Interim Analysis and Adaptive Design

Rui Qin and Yunling Du

Biostatistics and Data Management
Regeneron Pharmaceuticals
rui.qin@regeneron.com

June 22, 2016

Table of Contents

REGENERON
Biostatistics and Data Management

Clinical Trial

- A prospectively planned experiment for the purpose of evaluating a potentially beneficial therapy or treatment
- Conducted under as many controlled conditions as possible so that they provide definitive answers to pre-determined, well-defined questions
- Classic design requires such parameters to be pre-specified and fixed throughout a clinical trial
  - Sample size
  - Randomization ratio
  - Number of study arms
  - ... ...

REGENERON
Biostatistics and Data Management

Adaptations in Clinical Trial

Sometimes are necessary to

- reflect real medical practice on the actual patient population with the disease under study
- increase the probability of success for identifying clinical benefit of treatment

Include but not limited to

- Modifications of inclusion/exclusion criteria
- Adjustment of study dose or treatment
- Extension of study duration
- Changes in study endpoints
- Modifications in study design based on interim analysis

REGENERON
Biostatistics and Data Management

Interim Analysis (IA)

- *"Any examination of data obtained in a study while that study is still ongoing, and is not restricted to cases in which there are formal between-group comparisons"* – FDA Guidance on Adaptive Design (2010)
- Reasons for interim analysis
  - Ethical
  - Administrative
  - Economic
- Types of interim analysis
  - Efficacy vs. Safety vs. Other
  - Blinded vs. Unblinded
- Multiple stages are formed with interim analysis

REGENERON
Biostatistics and Data Management

Adaptive Design

- *"A study design that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data"* – FDA Guidance on Adaptive Design (2010)
- Adaptations based on interim analysis
  - Dose escalation/de-escalation
  - Early stopping for superiority or futility
  - Sample size re-estimation
  - Outcome-adaptive randomization
  - Study population enrichment
  - Drop or add study arms
  - ... ...

REGENERON
Biostatistics and Data Management

Types of Adaptive Designs

- Adaptive Dose-Finding Design
- Group Sequential Design
- Sample Size Re-estimation
- Adaptive Randomization Design
- Drop-Loser and/or Add-Arm Design
- Biomarker-Adaptive Design
- . . . . . .
- Bayesian Design

REGENERON
Biostatistics and Data Management

Statistical Aspects

- Type I error – α control and determination of stopping boundaries
- Type II error – β control and calculation of power or sample size
- Trial monitoring – make decisions based on conditional power (or futility index)
- Analysis after completion of study – calculation of adjusted p-values, unbiased point estimates and confidence intervals

REGENERON
Biostatistics and Data Management

Pro's and Con's

adaptive-pros-and-cons-2.jpg

REGENERON
Biostatistics and Data Management

Group Sequential Design (GSD)

group-sequential-design-v11_EN.png

REGENERON
Biostatistics and Data Management

A Phase III NSCLC Trial

Consider to design a phase III clinical trial for an experimental therapy vs. standard chemotherapy (control) in non-small cell lung cancer (NSCLC) patients, the primary endpoint is overall survival (OS)

- $OS_{ctrl} = 12$ *months*
- The clinically meaningful effect size $HR = 0.75$, (i.e. $OS_{trt} = 16$ *months*)
- Type I error $\alpha = 2.5\%$ (one-sided)
- Power $1 - \beta = 90\%$
- Accrual period of 48 months
- Minimum follow-up period of 12 months

REGENERON
Biostatistics and Data Management

Classic Design with Fixed Sample Size

- Pre-specify accrual and drop out rates
- Total study duration is at least 60 months!
- Sample size
  - Required number of events is 507
  - Required number of patients is 718
- No (formal) interim analysis
- Must wait till the study end to analyze data and make decisions

REGENERON
Biostatistics and Data Management

GSD with Interim Analysis

Can we evaluate efficacy results earlier to make decisions?

