Operating characteristic guided design of group-sequential trials

Aniko Szabo & Sergey Tarima

Medical College of Wisconsin

GMDS & CEN-IBS 2020 Satellite Meeting

Table of Contents

COVID-19 vaccine trials

- Goal: evaluate the vaccine efficacy (VE) defined as $1 - HR$ or $1 - RR$
- Multiple companies running randomized trials
- Basic design features are similar

Characteristic	Value
H_0	VE $= 30\%$
H_a	VE $= 60\%$
α	one-sided 2.5%
Power	90%

COVID-19 vaccine trials

Characteristic	Moderna[1]	Pfizer[2]	AstraZeneca[3]
Target #events	151	164	150
Interim analyses	2	4	1
Type	Efficacy	Efficacy+Futility	Efficacy
Timing	35%, 70%	$\sim$ 19%, 38%, 56%, 73%	50%
α-spending	Lan-DeMets OF	Bayesian PP	Lan-DeMets OF

[1] ModernaTx. *A Phase 3, Randomized, Stratified, Observer-Blind,Placebo-Controlled Study to Evaluate the Efficacy, Safety, andImmunogenicity of mRNA-1273 SARS-CoV-2 Vaccine inAdults Aged 18 Years and Older*. Aug. 2020. URL: https://www.modernatx.com/sites/default/files/mRNA-1273-P301-Protocol.pdf.

[2] Pfizer. *A Phase 1/2/3, Placebo-Controlled, Randomized, Observer-Blind, Dose-Finding Study To Evaluate The Safety, Tolerability, Immunogenicity, And Efficacy Of SARS-CoV-2 RNA Vaccine Candidates Against COVID-19 In Healthy Individuals*. 2020. URL: https://pfe-pfizercom-d8-prod.s3.amazonaws.com/2020-09/C4591001_Clinical_Protocol.pdf.

[3] AstraZeneca. *A Phase III Randomized, Double-blind, Placebo-controlledMulticenter Study in Adults to Determine the Safety, Efficacy, andImmunogenicity of AZD1222, a Non-replicating ChAdOx1 VectorVaccine, for the Prevention of COVID-19*. Sept. 2020. URL: https://s3.amazonaws.com/ctr-med-7111/D8110C00001/52bec400-80f6-4c1b-8791-0483923d0867/c8070a4e-6a9d-46f9-8c32-cece903592b9/D8110C00001_CSP-v2.pdf.

COVID-19 vaccine trials

- Same hypotheses, significance level, power
- Large variability in interim analysis plans
- Timing of analyses and choice of spending function not really justified
- Operating characteristics, defined as the probability of stopping at an interim analysis are given

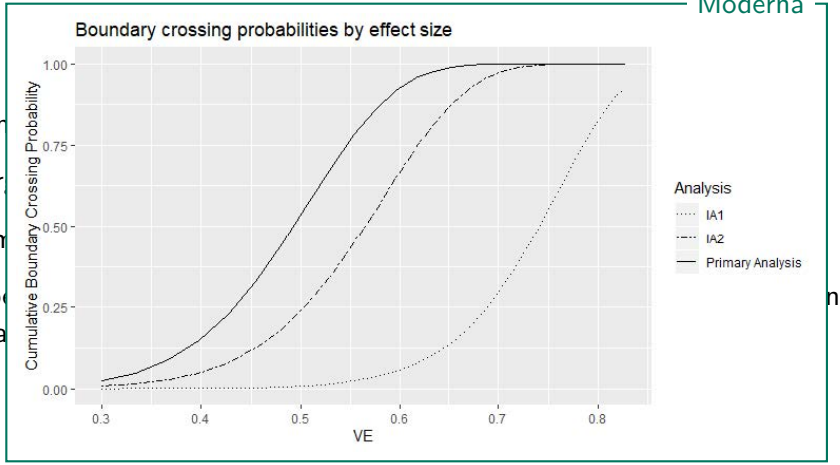
COVID-19 vaccine trials

- Sar
- Lar
- Tim
- Op

ana

nterim

Moderna



COVID-19 vaccine trials

- Sar
- Lar
- Tim
- Op

ana

nterim

Pfizer

Table 7. Statistical Design Operating Characteristics: Probability of Success or Failure for Interim Analyses

