Efficient Two-Stage Adaptive Designs with Sample Size Adjustment

Qing Liu[1], Gang Li[1], Keaven M. Anderson[2] and Pilar Lim[1]
Contact: qliu2@its.jnj.com

[1]Janssen Pharmaceutical Companies of Johnson & Johnson

[2]Merck Research Laboratories

NJ ASA Chapter Annual Symposium
June 3, 2011

Outline

Introduction - Sample Size Problem

Consider a clinical trial to study the efficacy and safety of new drug where patients are randomized to receive either a treatment with the new drug or a control with a placebo or existing treatment

A key design element is to determine the required sample size (e.g, number of patients or events)

- Basic to all clinical trials

Introduction - Sample Size Problem

Consider a clinical trial to study the efficacy and safety of new drug where patients are randomized to receive either a treatment with the new drug or a control with a placebo or existing treatment

A key design element is to determine the required sample size (e.g, number of patients or events)

- Basic to all clinical trials
- Pre-specified type 1 error rate and desired power

Introduction - Sample Size Problem

Consider a clinical trial to study the efficacy and safety of new drug where patients are randomized to receive either a treatment with the new drug or a control with a placebo or existing treatment

A key design element is to determine the required sample size (e.g, number of patients or events)

- Basic to all clinical trials
- Pre-specified type 1 error rate and desired power
- Uncertainty in effect size and variability

Introduction - Sample Size Problem

Consider a clinical trial to study the efficacy and safety of new drug where patients are randomized to receive either a treatment with the new drug or a control with a placebo or existing treatment

A key design element is to determine the required sample size (e.g, number of patients or events)

- Basic to all clinical trials
- Pre-specified type 1 error rate and desired power
- Uncertainty in effect size and variability
- Fixed sample size design

Introduction - Sample Size Problem

Consider a clinical trial to study the efficacy and safety of new drug where patients are randomized to receive either a treatment with the new drug or a control with a placebo or existing treatment

A key design element is to determine the required sample size (e.g, number of patients or events)

- Basic to all clinical trials
- Pre-specified type 1 error rate and desired power
- Uncertainty in effect size and variability
- Fixed sample size design
- Group sequential design

Introduction - Sample Size Problem

Consider a clinical trial to study the efficacy and safety of new drug where patients are randomized to receive either a treatment with the new drug or a control with a placebo or existing treatment

A key design element is to determine the required sample size (e.g, number of patients or events)

- Basic to all clinical trials
- Pre-specified type 1 error rate and desired power
- Uncertainty in effect size and variability
- Fixed sample size design
- Group sequential design
- Adaptive design

Introduction - Motivating Example

Phase 3 comfirmative trial for patient with stroke to demonstrate a new treatment more effective than an existing treatment

- Success defined as achieving individual scores of zero or one by mRS

Introduction - Motivating Example

Phase 3 comfirmative trial for patient with stroke to demonstrate a new treatment more effective than an existing treatment

- Success defined as achieving individual scores of zero or one by mRS
- Pilot trial with success rate $p_T = 56\%$

Introduction - Motivating Example

Phase 3 comfirmative trial for patient with stroke to demonstrate a new treatment more effective than an existing treatment

- Success defined as achieving individual scores of zero or one by mRS
- Pilot trial with success rate $p_T = 56\%$
- Historical success rate of existing treatment $p_C = 34\%$

Introduction - Motivating Example

Phase 3 comfirmative trial for patient with stroke to demonstrate a new treatment more effective than an existing treatment

- Success defined as achieving individual scores of zero or one by mRS
- Pilot trial with success rate $p_T = 56\%$
- Historical success rate of existing treatment $p_C = 34\%$
- FDA requires a minimum effect size (i.e., treatment difference) at least 10%

Introduction - Motivating Example

Phase 3 comfirmative trial for patient with stroke to demonstrate a new treatment more effective than an existing treatment

- Success defined as achieving individual scores of zero or one by mRS
- Pilot trial with success rate $p_T = 56\%$
- Historical success rate of existing treatment $p_C = 34\%$
- FDA requires a minimum effect size (i.e., treatment difference) at least 10%
- Type 1 error rate $\alpha = .025$ and power $1 - \beta = .8$