- If the experimental therapy truly works, can we complete study early to claim efficacy? – Superiority
- If the experimental therapy does not work, can we terminate study early to avoid harmful patient exposure? – Futility

Solution: Group sequential design with interim analysis

REGENERON
Biostatistics and Data Management

GSD with Interim Analysis

- How many interim analysis?
  - Not too many as interim analysis takes time and efforts
- When to conduct interim analysis?
  - Not too early as information may be too limited for making decisions, at least 25%–35%
  - Not too late (relative to study duration) as benefit of interim analysis diminishes
- Types of interim analysis?
  - Superiority only
  - Futility only
  - Both superiority and futility (binding vs. non-binding)

REGENERON
Biostatistics and Data Management

Statistical Issues

- Repeated significance testing with interim analysis
  - Claim efficacy after 1st interim analysis
  - Claim efficacy after 2nd interim analysis if study continues after 1st interim analysis
  - .....
  - Claim efficacy after final analysis if study continues after all interim analysis
- Multiple looks of superiority inflate family-wise error rate (FWER) of type I error (introduce bias)
- Multiple looks of futility inflate FWER of type II error (decrease power)
- Implementation of interim analysis for confirmatory trials must be done by an independent data monitoring committee (IMDC)

REGENERON
Biostatistics and Data Management

GSD for NSCLC Trial

Consider to modify the classic design for NSCLC trial to a group sequential design with

- One interim analysis at 50% information (i.e. number of events)
- Both superiority and futility at interim analysis
- FWER control methods
  - Pocock bounds
  - O'Brien-Fleming bounds
  - Spending function approach (Hwang-Shih-DeCani family)

REGENERON
Biostatistics and Data Management

GSD with Pocock Bounds

- Terminate study for superiority if $HR \leq 0.79$ at interim analysis
- Terminate study for futility if $HR \geq 0.89$ at interim analysis
- Continue study if $0.79 < HR < 0.89$ at interim analysis
- Claim efficacy if $HR \leq 0.84$ after final analysis
- Required number of events increases (from 507) to 637

REGENERON
Biostatistics and Data Management

GSD with Pocock Bounds

Analysis	Value	Efficacy	Futility
IA 1: 50%	Z	2.1570	1.0313
N: 768	p (1-sided)	0.0155	0.1512
Events: 319	HR at bound	0.7852	0.8908
Month: 36	P(Cross) if H_0 true (HR=1)	0.0155	0.8488
	P(Cross) if H_1 true (HR=0.75)	0.6600	0.0620
Final	Z	2.2010	2.2010
N: 902	p (1-sided)	0.0139	0.0139
Events: 637	HR at bound	0.8399	0.8399
Month: 60	P(Cross) if H_0 true (HR=1)	0.0229	0.9771
	P(Cross) if H_1 true (HR=0.75)	0.9000	0.1000

- Approx. 66% chance to claim superiority at IA if therapy is efficacious
- Approx. 85% chance to claim futility at IA if therapy is not efficacious
- Duration of study reduced to 36 months if either superiority or futility is claimed at IA

GSD with O'Brien-Fleming Bounds

- Terminate study for superiority if $HR \leq 0.69$ at interim analysis
- Terminate study for futility if $HR \geq 0.97$ at interim analysis
- Continue study if $0.69 < HR < 0.97$ at interim analysis
- Claim efficacy if $HR \leq 0.84$ after final analysis
- Required number of events increases (from 507) to 520

REGENERON
Biostatistics and Data Management

GSD with O'Brien-Fleming Bounds

Analysis	Value	Efficacy	Futility
IA 1: 50%	Z	2.9626	0.2670
N: 626	p (1-sided)	0.0015	0.3947
Events: 260	HR at bound	0.6923	0.9674
Month: 36	P(Cross) if H_0 true (HR=1)	0.0015	0.6053
	P(Cross) if H_1 true (HR=0.75)	0.2604	0.0200
Final	Z	1.9686	1.9686
N: 736	p (1-sided)	0.0245	0.0245
Events: 520	HR at bound	0.8413	0.8413
Month: 60	P(Cross) if H_0 true (HR=1)	0.0243	0.9757
	P(Cross) if H_1 true (HR=0.75)	0.9000	0.1000

