Statistical Considerations for Phase II Trials

Emily Van Meter, PhD
Biostatistics Shared Resource Facility, Markey Cancer Center
Assistant Professor, Division of Cancer Biostatistics
University of Kentucky

2014 Society of Clinical Trials Conference
Workshop P2: Enhancing the Design and Conduct of Phase II Studies
Philadelphia, PA
May 18, 2014

UKHealthCare
Markey Cancer Center

Objectives and Goals

Traditional Designs

Randomized Designs

PHASE II DESIGNS

Portions of this presentation adapted from:
Shyr, Y, 2012 ASCO/AACR Clinical
Cancer Research Methods Workshop

OBJECTIVES AND GOALS OF PHASE II TRIALS

Raise standards for preclinical cancer research

- Our ability to translate cancer research to clinical success has been remarkably low.
- Sadly, clinical trials in oncology have the highest failure rate compared with other therapeutic areas.

REPRODUCIBILITY OF RESEARCH FINDINGS

Preclinical research generates many secondary publications, even when results cannot be reproduced.

Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development program.

Why?

- Difficult nature of the disease and limitation of preclinical tools (i.e. cell-line and mouse models)
- Issues related to clinical trial design
  - Uncontrolled phase II studies
  - Reliance on standard criteria for evaluation tumor response
  - Challenges of selecting patient prospectively

Background

- Advent of molecularly targeted agents in oncology
- Not targeting cell killing (tumor response endpoint)
- PFS and OS are more appropriate endpoints
- Tumor response can fail to predict survival benefit

Background

- Choice of Randomized vs. Single Arm (compared with Historical Controls) depends on circumstances of individual trial
- Historical controls – appropriate for tumor response unaffected by miscellaneous prognostic factors; more readily available estimates from historical data

Background

- Randomized Trials - Protects against biases as explanation for results from phase II trials
- Randomized Trials – needed for PFS or OS
  - PFS or OS sensitive to historical data due to:
    - Changes in standard of care over time
    - Inter-institutional variability in follow-up
    - Differences in prognostic factors
- But randomization will require a lot more patients than a single arm study!

ASCO Perspective: Raising the Bar for Clinical Trials by Defining Clinically Meaningful Outcomes

- It is necessary to observe extremely strong signals in phase II studies
  - If we expect clinically meaningful outcomes to be achieved in subsequent phase III studies
- Sometimes results from phase II trials are more optimistic than warranted
- It is even possible that phase III studies will not be necessary if results from well-conducted phase II trials demonstrate exceptional activity that clearly benefits patients.[11]

Ellis LM, Bernstein DS, et al. JCO 2014: 32(12)

Recommendations

Table 1. Summary of Recommended Targets for Meaningful Clinical Trial Goals

Cancer Type	Patient Population	Current Baseline Median OS (months)	Primary End Point: Improvement Over Current OS That Would Be Clinically Meaningful (months)	Primary End Point: Target HRs	Secondary End Point: Improvement in 1-Year Survival Rate (%)*	Secondary End Point: Improvement in PFS (months)
Pancreatic cancer	FOLFIRINOX-eligible patients	10 to 11[19]	4 to 5	0.67 to 0.69	48 → 63	4 to 5
Pancreatic cancer	Gemcitabine or gemcitabine/nab-paclitaxel–eligible patients	8 to 9[20,21]	3 to 4	0.6 to 0.75	35 → 50	3 to 4
Lung cancer	Nonsquamous cell carcinoma	13[22]	3.25 to 4	0.76 to 0.8	53 → 61	4
Lung cancer	Squamous cell carcinoma	10[23]	2.5 to 3	0.77 to 0.8	44 → 53	3
Breast cancer	Metastatic triple negative, previously untreated for metastatic disease	18[24,25]	4.5 to 6	0.75 to 0.8	63 → 71	4
Colon cancer	Disease progression with all prior therapies (or not a candidate for standard second- or third-line options)	4 to 6[26]	3 to 5	0.67 to 0.67	25 → 35	3 to 5

Abbreviations: FOLFIRINOX, leucovorin, fluorouracil, irinotecan, and oxaliplatin; HR, hazard ratio; OS, overall survival; PFS, progression-free survival.
*Current → target.