Introduction - Motivating Example

Phase 3 comfirmative trial for patient with stroke to demonstrate a new treatment more effective than an existing treatment

- Success defined as achieving individual scores of zero or one by mRS
- Pilot trial with success rate $p_T = 56\%$
- Historical success rate of existing treatment $p_C = 34\%$
- FDA requires a minimum effect size (i.e., treatment difference) at least 10%
- Type 1 error rate $\alpha = .025$ and power $1 - \beta = .8$
- $2 : 1$ randomization ratio to treatment or control

Introduction - Motivating Example (Continued)

Table: *Sample size for fixed sample design*

p_C	p_T	$p_T - p_C$	Δ	n_C	n_T	n
.34	.54	.20	.2313	74	148	222
.34	.49	.15	.1745	129	258	387
.34	.44	.10	.1175	285	570	855

1. n_C, n_T are sample sizes for the control and treatment groups
2. $n = n_C + n_T$ is the total sample size
3. Δ is the canonical parameter defined in the appendix

Introduction - Alternative Designs

Extended Group Sequential Design

- Significance boundary to reject the null hypothesis with possibility to continue to quantify secondary efficacy or safety

Two-Stage Adaptive Design (TSD)

Introduction - Alternative Designs

Extended Group Sequential Design

- Significance boundary to reject the null hypothesis with possibility to continue to quantify secondary efficacy or safety
- Non-binding futility boundary to stop for lack of efficacy

Two-Stage Adaptive Design (TSD)

Extended Group Sequential Design

- Significance boundary to reject the null hypothesis with possibility to continue to quantify secondary efficacy or safety
- Non-binding futility boundary to stop for lack of efficacy
- Sequential p-values, confidence intervals, and point estimates (Liu and Anderson, 2008, JASA).

Two-Stage Adaptive Design (TSD)

Introduction - Alternative Designs

Extended Group Sequential Design

- Significance boundary to reject the null hypothesis with possibility to continue to quantify secondary efficacy or safety
- Non-binding futility boundary to stop for lack of efficacy
- Sequential p-values, confidence intervals, and point estimates (Liu and Anderson, 2008, JASA).
- Difficult to apply due to excessive over-running

Two-Stage Adaptive Design (TSD)

Introduction - Alternative Designs

Extended Group Sequential Design

- Significance boundary to reject the null hypothesis with possibility to continue to quantify secondary efficacy or safety
- Non-binding futility boundary to stop for lack of efficacy
- Sequential p-values, confidence intervals, and point estimates (Liu and Anderson, 2008, JASA).
- Difficult to apply due to excessive over-running

Two-Stage Adaptive Design (TSD)

- Single interim analysis with significance and futility boundary

Introduction - Alternative Designs

Extended Group Sequential Design

- Significance boundary to reject the null hypothesis with possibility to continue to quantify secondary efficacy or safety
- Non-binding futility boundary to stop for lack of efficacy
- Sequential *p*-values, confidence intervals, and point estimates (Liu and Anderson, 2008, JASA).
- Difficult to apply due to excessive over-running

Two-Stage Adaptive Design (TSD)

- Single interim analysis with significance and futility boundary
- Option to increase the sample size

Introduction - Alternative Designs

Extended Group Sequential Design

- Significance boundary to reject the null hypothesis with possibility to continue to quantify secondary efficacy or safety
- Non-binding futility boundary to stop for lack of efficacy
- Sequential *p*-values, confidence intervals, and point estimates (Liu and Anderson, 2008, JASA).
- Difficult to apply due to excessive over-running

Two-Stage Adaptive Design (TSD)

- Single interim analysis with significance and futility boundary
- Option to increase the sample size
- Combination tests or "likelihood ratio" test

Introduction - Alternative Designs

Extended Group Sequential Design

- Significance boundary to reject the null hypothesis with possibility to continue to quantify secondary efficacy or safety
- Non-binding futility boundary to stop for lack of efficacy
- Sequential *p*-values, confidence intervals, and point estimates (Liu and Anderson, 2008, JASA).
- Difficult to apply due to excessive over-running

Two-Stage Adaptive Design (TSD)