- Approx. 26% chance to claim superiority at IA if therapy is efficacious
- Approx. 61% chance to claim futility at IA if therapy is not efficacious
- Duration of study reduced to 36 months if either superiority or futility is claimed at IA

Pocock vs O'Brien-Fleming Bounds

Design Parameters	Pocock	O'Brien-Fleming
HR bounds at IA	(0.79, 0.89)	(0.69, 0.97)
α spending at IA	0.0155	0.0015
Pr(stop for superiority) at IA	66%	26%
Pr(stop for futility) at IA	85%	61%
Events/Sample Size	637 / 902	520 / 736

- Pocock bounds spends more α at IA , thus more aggressive to claim superiority/futility
- O'Brien-Fleming bounds is more conservative in claiming efficacy/futility at IA, reserving more α for final analysis
- O'Brien-Fleming bounds requires fewer event/sample size than that of Pocock bounds

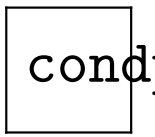
Flexible GSD with Spending Function

- Balance of aggressive/conservative IA
- Spending α as a function of the observed information levels
- Interim analysis may occur at any times with spending function
- Number of interim analyses may change
- Operational and logistical restrictions

REGENERON
Biostatistics and Data Management

Conditional Power

Given a normal test statistic from IA, the conditional power curves under observed effect size (ES), H_0 and H_1



- Probability of rejecting H_0 (claim efficacy) during the rest of the trial based on accumulated data at IA
- Commonly used for monitoring an ongoing trial
- Maybe utilized for sample size re-estimation

REGENERON
Biostatistics and Data Management

Sample Size Re-estimation

promising.jpg

[3]

[3] Cytel Presentation (2012)

Sample Size Calculation

- Sample size calculation based on early phase trial results or historical data at the design stage
  - A clinically meaningful effect size
  - Variability associated with the effect size (and other nuisance parameters)
- What if the effect size and/or the associated variability were incorrectly specified in the NSCLC trial?
  - If $OS_{ctrl} = 15$ months and $OS_{trt} = 20$ months (still HR=0.75)
  - number of event 507, sample size 795 to achieve 90% power
  - If $OS_{ctrl} = 12$ months and $OS_{trt} = 15$ months (HR=0.80)
  - number o event 844, sample size 1300 to achieve 90% power

REGENERON
Biostatistics and Data Management

Sample Size Re-estimation

Can we plan a sample size re-estimation after interim analysis to overcome under-power or over-power in initial design of the NSCLC trial?

- If the study is under-powered based on interim analysis, increase sample size
- If the study is over-powered based on interim analysis, reduce sample size (though rarely done)

Solution: Sample size re-estimation after interim analysis

- N-adjusted clinical trial design (straightforward)
- Integrated with group sequential design (complex)
- Types of sample size re-estimation based on interim analysis
  - Blinded
  - Unblinded

REGENERON
Biostatistics and Data Management

Blinded SSR for NSCLC Trial

Consider to modify the classic design for NSCLC trial to a sample size re-estimation design with

- Sample size re-estimation after an interim analysis at 50% information (i.e. number of events)
- Interim analysis is blinded without any knowledge of treatment assignment
- Interim analysis is not intended for superiority or futility
- Significance level does not need to be adjusted for blinded interim analysis

REGENERON
Biostatistics and Data Management

Blinded SSR for NSCLC Trial

- Specified maximum sample size inflation was 100%
- Assumed enrollment overrun at interim analysis was 25 patients
- Observed unblinded median $OS = 17.5$ months at the interim analysis
- A heuristic calculation