Vaccine Efficacy (%)	Interim Analysis 1 (Total Cases = 32)		Interim Analysis 2 (Total Cases = 62)		Interim Analysis 3 (Total Cases = 92)		Interim Analysis 4 (Total Cases = 120)
	Probability of Success (Cases in Vaccine Group ≤6)	Probability of Failure (Cases in Vaccine Group ≥15)	Probability of Success (Cases in Vaccine Group ≤15)	Probability of Failure (Cases in Vaccine Group ≥26)	Probability of Success (Cases in Vaccine Group ≤25)	Probability of Failure (Cases in Vaccine Group ≥35)	Probability of Success (Cases Vaccine Group ≤35)
30	0.006	0.315	0.003	0.231	0.002	0.239	0.002
50	0.054	0.078	0.051	0.056	0.063	0.103	0.075
60	0.150	0.021	0.160	0.010	0.175	0.019	0.160
70	0.368	0.003	0.310	<0.001	0.195	0.001	0.085
80	0.722	<0.001	0.238	<0.001	0.037	<0.001	0.003

COVID-19 vaccine trials

- Same hypotheses, significance level, power
- Large
- Timi
- Oper

AstraZeneca

An interim efficacy analysis . . . will give >70% power to detect a VE of 70% and >90% power to detect a VE of 75%

d

interim analysis are given

Operating characteristic guided design

Aniko Szabo

Sergey Tarima

Main idea

- Select the interim analysis times and α-spending to achieve the desired operating characteristics
- Optimize a useful criterion (eg sample size) if multiple designs are available

Build on existing ideas

- Group-sequential designs and calculations are well established[4]
- Optimal designs with minimal expected sample size[5,6]
- Efficacy boundary with target power for a sequence of alternatives
  - Fixing the timing of the tests[7]
  - Fixing the α-spending sequence[8]

We will add:

- Optimization while maintaining target power
- Futility boundary based on sequence of 'nulls'

---

[4] C. Jennison and B.W. Turnbull. *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC Interdisciplinary Statistics. New York: CRC Press, 1999.

[5] John D. Eales. "Optimal two-sided group sequential tests". In: *Sequential Analysis* 14.4 (1995), pp. 273–286.

[6] Christopher Jennison and Bruce W. Turnbull. "Efficient group sequential designs when there are several effect sizes under consideration.". In: *Statistics in Medicine* 25 (6 Mar. 2006), pp. 917–932.

[7] Helmut Schäfer and Hans-Helge Müller. "Construction of group sequential designs in clinical trials on the basis of detectable treatment differences.". In: *Statistics in Medicine* 23 (9 May 2004), pp. 1413–1424.

[8] Sergey Tarima and Nancy Flournoy. "Effect of Interim Adaptations in Group Sequential Designs". In: *arXiv preprints* (Aug. 2019). arXiv: 1908.01411 [stat.ME].

Setup and notations

Most study design settings can be reduced to/appoximated by the following special case:

$$X_i \sim N(\theta, \sigma^2) \text{ iid with known } \sigma$$

Hypothesis:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_A : \theta \geq \theta_0$$

Constraints:

- Type I error α
- Power π at $\theta = \theta_A$

Test statistic

Based on n observed values:

$$Z(n) = (\bar{X}_n - \theta_0)\sqrt{n}/\sigma = \sqrt{\mathcal{I}_n}(\bar{X}_n - \theta_0)$$

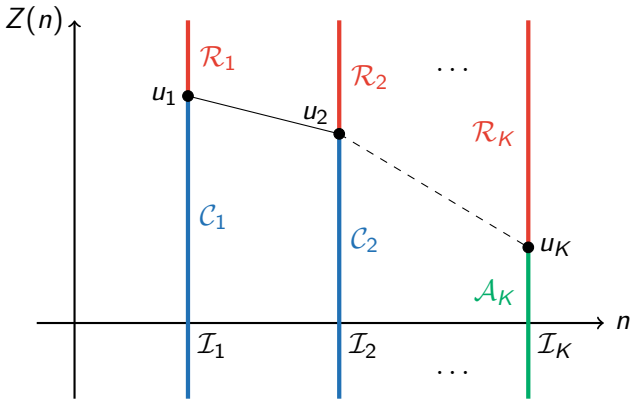
where $\mathcal{I}_n = [\sigma^2/n]^{-1}$ is the information*

$$Z(n) \sim N(\sqrt{\mathcal{I}_n}(\theta - \theta_0), 1)$$

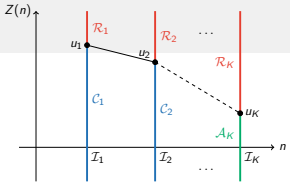
$$Cov(Z_{n_1}, Z_{n_2}) = \sqrt{\mathcal{I}_{n_1}/\mathcal{I}_{n_2}} \text{ for } n_1 \leq n_2$$