- Single interim analysis with significance and futility boundary
- Option to increase the sample size
- Combination tests or "likelihood ratio" test
- Controversial due to efficiency issues

Introduction - Debates and Insights

Inefficiency of Adaptive Designs

- Adaptive designs have various practical appeals

Insights

Introduction - Debates and Insights

Inefficiency of Adaptive Designs

- Adaptive designs have various practical appeals
- Combination tests (e.g., Cui, Hung and Wang, 1999; Bauer and Khöne, 1993) lack sufficiency

Insights

Introduction - Debates and Insights

Inefficiency of Adaptive Designs

- Adaptive designs have various practical appeals
- Combination tests (e.g., Cui, Hung and Wang, 1999; Bauer and Khöne, 1993) lack sufficiency
- Tsiatis and Mehta (2003) claims group sequential designs are uniformly more powerful

Insights

Introduction - Debates and Insights

Inefficiency of Adaptive Designs

- Adaptive designs have various practical appeals
- Combination tests (e.g., Cui, Hung and Wang, 1999; Bauer and Khöne, 1993) lack sufficiency
- Tsiatis and Mehta (2003) claims group sequential designs are uniformly more powerful
- The new FDA guidance on adaptive designs (2010) submits to these criticisms.

Insights

Introduction - Debates and Insights

Inefficiency of Adaptive Designs

- Adaptive designs have various practical appeals
- Combination tests (e.g., Cui, Hung and Wang, 1999; Bauer and Khöne, 1993) lack sufficiency
- Tsiatis and Mehta (2003) claims group sequential designs are uniformly more powerful
- The new FDA guidance on adaptive designs (2010) submits to these criticisms.

Insights

- Combination tests are analytically non-optimal but contribute little to inefficiency

Introduction - Debates and Insights

Inefficiency of Adaptive Designs

- Adaptive designs have various practical appeals
- Combination tests (e.g., Cui, Hung and Wang, 1999; Bauer and Khöne, 1993) lack sufficiency
- Tsiatis and Mehta (2003) claims group sequential designs are uniformly more powerful
- The new FDA guidance on adaptive designs (2010) submits to these criticisms.

Insights

- Combination tests are analytically non-optimal but contribute little to inefficiency
- Culprit is the sample size procedure using distorted *conditional power at the current trend*

New Likelihood Approach - Overview

Three Pillars

- Likelihood ratio test (Li, *et al.*, 2002)

Design

New Likelihood Approach - Overview

Three Pillars

- Likelihood ratio test (Li, *et al.*, 2002)
- Minimum effect size (Chi and Liu, 1999; Liu and Chi, 2001)

Design

New Likelihood Approach - Overview

Three Pillars

- Likelihood ratio test (Li, *et al.*, 2002)
- Minimum effect size (Chi and Liu, 1999; Liu and Chi, 2001)
- Pseudo group sequential design based sample size procedure

Design

New Likelihood Approach - Overview

Three Pillars

- Likelihood ratio test (Li, *et al.*, 2002)
- Minimum effect size (Chi and Liu, 1999; Liu and Chi, 2001)
- Pseudo group sequential design based sample size procedure

Design

- Two-Stage Design (TSD) with pre-specified first stage sample size n_1

New Likelihood Approach - Overview

Three Pillars

- Likelihood ratio test (Li, *et al.*, 2002)
- Minimum effect size (Chi and Liu, 1999; Liu and Chi, 2001)
- Pseudo group sequential design based sample size procedure

Design

- Two-Stage Design (TSD) with pre-specified first stage sample size n_1
- Interim analysis has non-binding futility and significance boundaries $a_1 < b_1$

New Likelihood Approach - Overview

Three Pillars

- Likelihood ratio test (Li, *et al.*, 2002)
- Minimum effect size (Chi and Liu, 1999; Liu and Chi, 2001)
- Pseudo group sequential design based sample size procedure

Design

- Two-Stage Design (TSD) with pre-specified first stage sample size n_1
- Interim analysis has non-binding futility and significance boundaries $a_1 < b_1$
- Second stage sample size is $\tilde{n}_2(Z_1)$ where Z_1 is the first stage test statistic