Stage	IA	FA	SSR
No. of Events	254	507	507
Sample Size	359	718	794
Overrun	25	0	0

- Hence, blinded SSR suggest to increase sample size (initial design) by $794 - 718 = 76$ patients for final analysis (FA)

REGENERON
Biostatistics and Data Management

Unblinded SSR for NSCLC Trial

- May provide more accurate sample-size estimation based on the estimated effect size at interim analysis
- Bias results from knowledge of observed effect size at interim analysis
- Statistical approaches to control FWER
  - Combination test
  - Conditional error function
  - Conditional power (CP)

REGENERON
Biostatistics and Data Management

GSD with Unblinded SSR

Table: Asymmetric two-sided group sequential design with non-binding futility bound, sample size 833. Efficacy bounds derived using a HSD spending function with gamma = -4. Futility bounds derived using a HSD spending function with gamma = 1.

Analysis	Value	Efficacy	Futility
IA 1: 50%	Z	2.7500	0.9316
N: 417	p (1-sided)	0.0030	0.1758
	HR at bound	0.7257	0.8971
	P(Cross) if HR=1	0.0030	0.8242
	P(Cross) if HR=0.75	0.3889	0.0622
Final	Z	1.9811	1.9811
N: 833	p (1-sided)	0.0238	0.0238
	HR at bound	0.8493	0.8493
	P(Cross) if HR=1	0.0211	0.9789
	P(Cross) if HR=0.75	0.9000	0.1000

REGENERON
Biostatistics and Data Management

GSD with Unblinded SSR

- Maximum sample size inflation is specified as 100%
- Assumed enrollment overrrun at IA is 25 patients
- Promising zone in CP interval $(0.36, 0.9)$ where SSR to be conducted

REGENERON
Biostatistics and Data Management

GSD with Unblinded SSR

ssr.pdf

GSD with Unblinded SSR

Based on the GSD with interim analysis for superiority and futility

- Futility – Stop after interim analysis with actual sample size of 417
- Superiority – Stop after interim analysis with actual sample size of 417

Otherwise, based on the conditional power at the interim analysis,

- CP < 0.36 unfavorable – Continue the study after interim analysis without SSR, sample size is still 833
- CP $\in [0.36, 0.9]$ promising zone – Increase sample size $833 < N^* \leq 833 \times 2 = 1666$
- CP > 0.9 favorable – Continue the study after interim analysis without SSR, sample size is still 833

REGENERON
Biostatistics and Data Management

Blinded vs. Unblinded SSR

Design Parameters	Blinded	Unblinded
FWER Control	No adjustment	Adjustment
Stat Methods	Straightforward	Complex
Implementation	In-house	External
FDA guidance	Well-understood	Less well-understood

REGENERON
Biostatistics and Data Management

Regulatory Guidelines

- PhRMA (2006) – Adaptive designs in clinical drug development - an executive summary of the PhRMA working group
- EMA (2007) – Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design
- FDA (2010) – Guidance for the use of Bayesian statistics in medical device clinical trials
- FDA (2010) – Adaptive design clinical trials for drugs and biologics
- FDA (2015) – Adaptive designs for medical device clinical studies

REGENERON
Biostatistics and Data Management

FDA Draft Guideline 2010

- Distributed in February, 2010, expect to publish final document in 2017
- Endorsed by both CDER and CBER for drugs and biologics
- Well-understood designs
  - Group sequential design
  - Sample size re-estimation with blinded interim analysis
- Less well-understood designs
  - Adaptive dose-selection, sample size re-estimation with unblinded interim analysis, adaptive randomization, adaptive population, endpoint selection, ...