* "Information" $\mathcal{I}_n$ will be used as a rescaling of n that unifies many settings

Group-sequential design



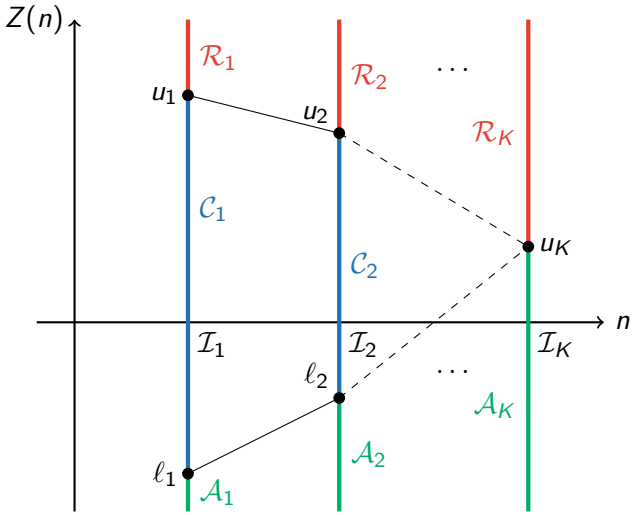
Target operating characteristics



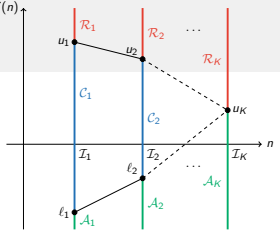
For $\theta_{A1} \geq \theta_{A2} \geq \cdots \geq \theta_{AK} = \theta_A$:

Characteristic	Event	θ	Target
Overall type I error	$\bigcup_{i=1}^{K} \left\{ \mathcal{C}_1 \cap \cdots \cap \mathcal{C}_{i-1} \cap \mathcal{R}_i \right\}$	θ_0	$\leq \alpha$
Overall power	—— " ——	θ_A	$\geq \pi$
Stop by stage k for efficacy	$\bigcup_{i=1}^{k} \left\{ \mathcal{C}_1 \cap \cdots \cap \mathcal{C}_{i-1} \cap \mathcal{R}_i \right\}$	θ_{Ak}	$\geq \pi_E$

Group-sequential design with futility boundary



Target operating characteristics with futility

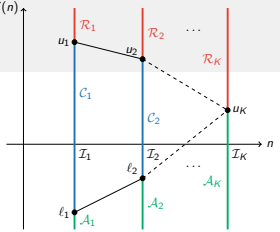


Binding lower boundary

For $\theta_{A1} \geq \theta_{A2} \geq \cdots \geq \theta_{AK} = \theta_A$
and $\theta_{01} \leq \theta_{02} \leq \cdots \leq \theta_{0K} = \theta_0$:

Characteristic	Event	θ	Target
Overall type I error	$\bigcup_{i=1}^{K} \left\{\mathcal{C}_1 \cap \cdots \cap \mathcal{C}_{i-1} \cap \mathcal{R}_i\right\}$	θ_0	$\leq \alpha$
Overall power	—— " ——	θ_A	$\geq \pi$
Stop by stage k for efficacy	$\bigcup_{i=1}^{k} \left\{\mathcal{C}_1 \cap \cdots \cap \mathcal{C}_{i-1} \cap \mathcal{R}_i\right\}$	θ_{Ak}	$\geq \pi_E$
Stop by stage k for futility	$\bigcup_{i=1}^{K} \left\{\mathcal{C}_1 \cap \cdots \cap \mathcal{C}_{i-1} \cap \mathcal{A}_i\right\}$	θ_{0k}	$\geq \pi_F$

Target operating characteristics with futility



Non-binding lower boundary

For $\theta_{A1} \geq \theta_{A2} \geq \cdots \geq \theta_{AK} = \theta_A$
and $\theta_{01} \leq \theta_{02} \leq \cdots \leq \theta_{0K} = \theta_0$:

Characteristic	Event	θ	Target
Overall type I error	$\bigcup_{i=1}^{K} \left\{ \overline{\mathcal{R}}_1 \cap \cdots \cap \overline{\mathcal{R}}_{i-1} \cap \mathcal{R}_i \right\}$	θ_0	$\leq \alpha$
Overall power	—— " ——	θ_A	$\geq \pi$
Stop by stage k for efficacy	$\bigcup_{i=1}^{k} \left\{ \overline{\mathcal{R}}_1 \cap \cdots \cap \overline{\mathcal{R}}_{i-1} \cap \mathcal{R}_i \right\}$	θ_{Ak}	$\geq \pi_E$
Stop by stage k for futility	$\bigcup_{i=1}^{K} \left\{ \mathcal{C}_1 \cap \cdots \cap \mathcal{C}_{i-1} \cap \mathcal{A}_i \right\}$	θ_{0k}	$\geq \pi_F$

Existence theorem

Aniko Szabo

Sergey Tarima

Theorem

A design satisfying the operating characteristic requirements exists, if the following conditions are satisfied:

C1. $\pi_E \leq \pi$

C2. $\theta_{A1} \geq \theta_{A2} \geq \cdots \geq \theta_{A,K-1} \geq \theta_A$

Additionally, if a futility boundary is needed:

C3. $\pi_F \leq 1 - \alpha$

C4. $\theta_{01} \leq \theta_{02} \leq \cdots \leq \theta_{0,K-1} \leq \theta_0$

Proof by induction on K

- For $K = 1$ – known one-stage sample size calculation

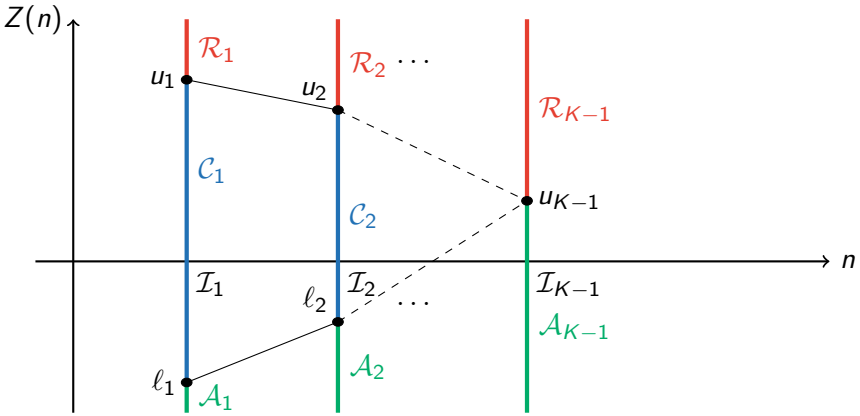
Proof by induction on K

- For $K = 1$ – known one-stage sample size calculation
- Construct a design for $K - 1$ stages with
  - Significance level $\alpha_{K-1} < \alpha$ at θ_0
  - Overall and stage-specific power π_E under $\theta_{A1}, \ldots, \theta_{A,K-1}$
  - Stage-specific futility stopping probabilities π_F under $\theta_{01}, \ldots, \theta_{0,K-2}$, if desired

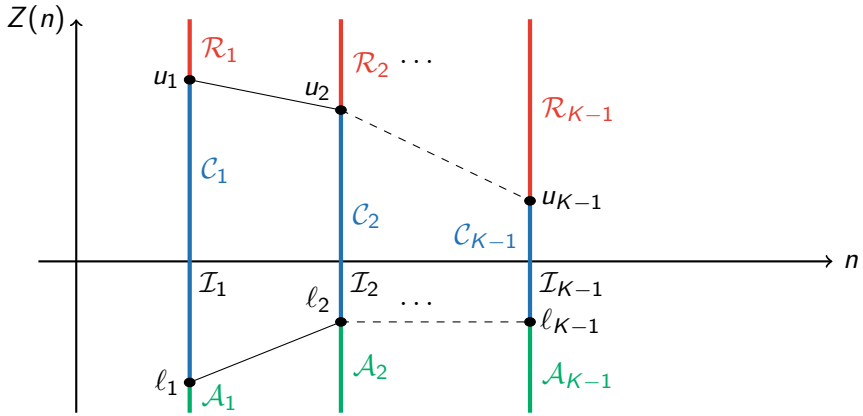
Proof by induction on K

- For $K = 1$ – known one-stage sample size calculation
- Construct a design for $K - 1$ stages with
  - Significance level $\alpha_{K-1} < \alpha$ at θ_0
  - Overall and stage-specific power π_E under $\theta_{A1}, \ldots, \theta_{A,K-1}$
  - Stage-specific futility stopping probabilities π_F under $\theta_{01}, \ldots, \theta_{0,K-2}$, if desired
- Add lower boundary ℓ_{K-1}