- The goals established will likely require biomarker enrichment strategies to achieve them
- Validated biomarkers are not currently available to select patients for treatment with specific drugs
- We expect that over time, such biomarkers will be identified and that the goals set forth by these working groups will be achievable

Goals of Phase II Trials

- Provide initial assessment of efficacy or 'clinical activity'
  - Screen out ineffective drugs
  - Identify promising new drugs for further evaluation
- Further define safety and toxicity
  - Type
  - Frequency

Important Design Considerations in Phase II trials

- Minimize cost of the trial
  - Minimize number of patients exposed to an ineffective treatment
  - Enroll as few patients as “necessary” to show benefit or failure

Phase II trials

- Trials of investigational new drugs
  - Assess response for a particular disease type in addition to safety
- Phase II pilot studies
  - Usually done to assess previously tested treatments
    - Using a new treatment schedule
    - Using the drug in combination with other agents

Frequentist approaches

Two-stage Designs

Bayesian Methods

TRADITIONAL SINGLE-ARM PHASE II DESIGNS

Sample Size Considerations

- There are many different ways to calculate sample size!
  - Dependent on primary endpoint selection
    - Usually use a single response variable for the primary endpoint
    - Any additional variables will need to be included in the sample size calculation

- Factors that influence sample size
  - Type I error (α): Usually set at 5%
  - Power (1-β): Usually set 80-90%
  - Difference between treatment sizes
    - Small difference between groups: **LARGER SAMPLE SIZE!**
    - Bigger differences: **SMALLER SAMPLE SIZE!**
  - Patient accrual, participation, and loss to follow-up

Standard Single Arm Phase II Study

- **Single arm**:
- Comparison is "**fixed**" constant
- Binary endpoint (clinical response vs. no response)
- Simple set-up:

$$\alpha = 0.10$$

$$\beta = 0.10 \ (\text{power} = 0.90)$$

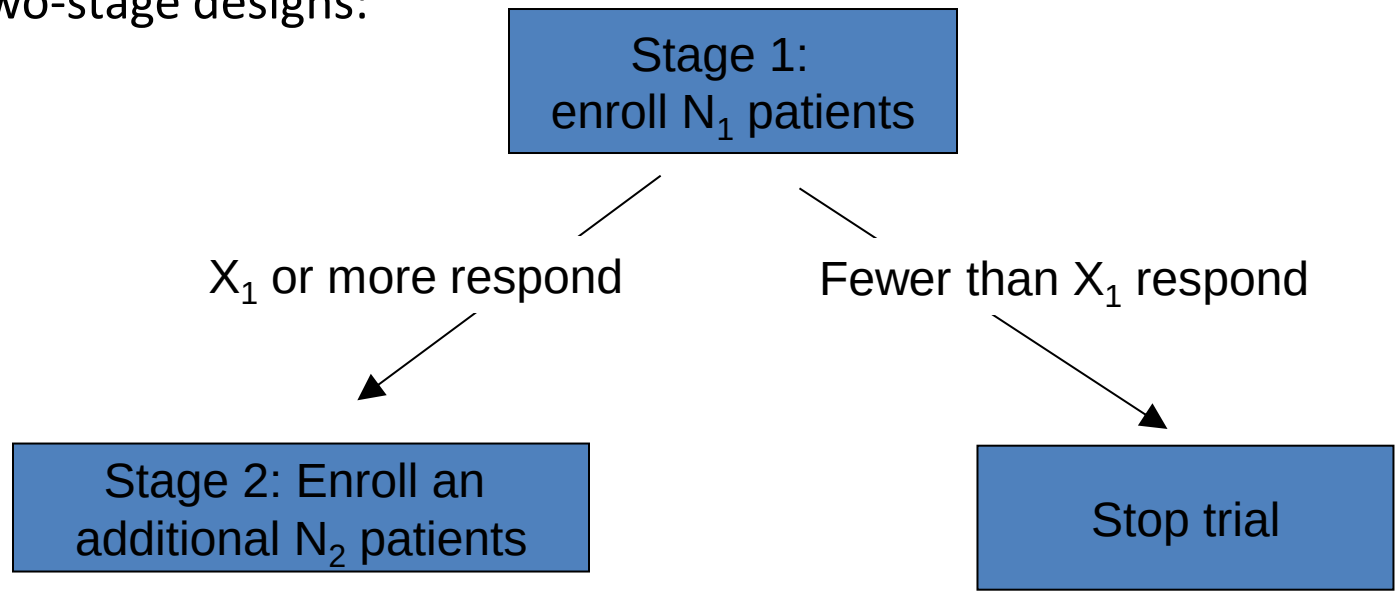
$$H_0 : p = 0.20 \ (\text{null reponse rate})$$

$$H_1 : p = 0.40 \ (\text{target response rate})$$

- Based on design parameters:
  - N=36
  - Conclude effective if 11 or more responses (i.e., observed response rate of ≥0.31)