REGENERON
Biostatistics and Data Management

Challenges

- Requirements of pre-specified vs. unplanned adaptations
- Timing of interim analysis vs. patient accrual
- Time and efforts in designing a complex adaptive design

REGENERON
Biostatistics and Data Management

Operations

- Early interaction with FDA
- Extensive simulation studies for evaluation
- Documentation in protocol and SAP
- Available software and/or packages

REGENERON
Biostatistics and Data Management

Summary

- Not intended for adaptations due to poor planning in design stage
- Improves efficiency when used appropriately
- Currently more acceptable in early-phase drug development when information is limited
- Important to communicate with clinical colleagues and FDA

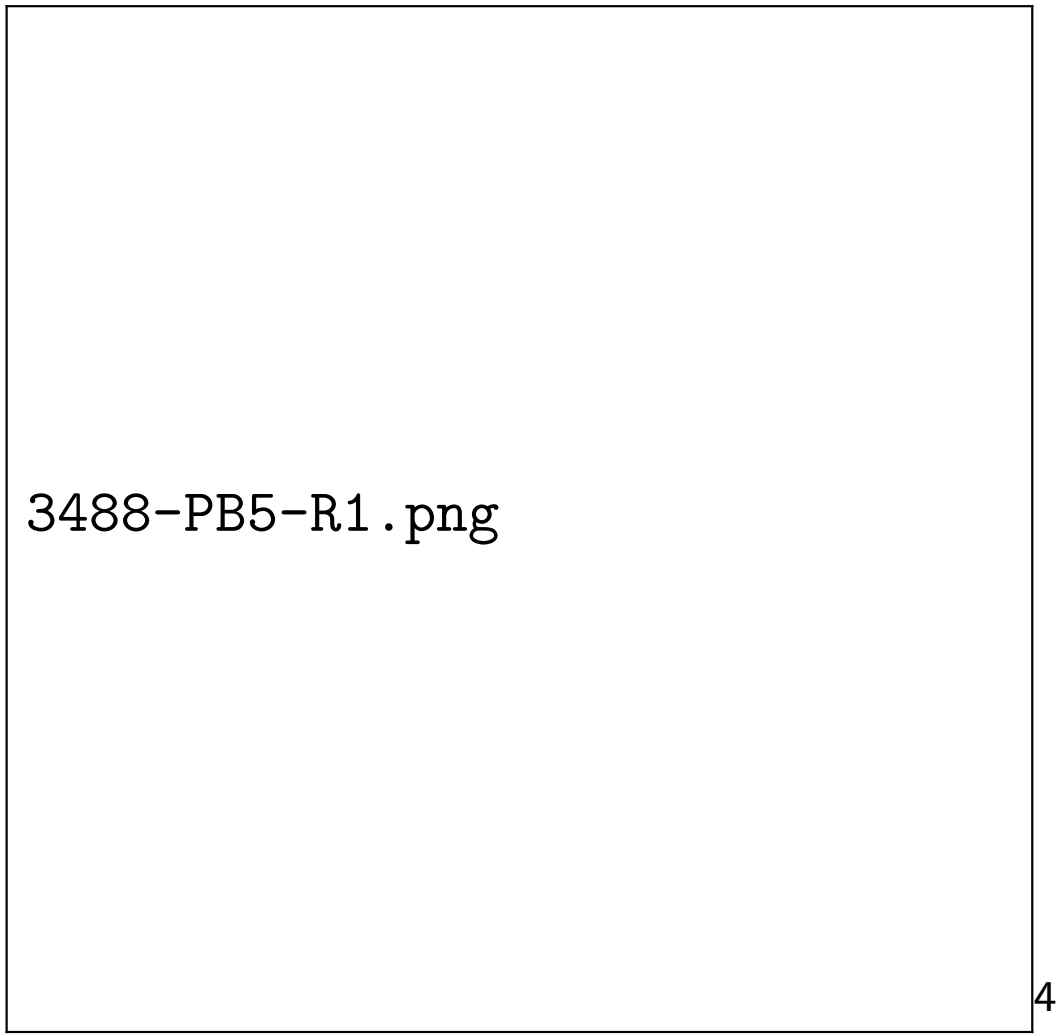
REGENERON
Biostatistics and Data Management