New Likelihood Approach - Test Procedure

- H_0: $\Delta \leq 0$ versus H_A: $\Delta > 0$

New Likelihood Approach - Test Procedure

- H_0: $\Delta \leq 0$ versus H_A: $\Delta > 0$
- Stage-wise test statistics Z_1 and X_2 such that

$$E(Z_1) = n_1^{1/2}\Delta \text{ and } Var(Z_1) = 1$$

and

$$E(X_2) = \{\tilde{n}_2(Z_1) - n_1\}^{1/2}\Delta \text{ and } Var(X_2) = 1$$

New Likelihood Approach - Test Procedure

- H_0: $\Delta \leq 0$ versus H_A: $\Delta > 0$
- Stage-wise test statistics Z_1 and X_2 such that

$$E(Z_1) = n_1^{1/2}\Delta \text{ and } Var(Z_1) = 1$$

  and

$$E(X_2) = \{\tilde{n}_2(Z_1) - n_1\}^{1/2}\Delta \text{ and } Var(X_2) = 1$$

- Let $\lambda(Z_1) = n_1/\tilde{n}_2(Z_1)$. The final test statistic is

$$Z_2 = \{\lambda(Z_1)\}^{1/2} Z_1 + \{1 - \lambda(Z_1)\}^{1/2} X_2$$

New Likelihood Approach - Test Procedure

- H_0: $\Delta \leq 0$ versus H_A: $\Delta > 0$
- Stage-wise test statistics Z_1 and X_2 such that

$$E(Z_1) = n_1^{1/2}\Delta \;\text{ and }\; Var(Z_1) = 1$$

  and

$$E(X_2) = \{\tilde{n}_2(Z_1) - n_1\}^{1/2}\Delta \;\text{ and }\; Var(X_2) = 1$$

- Let $\lambda(Z_1) = n_1/\tilde{n}_2(Z_1)$. The final test statistic is

$$Z_2 = \{\lambda(Z_1)\}^{1/2} Z_1 + \{1 - \lambda(Z_1)\}^{1/2} X_2$$

- For type 1 error rate α, choose upper bound $b_2^{\dagger}$ such that

$$P_{\Delta=0}\{Z_1 \geq b_1 \cup Z_2 \geq b_2^{\dagger}\} = \alpha$$

New Likelihood Approach - Sample Size Procedure

Sample Size Specification

- Minimum effect size δ_{min}

Pseudo Group Sequential Design (pGSD)

New Likelihood Approach - Sample Size Procedure

Sample Size Specification

- Minimum effect size δ_{min}
- For type 2 error rate β, choose sample size $\tilde{n}_2(Z_1)$ such that

$$P_{\Delta=\delta_{\text{min}}}\{Z_1 \geq b_1 \cup Z_2 \geq b_2^{\dagger}\} = 1 - \beta$$

Pseudo Group Sequential Design (pGSD)

New Likelihood Approach - Sample Size Procedure

Sample Size Specification

- Minimum effect size δ_{min}
- For type 2 error rate β, choose sample size $\tilde{n}_2(Z_1)$ such that

$$P_{\Delta=\delta_{\text{min}}}\{Z_1 \geq b_1 \cup Z_2 \geq b_2^{\dagger}\} = 1 - \beta$$

Pseudo Group Sequential Design (pGSD)

- First interim of pGSD matches the interim of the two-stage design

New Likelihood Approach - Sample Size Procedure

Sample Size Specification

- Minimum effect size δ_{min}
- For type 2 error rate β, choose sample size $\tilde{n}_2(Z_1)$ such that

$$P_{\Delta=\delta_{\text{min}}}\{Z_1 \geq b_1 \cup Z_2 \geq b_2^{\dagger}\} = 1 - \beta$$

Pseudo Group Sequential Design (pGSD)

- First interim of pGSD matches the interim of the two-stage design
- Overall power of $1 - \beta$ at the minimum effect size δ_{min}

New Likelihood Approach - Sample Size Procedure

Sample Size Specification

- Minimum effect size δ_{min}
- For type 2 error rate β, choose sample size $\tilde{n}_2(Z_1)$ such that

$$P_{\Delta=\delta_{\text{min}}}\{Z_1 \geq b_1 \cup Z_2 \geq b_2^\dagger\} = 1 - \beta$$