Building group-sequential boundary: ℓ_{K-1}

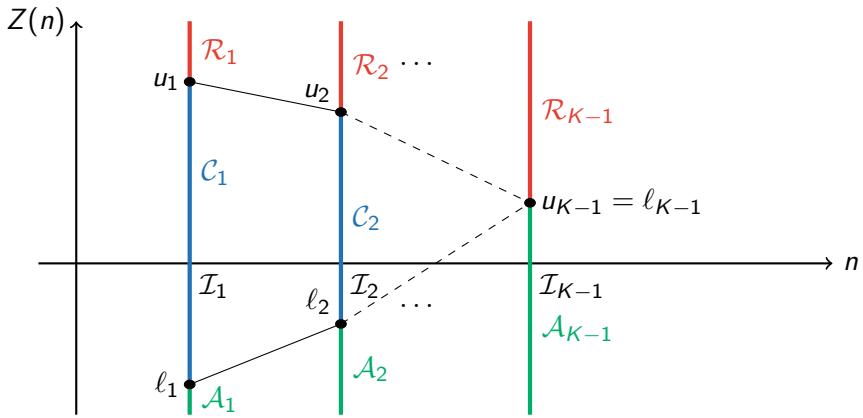


Building group-sequential boundary: ℓ_{K-1}



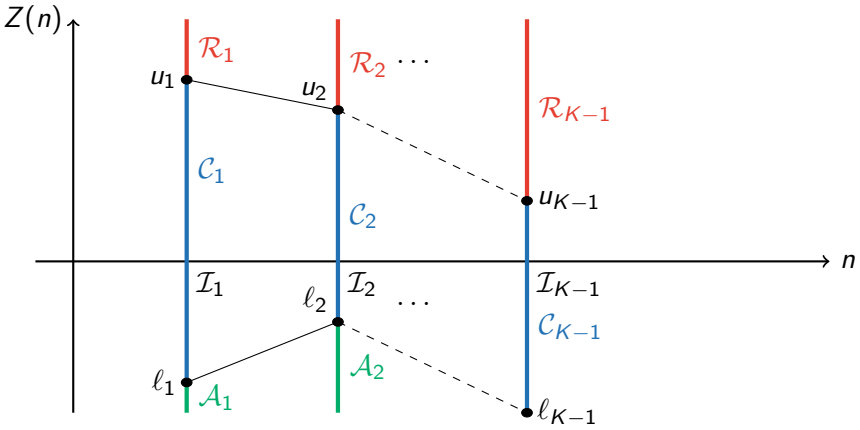
Need: $P(\text{accept by (K-1) with } \ell^*_{K-1} | \theta_{0,K-1}) = \pi_F$

Building group-sequential boundary: ℓ_{K-1}



$$P(\text{accept by } K-1 \text{ with } \ell_{K-1} = u_{K-1} | \theta_{0,K-1}) =$$
$$1 - P(\text{reject by } K-1 | \theta_{0,K-1}) \geq 1 - P(\text{reject by } K-1 | \theta_0) =$$
$$1 - \alpha_{K-1} \geq 1 - \alpha \geq \pi_F$$

Building group-sequential boundary: ℓ_{K-1}

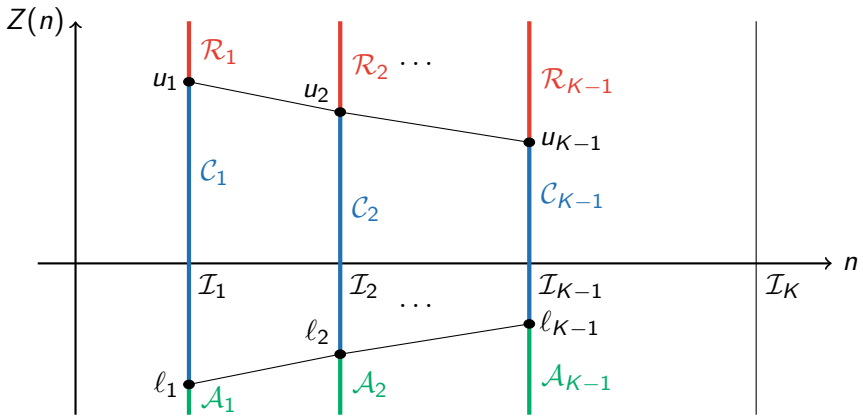


$$P(\text{accept by } K-1 \text{ with } \ell_{K-1} = -\infty | \theta_{0,K-1}) =$$
$$P(\text{accept by } K-2 \,|\theta_{0,K-1}) \le P(\text{accept by } K-2 \,|\theta_{0,K-2}) = \pi_F$$