References

- Jennison and Turnbull (2000) *Group Sequential Methods with Applications to Clinical Trials*, Chapman & Hall
- Chang (2008) *Adaptive Design Theory and Implementation using SAS and R*, Chapman & Hall
- Mehta and Pocock (2009) *Adaptive increase in sample size when interim results are promising: a practical guide with examples*, Stat in Medicine
- FDA draft guidance (2010) *Adaptive design clinical trials for drugs and biologics*
- Anderson (2014) R package *gsDesign: Group Sequential Design*

REGENERON
Biostatistics and Data Management

Adaptive Dose-Finding Design

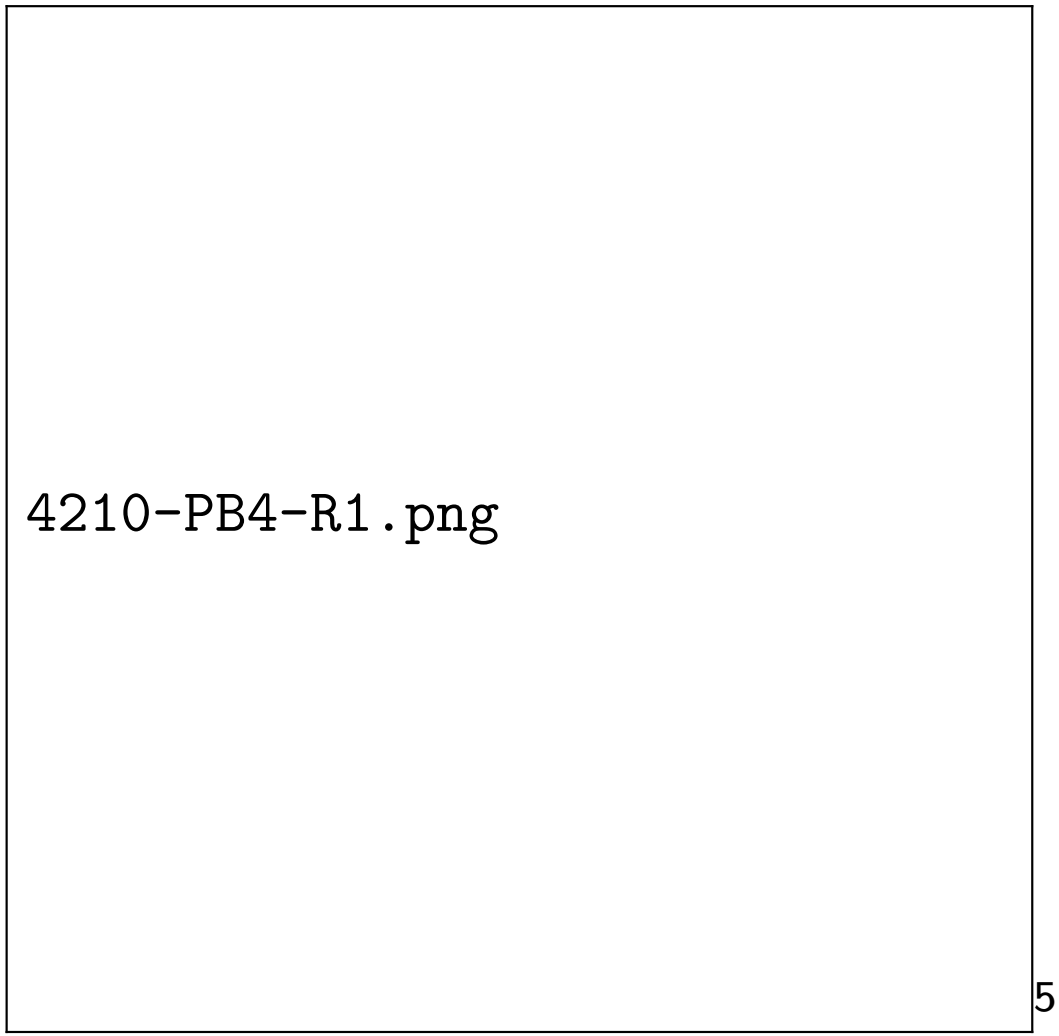


[4]