Pseudo Group Sequential Design (pGSD)

- First interim of pGSD matches the interim of the two-stage design
- Overall power of $1 - \beta$ at the minimum effect size δ_{min}
- Sample sizes for subsequent interim analyses provide the final sample size for the two-stage design

New Likelihood Approach - Sequential Inference

Sequential *p*-values

- Existing *p*-values for adaptive designs, except Liu and Pledger (2006), are not *p*-values and lead to logical and interpretability problems

Sequential Estimates

New Likelihood Approach - Sequential Inference

Sequential *p*-values

- Existing *p*-values for adaptive designs, except Liu and Pledger (2006), are not *p*-values and lead to logical and interpretability problems
- Sequential *p*-values of extended group sequential designs by Liu and Anderson (2008)

Sequential Estimates

New Likelihood Approach - Sequential Inference

Sequential *p*-values

- Existing *p*-values for adaptive designs, except Liu and Pledger (2006), are not *p*-values and lead to logical and interpretability problems
- Sequential *p*-values of extended group sequential designs by Liu and Anderson (2008)
- Equivalent to two-stage likelihood ratio test

Sequential Estimates

New Likelihood Approach - Sequential Inference

Sequential *p*-values

- Existing *p*-values for adaptive designs, except Liu and Pledger (2006), are not *p*-values and lead to logical and interpretability problems
- Sequential *p*-values of extended group sequential designs by Liu and Anderson (2008)
- Equivalent to two-stage likelihood ratio test
- Applicable to existing multiplicity procedures developed for single stage designs

Sequential Estimates

New Likelihood Approach - Sequential Inference

Sequential *p*-values

- Existing *p*-values for adaptive designs, except Liu and Pledger (2006), are not *p*-values and lead to logical and interpretability problems
- Sequential *p*-values of extended group sequential designs by Liu and Anderson (2008)
- Equivalent to two-stage likelihood ratio test
- Applicable to existing multiplicity procedures developed for single stage designs

Sequential Estimates

- Sequential confidence intervals of extended group sequential designs by Liu and Anderson (2008)

New Likelihood Approach - Sequential Inference

Sequential *p*-values

- Existing *p*-values for adaptive designs, except Liu and Pledger (2006), are not *p*-values and lead to logical and interpretability problems
- Sequential *p*-values of extended group sequential designs by Liu and Anderson (2008)
- Equivalent to two-stage likelihood ratio test
- Applicable to existing multiplicity procedures developed for single stage designs

Sequential Estimates

- Sequential confidence intervals of extended group sequential designs by Liu and Anderson (2008)
- Equivalent to two-stage likelihood ratio test and sequential *p*-values

New Likelihood Approach - Sequential Inference

Sequential *p*-values

- Existing *p*-values for adaptive designs, except Liu and Pledger (2006), are not *p*-values and lead to logical and interpretability problems
- Sequential *p*-values of extended group sequential designs by Liu and Anderson (2008)
- Equivalent to two-stage likelihood ratio test
- Applicable to existing multiplicity procedures developed for single stage designs

Sequential Estimates

- Sequential confidence intervals of extended group sequential designs by Liu and Anderson (2008)
- Equivalent to two-stage likelihood ratio test and sequential *p*-values
- Median unbiased estimates

Motivating Example - Design

- Over-running sample size is 100
- $\alpha = .025$ and $\beta = .2$
- $n_1 = 300$ and min $\tilde{n}_2(Z_1) = 450$
- $a_1 = .7672$ and $b_1 = 3.017$
- $b_1^\dagger = 1.973$

Table: *Information adaptation rule*

k	2	3	4	5	6	7	8
z_{1k}	2.4194	2.2316	2.0734	1.9263	1.7847	1.6475	1.5162
$\tilde{n}_2$	450	475	500	525	550	575	600
k	9	10	11	12	13	14	15
z_{1k}	1.3947	1.2879	1.1998	1.1319	1.0838	.7672	$-\infty$
$\tilde{n}_2$	625	650	675	700	725	750	400

k — the analysis number of pGSD

Motivating Example - Operating Properties

Table: *Operating properties*

pGSD				
Δ	0	.1175	.1745	.2313
Fut. Prob. IA	.7785	.1024	.0121	.0006
Sig. Prob. IA	.0013	.1631	.5021	.8387
Sig. Prob. Final	.0194	.7964	.9822	.9993
Exp. Info.	443.8	555.1	482.5	424.7