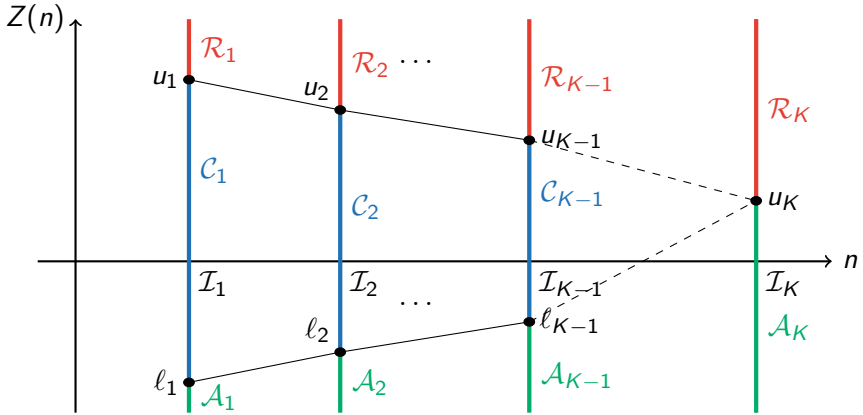
Proof by induction on K

- For $K = 1$ – known one-stage sample size calculation
- Construct a design for $K - 1$ stages with
  - Significance level $\alpha_{K-1} < \alpha$ at θ_0
  - Overall and stage-specific power π_E under $\theta_{A1}, \ldots, \theta_{A,K-1}$
  - Stage-specific futility stopping probabilities π_F under $\theta_{01}, \ldots, \theta_{0,K-2}$, if desired
- Add lower boundary ℓ_{K-1}
- Find parameters for stage K
  - Given $\mathcal{I}_K$, find u_K to guarantee type I error

Building group-sequential boundary: u_K

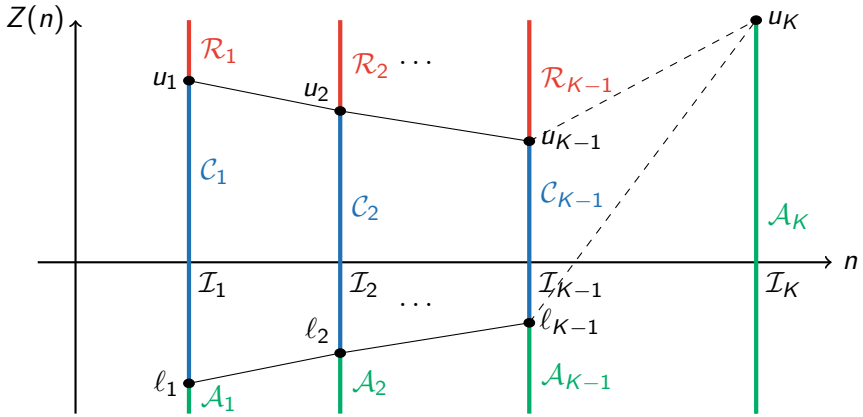


Building group-sequential boundary: u_K



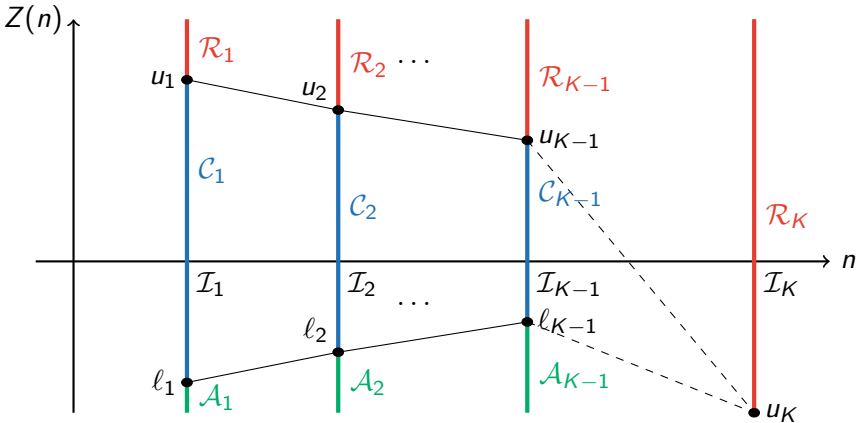
Need: $P(\text{reject by } K \text{ with } u_K^*|\theta_0) = \alpha$

