

Adaptive Phase II Clinical Trials - Rigorous Statistics and Flexible Science -

Tatsuki Koyama, Ph.D.

Cancer Biostatistics Center, Vanderbilt-Ingram Cancer Center
Department of Biostatistics, Vanderbilt University School of Medicine

April 18, 2008

Outline

- 1 Quotes
- 2 Phases of Clinical Trials
- 3 Notes on Experimental Designs
- 4 Adaptive Phase II Clinical Trials
- 5 Summary

“To call in the statistician after an experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of.”

Ronald Fisher
At the Indian Statistical Congress, ca 1938

from Lindley and Phillips (1976)

Suppose you are convinced that the coin you have is loaded (not fair), and you decide to test whether $P[\text{Heads}] = 0.5$.

You observed 9 Heads in 12 tosses in the following sequence:

HHTHHHHTTHHH

- case 1** If the experimental design were “Toss the coin 12 times and count the number of Heads”, then the p -value would be 0.073.
- case 2** If the experimental design were “Toss the coin until you see 9 Heads”, then the p -value would be 0.033.
- case 3** If there were no plan. If you stopped after 12 tosses because your coffee was ready, then the p -value would be \dots ?

“Statistics is hard.”

“Statistics is hard.”

Tatsuki Koyama

“Most statistics programs are designed by statisticians, for statisticians. These programs are feature-packed and powerful, but can overwhelm scientists with thick manuals, obscure statistical jargon and high prices. $\$ \% \#$ is different. $\$ \% \#$ is designed by a scientist for scientists.”

“You don’t know how Lab Science works,

“You don’t know how Lab Science works, Tatsuki.”

Somebody I don’t talk to anymore.

Clinical Trial Phases

Phase I Dose-finding

- Establish a safe dose and schedule of administration
- Determine the types of side effects and toxicity
- Assess evidence for efficacy

Phase II Preliminary evidence of efficacy

- Safety should be closely monitored
- May be one-arm against historical control
- May be many dose levels

Phase III Pivotal

- May be against an already available treatment

Phase IV Post-marketing surveillance

Sample Size Computation

Suppose you want to test

$$H_0 : \pi = 0.3$$

$$H_1 : \pi = 0.5$$

PLAN 1

- A software tells you to use $N = 50$ and reject H_0 if there are 21 or more “Successes” in 50.
- Then $\alpha = 0.048$ and Power ≈ 0.90 .

Repeat Until Significance

Suppose you want to test

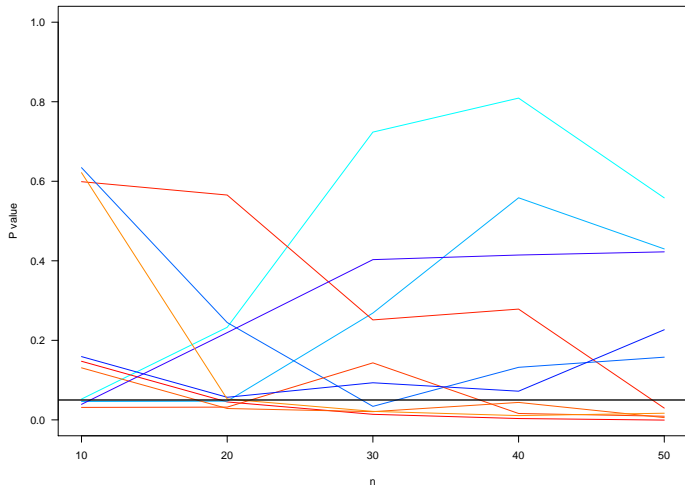
$$H_0 : \pi = 0.3$$

$$H_1 : \pi = 0.5$$

PLAN 2

- You have a sample of size 10 every week. Get the data and compute a P -value.
- If not significance ($P \geq .05$) \Rightarrow “not enough information” \Rightarrow get an additional sample of size 10.
- After $N = 50$, give up.

Red PLAN 1 / Blue PLAN 2



Type I error rate

The probabilities of incorrectly rejecting H_0 when H_0 is actually true (i.e., type I error rates) are

- PLAN 1 : 0.047
- PLAN 2 : 0.114

- FDA would not accept PLAN 2.
- “Sample until enough information” should be a good strategy.
i.e., “Stop when you have enough information.”

⇒ Bayesian?

Adaptive Designs

Definition by PhRMA working group

A clinical study design that uses accumulating data to decide how to modify aspects of the study as it continues, without undermining the validity and integrity of the trial.

Adaptive Designs

Definition by PhRMA working group

A clinical study design that uses accumulating data to decide how to modify aspects of the study as it continues, without undermining the validity and integrity of the trial.

Also ...

... changes are made “by design,” and not on an ad hoc basis; therefore, adaptation is a design feature aimed to enhance the trial, not remedy for inadequate planning.”

Types of Adaptive Designs

- Modify sample size
 - Early termination
- Change the primary endpoint
- Treatment selection
 - Drop ineffective arms
- Seamless Phase II/III trials

Pros and Cons of Adaptive Designs

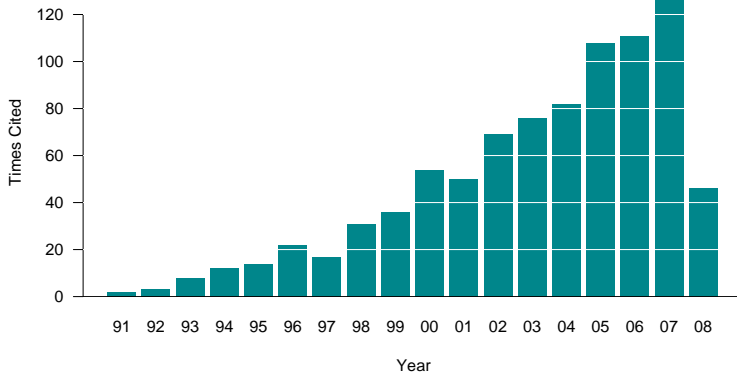
Pros

- Expedited decisions
“The ability to fail faster is an important advance in science.”
- Flexibility
- Reduced sample size (on average)
- Reduced up-front cost