TSD				
Δ	0	.1175	.1745	.2313
Fut. Prob. IA	.7785	.1024	.0121	.0006
Sig. Prob. IA	.0013	.1631	.5021	.8387
Sig. Prob. Final	.0244	.7964	.9806	.9993
Exp. Info.	457.9	513.4	451.6	412.4

1. Fut. Prob. IA — probability to stop for futility at IA
2. Sig. Prob. IA — probability to reject H IA
3. Sig. Prob. Final — cumulative probability to reject H at FA
4. Exp. Info. — expected information

Simulation - Operating Properties

Table: *Operating properties*

Δ	0	.1175	.1745	.2313
Fut. Prob. IA	.7804	.1013	.0107	.0004
Sig. Prob. IA	.0014	.1647	.5025	.8368
Sig. Prob. Final	.0244	.7930	.9808	.9995
Prob. Coverage	.9516	.9515	.9507	.9507
Exp. Info.	457.3	513.5	452.1	412.5

Prob. Coverage — probability of coverage

Simulation - Estimation

Table: *Point estimators*

Method	MedUE	MVUE	MedUE	MVUE
Δ	0		.1175	
Mean	-.00231	-.00058	.11853	.11705
SD	.047489	.06646	.04776	.05657
Bias	-.00231	-.00058	.00103	-.00048
$MSE^{1/2}$	.047545	.06646	.04777	.05657
% Bias	n/a	n/a	.00877	-.00411
% MAE	n/a	n/a	.31770	.37237
Δ	.1745		.2313	
Mean	.17578	.17387	.23156	.23065
SD	.04760	.06377	.04884	.06965
Bias	.00128	-.00063	.00026	-.00065
$MSE^{1/2}$	.04762	.06377	.04884	.06965
% Bias	.00733	-.00361	.00112	-.00281
% MAE	.21861	.28834	.16987	.23992

1. MedUE - median unbiased estimate
2. MAE - mean absolute error defined as mean of absolute bias divided by the effect size

Optimization - First Generation

Table: *Generation I ($R = 1.3158$)*

	pGSD			
Δ	0	.1175	.1745	.2313
Sig. Prob. Final	.0194	.7964	.9822	.9993
Exp. Info.	443.8	555.1	482.5	424.7
	TSD			
Δ	0	.1175	.1745	.2313
Sig. Prob. Final	.0244	.7964	.9806	.9993
Exp. Info.	457.9	513.4	451.6	412.4

$R = n_{2K}/n_T$ for which n_{2K} is the maximum information and $n_T = 570$ is the naive information taken from Table 1

Optimization - Second Generation

Table: *Generation II ($R = 1.2474$)*

	pGSD			
Δ	0	.1175	.1745	.2313
Sig. Prob. Final	.0213	.7964	.9826	.9994
Exp. Info.	444.1	539.7	471.7	421.0
	TSD			
Δ	0	.1175	.1745	.2313
Sig. Prob. Final	.0245	.7964	.9805	.9993
Exp. Info.	456.5	510.8	447.7	410.4

Optimization - Last Generation

Table: *Generation III* ($R = 1.1930$)

	pGSD			
Δ	0	.1175	.1745	.2313
Sig. Prob. Final	.0215	.7964	.9826	.9994
Exp. Info.	443.1	532.6	465.9	418.5
	TSD			
Δ	0	.1175	.1745	.2313
Sig. Prob. Final	.0245	.7964	.9809	.9993
Exp. Info.	453.8	510.1	447.1	409.7

Discussion

- Desirable operating characteristics
- Sequential inference
- Flexible and easy to implement
- More efficient then group sequential designs
- Optimization procedure can be applied to optimize other adaptive designs
- Not shown that the likelihood ratio test is an optimal analysis
- Avoid conditional power at current trend
- Use Cui, Hung and Wang (1999) when sample size adjustment is unplanned