[4] Figure 3. of Braun (2014) Chinese Clinical Oncology

REGENERON
Biostatistics and Data Management

Adaptive Randomization Design

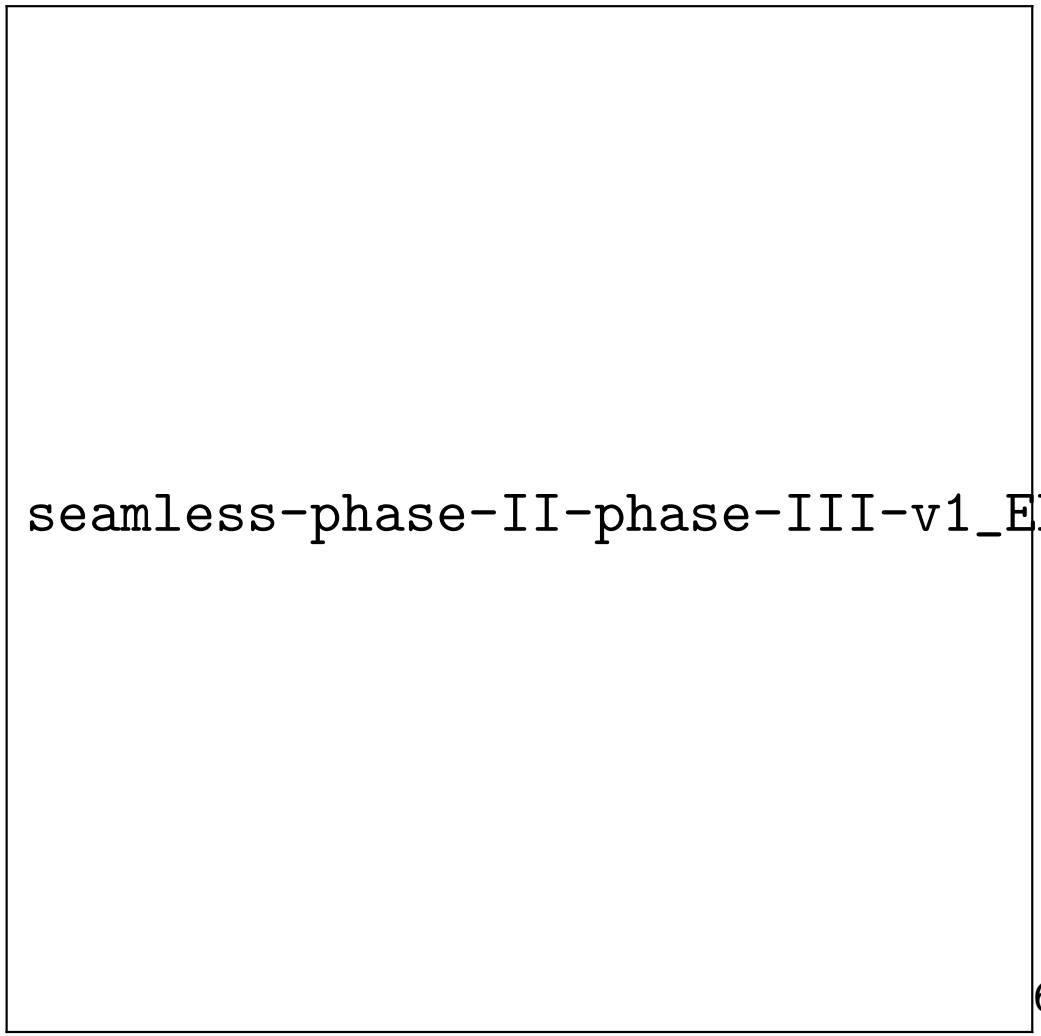


[5]