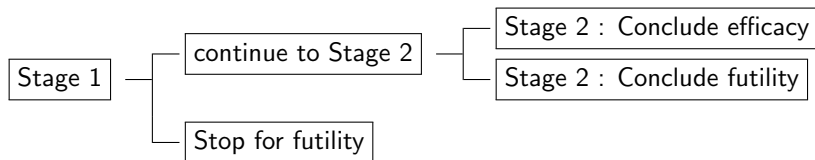
Cons

- Complexity
- Opportunities to cheat
(It's difficult to maintain the integrity.)
- Increased maximum sample size (worst case scenario)
- Difficult inference from the data
- Necessity to halt accrual

Simon's Designs



Flowchart



Example : Simon's Design

$$H_0 : \pi = 0.3$$

$$H_1 : \pi = 0.5$$

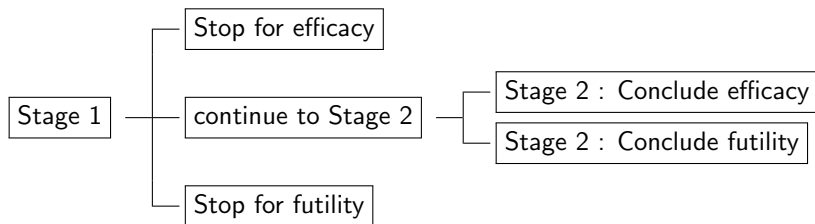
(Sample size for a single stage design is 53.)

Stage 1 Take a sample of size 24. If there are 7 or fewer positive responses, then stop for futility, otherwise continue to stage 2.

Stage 2 Take an additional sample of size 29 (total is 53). If there are 22 or more positive responses total, conclude efficacy.

True π	Stage 1 Probabilities		Stage 2 Probability	E[Sample Size]
	Futility	Continue	Efficacy	
0.3	0.565	0.435	0.047	36.6
0.5	0.032	0.968	0.902	52.1

Early Stopping for Efficacy Also



Example : with Early Stopping for Efficacy

$$H_0 : \pi = 0.3$$

$$H_1 : \pi = 0.5$$

- Stage 1** Take a sample of size 25. If there are 7 or fewer positive responses stop for futility, and if there are 13 or more positive responses stop for efficacy, otherwise continue to Stage 2.
- Stage 2** Take an additional sample of size 27 (total is 52). If there are 22 or more positive responses total, conclude efficacy.

True π	Stage 1 Probabilities			Stage 2 Probability	Power	E[Sample Size]
	Futility	Continue	Efficacy	Efficacy		
0.3	0.565	0.424	0.011	0.032	0.043	35.9
0.5	0.032	0.549	0.419	0.467	0.886	39.4

Comparisons

Single stage design

True π	Stage 1 Probabilities			Stage 2 Probability	Power	E[Sample Size]
	Futility	Continue	Efficacy	Efficacy		
0.3	—	—	—	—	0.049	53
0.5	—	—	—	—	0.916	53

Early Stopping for Futility

True π	Stage 1 Probabilities			Stage 2 Probability	Power	E[Sample Size]
	Futility	Continue	Efficacy	Efficacy		
0.3	0.565	0.435	—	0.047	0.047	36.6
0.5	0.032	0.968	—	0.902	0.902	52.1

Early Stopping for Futility and Efficacy

True π	Stage 1 Probabilities			Stage 2 Probability	Power	E[Sample Size]
	Futility	Continue	Efficacy	Efficacy		
0.3	0.565	0.424	0.011	0.032	0.043	35.9
0.5	0.032	0.549	0.419	0.467	0.886	39.4

Abused Simon's Designs

Accrual into this study was stopped as soon as the number of responses required by the statistical design was achieved.

Baldini et al. *Cancer chemotherapy and pharmacology* 2002

"... the study proceeded to the second step and it was completed when a total of seven responses clearly exceeded the six response target in the global pt series."

Caraglia et al. *Cancer chemotherapy and pharmacology* 2006

Inference from Simon's Designs

STATISTICS IN MEDICINE

Statist. Med. (2007)

Published online in Wiley InterScience

(www.interscience.wiley.com) DOI: 10.1002/sim.3123



Proper inference from Simon's two-stage designs

Tatsuki Koyama^{*,†} and Heidi Chen

*Department of Biostatistics, Vanderbilt University School of Medicine, 571 Preston Building, Nashville,
TN 37232-6848, U.S.A.*

Oh my!

A MinMax two-stage accrual design described by Simon will ensure that the total number of the patients exposed to this therapy is minimized. ... Initially, 66 eligible patients will be entered into the study. If 40 or more patients reach 100-day survival, the trial will continue until 68 patients have been treated. ... This design provides 80% statistical power to detect a difference of 15% (65% vs. 50% of the historical control) with a significance level less than 0.05 (type I error).

Something I reviewed.

Summary

- After all, statistics can't salvage a bad experiment.
- Rigorous statistics can be flexible; just plan ahead!

Adaptive Phase II Clinical Trials - Rigorous Statistics and Flexible Science -

Tatsuki Koyama, Ph.D.

Cancer Biostatistics Center, Vanderbilt-Ingram Cancer Center
Department of Biostatistics, Vanderbilt University School of Medicine

April 18, 2008