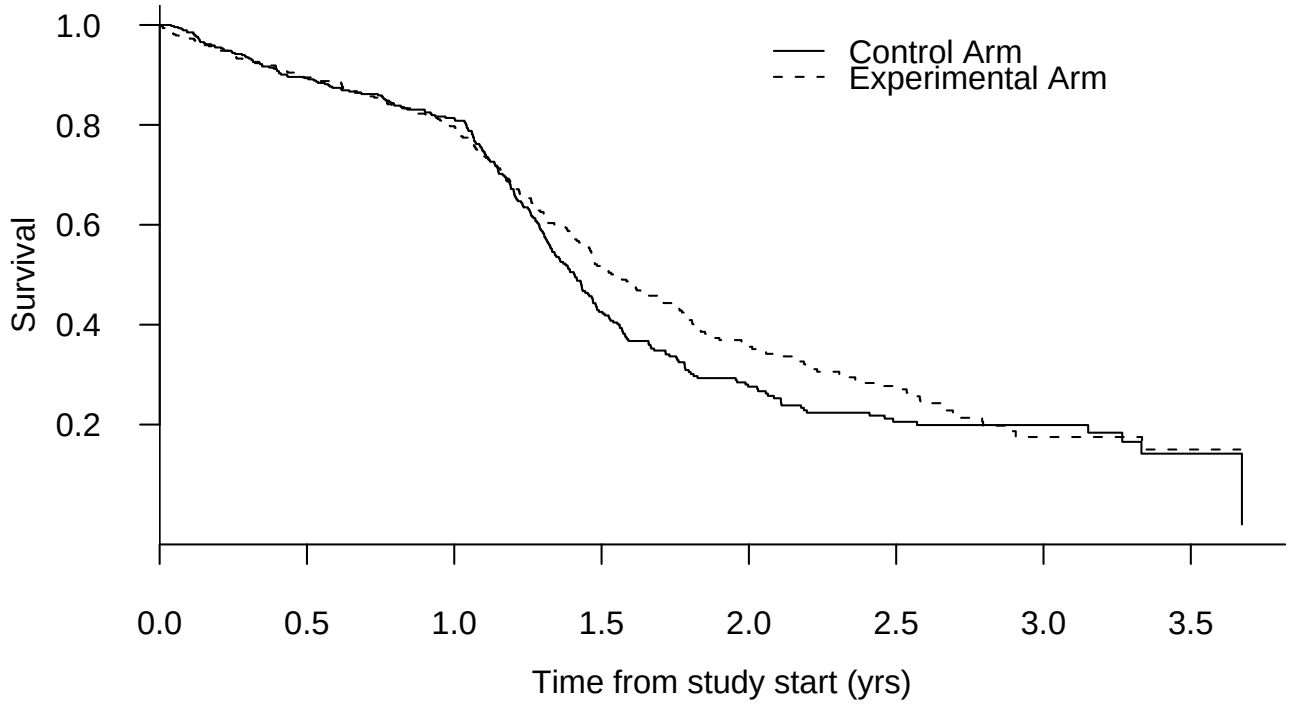


Control	490 (0)	296 (78)	68 (230)	16 (253)	1 (255)
Treatment	482 (0)	304 (66)	75 (222)	23 (245)	1 (252)
Total	972 (0)	600 (144)	143 (452)	39 (498)	2 (507)

UCIrvine
University of California, Irvine

Alternatives used for simulation study

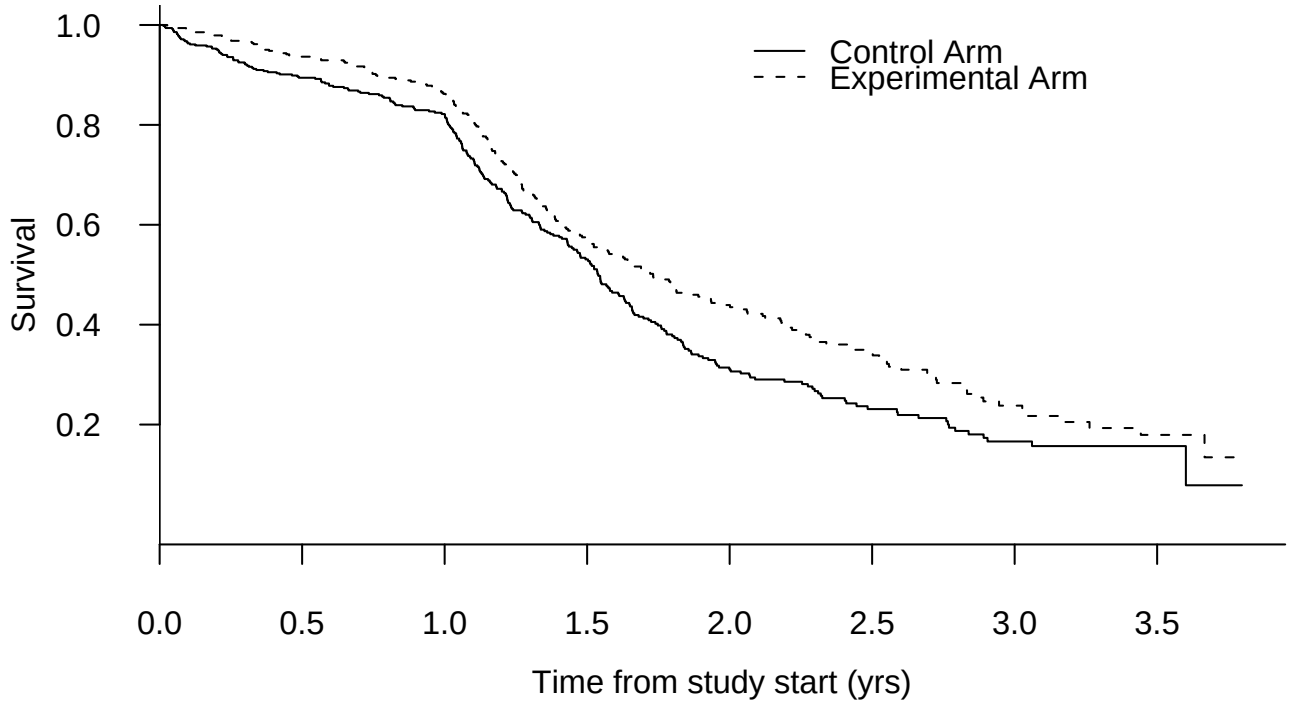
One simulated dataset under 50% mixing (Intermediate hypothesis)



Control	480 (0)	288 (79)	64 (241)	20 (256)	1 (260)
Treatment	481 (0)	282 (84)	80 (218)	13 (246)	1 (247)
Total	961 (0)	570 (163)	144 (459)	33 (502)	2 (507)

Alternatives used for simulation study

One simulated dataset under 100% mixing (Powered alternative)



Control	497 (0)	320 (80)	80 (246)	20 (273)	1 (275)
Treatment	492 (0)	310 (58)	103 (190)	27 (226)	1 (232)
Total	989 (0)	630 (138)	183 (436)	47 (499)	2 (507)

UCIrvine
University of California, Irvine