Building group-sequential boundary: u_K



$$P(\text{reject by } K \text{ with } u_K = \infty | \theta_0) =$$
$$P(\text{reject by } K-1 | \theta_0) = \alpha_{K-1} \leq \alpha$$

Building group-sequential boundary: u_K



$$P(\text{reject by } K \text{ with } u_K = -\infty | \theta_0) = 1 - P(\text{accept by } K-1 | \theta_0) \geq$$
$$1 - P(\text{accept by } K-1 | \theta_{0,K-1}) = 1 - \pi_F \geq \alpha$$

Proof by induction on K

- For $K = 1$ – known one-stage sample size calculation
- Construct a design for $K - 1$ stages with
  - Significance level $\alpha_{K-1} < \alpha$ at θ_0
  - Overall and stage-specific power π_E under $\theta_{A1}, \ldots, \theta_{A,K-1}$
  - Stage-specific futility stopping probabilities π_F under $\theta_{01}, \ldots, \theta_{0,K-2}$, if desired
- Add lower boundary ℓ_{K-1}
- Find parameters for stage K
  - Given $\mathcal{I}_K$, find u_K to guarantee type I error
  - Find $\mathcal{I}_K$ to guarantee power
    - With a binding lower bound $\mathcal{I}_{K-1}$ might have to be increased

Existence theorem

We have actually proven a stronger result

Theorem

Under conditions C1 – C4, given an α-spending sequence $\alpha_1 \leq \cdots \leq \alpha_{K-1}$, a well-defined ($\approx$ unique) design satisfying the operating characteristic requirements exists such that additionally for $k = 1, \ldots, K-1$,

$$P\left(\bigcup_{i=1}^{k} \left\{\mathcal{C}_1 \cap \cdots \cap \mathcal{C}_{i-1} \cap \mathcal{R}_i\right\} \middle| \theta_0\right) \leq \alpha_k$$

Optimization

Aniko Szabo

Sergey Tarima

Following Jennison and Turnbull[9], we optimize the expected sample size averaged over values of interest for θ

Optimality criterion

$$ASN = \sum_i w_i E[N|\theta = \theta_i^*]$$

- Original proposal: $\theta_0, \theta_A, \theta_0 + L(\theta_A - \theta_0)$
- Our setting: $\theta_0, \theta_A, \theta_{A1}, \ldots, \theta_{A,K-1}$
  - With futility boundary, can include $\theta_{01}, \ldots, \theta_{0,K-1}$
- For a non-binding futility boundary, we do not include the crossing of the lower bound in the calculation of $E[N]$

[9] Jennison and Turnbull, "Efficient group sequential designs when there are several effect sizes under consideration."

Implementation

Optimization can be implemented

- by multivariate optimization over the error-spending sequence $\{\alpha_k\}$
- by backward recursive univariate optimization over the last term α_{K-1}

Since

$$\mathcal{I} \propto n \quad \text{and} \quad n \propto (\theta_A - \theta_0)^{-2},$$

effect sizes can be scaled to $\theta_A - \theta_0$: $\quad r_k^E = \dfrac{\theta_{Ak} - \theta_0}{\theta_A - \theta_0} > 1; \quad r_k^F = \dfrac{\theta_{0k} - \theta_0}{\theta_A - \theta_0} < 0$

to get

$$n(\theta_0, \theta_1) = \frac{\mathcal{I}(0,1)}{\mathcal{I}_{fix}} \frac{n_{fix}}{(\theta_A - \theta_0)^2} = \frac{\mathcal{I}(0,1)}{(z_\alpha + z_{1-\beta})^2} n_{fix},$$

where $\mathcal{I}(0,1)$ is computed assuming $\theta_0 = 0, \theta_A = 1, \theta_{Ak} = r_k^E, \theta_{0k} = r_k^F$ and $_{fix}$ refers to the one-stage study with the target significance level and power