[5]Figure 2. of Zang (2014) Chinese Clinical Oncology

REGENERON
Biostatistics and Data Management

Drop-Loser Design

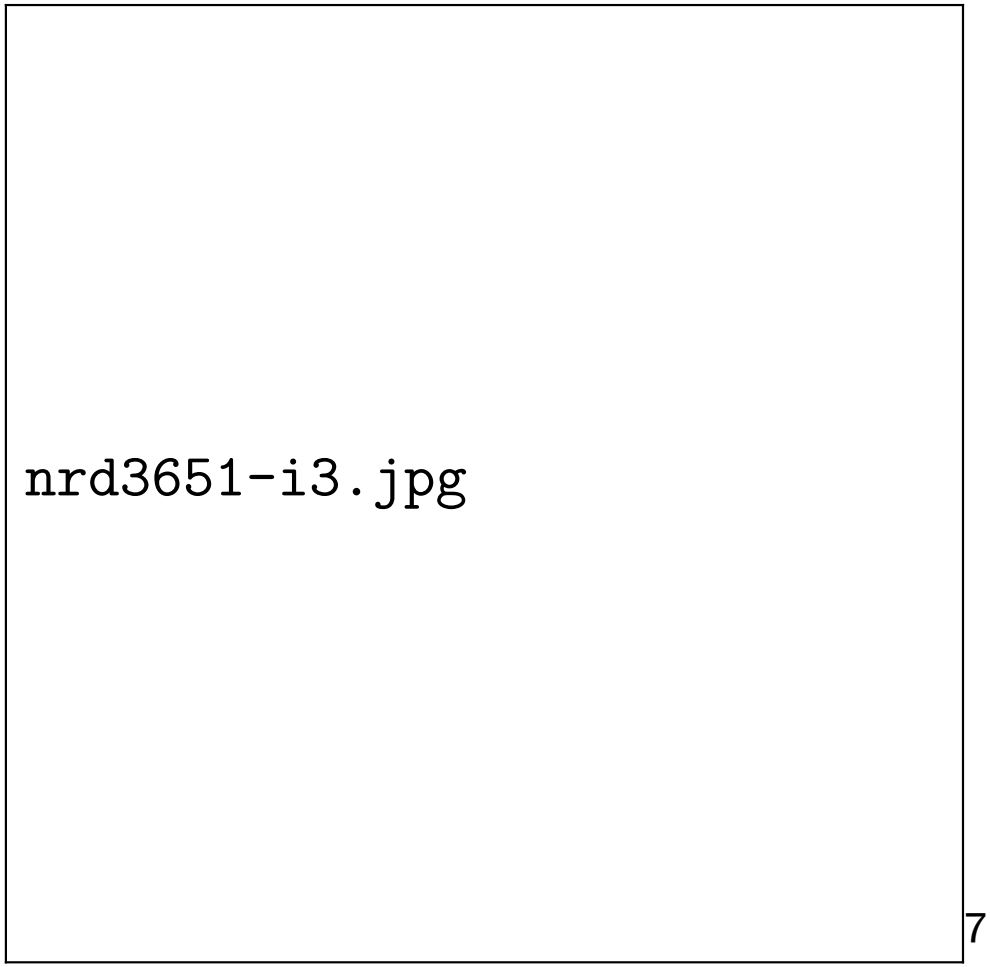


[6]

[6] Figure 3. of EUPATI (2015)

REGENERON
Biostatistics and Data Management

Biomarker-Adaptive Design



[7]

[7] Figure 5. of Kelloff (2012)

REGENERON
Biostatistics and Data Management

Bayesian Design

beyond-traditional-designs-in-early-drug-development-5-728

REGENERON
Biostatistics and Data Management