AstraZeneca trial – one interim efficacy analysis

- $H_0 : VE = 30\%$ versus $H_A : VE = 60\%$
- $n_{fix} = 149$
- One interim analysis to achieve 90% power at interim analysis if $VE = 75\%$

```
library(gsDesign)
gsDesign(k=2, test.type = 1, alpha=0.025, beta=0.1,
         timing = c(0.5), sfu = sfLDOF, n.fix = 149)
```

```
library(gsDesignOC)
ve_to_r <- function(ve, ve0=0.3, ve1=0.6){
     (log(1-ve) - log(1-ve0)) / (log(1-ve1)-log(1-ve0))}
gsDesignOC(n.stages = 2, rE.seq = c(ve_to_r(0.75), 1), n.fix = 149,
           sig.level = 0.025, power=0.9, power.efficacy = 0.9,
           r_EN = c(0,1,ve_to_r(0.75))
```

AstraZeneca trial – one interim efficacy analysis

	gsDesign			gsDesignOC		
Analysis	N	u	Spend	N	u	Spend
Interim	75	2.96	0.0015	58	2.42	0.0078
Final	150	1.97	0.0235	157	2.06	0.0171

Cumulative power at each analysis

	gsDesign			gsDesignOC		
VE	Interim	Final	$E(N)$	Interim	Final	$E(N)$
30%	0.15%	2.5%	149.4	0.79%	2.5%	155.6
60%	25.2%	90%	130.6	34.2%	90%	122.5
75%	89.7%	>99%	82.5	90%	>99%	67.2
Average			120.8			115.1

Moderna trial – two interim analyses

- Two interim efficacy analyses to achieve 80% power at IA1 and IA2, if $VE = 80\%, 70\%$
- Add non-binding futility analyses to stop with 80% probability at IA1 and IA2, if $VE = 0\%, 15\%$

```
gsDesign(k=3, test.type = 4, alpha=0.025, beta=0.1,
         timing = c(0.35, 0.7), sfu = sfLDOF, n.fix = 148)

gsDesignOC(n.stages = 3, rE.seq = ve_to_r(c(0.8,0.7,0.6)),
           rF.seq = ve_to_r(c(0, 0.15, 0.3)),
           futility.type = "non-binding", n.fix = 148,
           power.efficacy = 0.8, power.futility = 0.8,
           sig.level = 0.025, power=0.9,
           r_EN = ve_to_r(c(0.8,0.7,0.6,0.3,0.15,0)))
```

Moderna trial – two interim analyses

	gsDesign			gsDesignOC		
Analysis	N	u	ℓ	N	u	ℓ
Interim 1	56	3.61	-0.16	38	2.79	-0.19
Interim 2	112	2.44	1.06	73	2.63	-0.18
Final	159	2.00	2.00	153	2.03	2.03

Cumulative crossing probability at each analysis (in the relevant direction)

	gsDesign				gsDesignOC			
VE	IA1	IA2	Final	$E(N)$	IA1	IA2	Final	$E(N)$
0%	86.5%	>99.9%	>99.9%	158.6	80%	92.8%	>99.9%	152.7
15%	70.1%	98.0%	>99.9%	158.6	64.5%	80%	99.9%	152.6
30%	43.6%	86.2%	97.5%	158.3	42.4%	54.7%	97.5%	152.1
60%	5.2%	64.3%	90%	125.1	12.2%	37.7%	90%	118.2
70%	27.2%	96.5%	99.9%	97.6	37.0%	80%	99.5%	75.5
80%	79.7%	>99.9%	>99.9%	66.8	80%	99.3%	>99.9%	44.7
Average				127.5				116.0

Potential extensions

- Target power at interim analysis could also change: $\pi_{E1} \leq \cdots \leq \pi_{EK} = \pi$
  - Interesting setting: $\theta_{0k} = \theta$, but $\pi_{F1} < \cdots < \pi_{FK} = 1 - \alpha$
- Optimization could be constrained by logistics
  - restrict maximal sample size inflation
  - fix final significance level
  - different optimization criteria
- Other operating characteristics could be targeted

Summary

- We should design trials to get desired operating characteristics
- For group-sequential trials crossing probabilities under selected alternatives are of interest
  - Can allow for uncertainty about effect size
  - Stopping for futility can be incorporated with an ordered sequence of 'nulls'
- Optimization can be used to select among feasible designs
  - Eliminates the need to specify timing and spending function
- R package `gsDesignOC` available on GitHub