Two-Stage Designs

- *What if by the 15th patient you've seen no responses?*
- *Is it worth proceeding?*
- Maybe you should have considered a design with an **early stopping rule**
- Two-stage designs:



Revised Design

- Stage 1: enroll **19** patients
  - If **4** or more respond, proceed to stage 2
  - If **3** or fewer respond, stop
- Stage 2: enroll **17** more patients (total N=36)
  - If **11** or more of total respond, conclude effective
  - If **10** or fewer of total respond, conclude ineffective
- Design properties?

$$\alpha = 0.10$$

$$H_0 : p = 0.20 \text{ (null reponse rate)}$$

$$H_1 : p = 0.40 \text{ (target response rate)}$$

- What about power compared to standard single arm study?

Same as before!

Two-stage Designs

- Simon two-stage (1989)
  - Used in example
  - MANY designs fit the criteria
  - “Optimal” & “MinMax”
    - Minimum expected sample size under H0
    - Minimum maximum sample size
  - Preserves alpha and power, and permits early look

Two-stage Designs

- Gehan two-stage Design (1961)
  - It is a two-stage design for estimating the response rate but providing for early termination if the drug shows insufficient antitumor activity.
  - The design is most commonly used with a first stage of 14 patients. If no responses are observed, the trial is terminated.

- Fleming two-stage Design (1982)
  - Fleming's design is the only two-stage design that we cover that may have the early termination with the "accept the drug" conclusion.

Bayesian Interim Monitoring in Phase II Trials

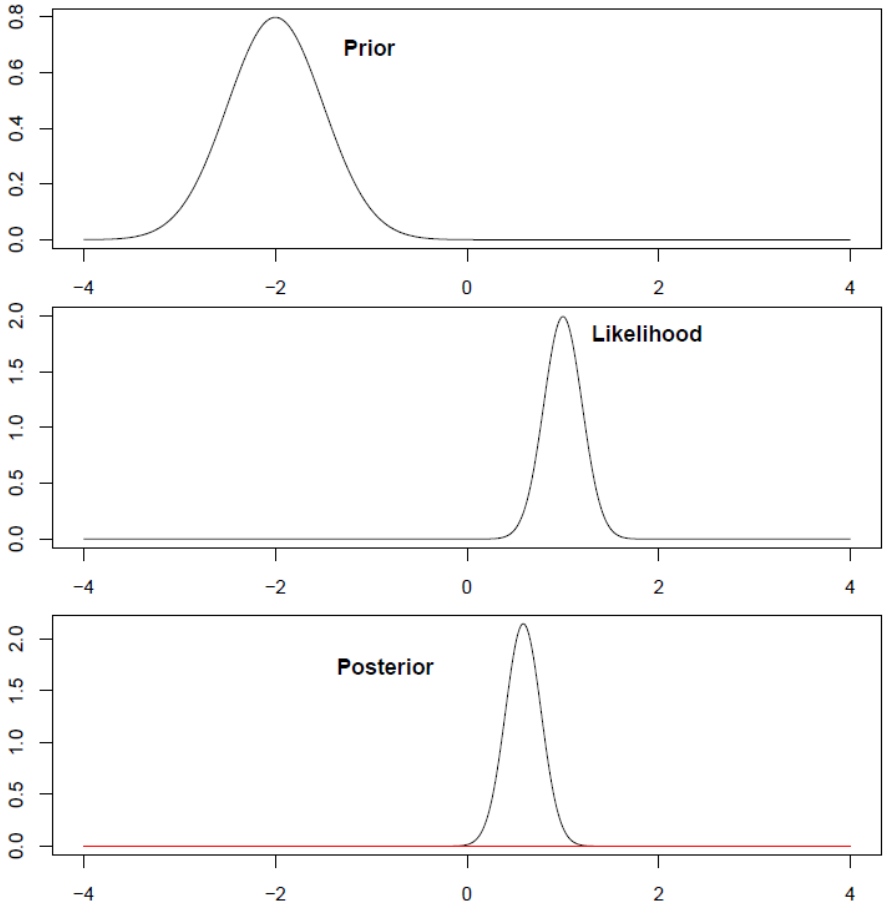
- What is Bayesian analysis???
  - To see the future, one must look at the past…
  - Bayes theorized that the probability of future events could be calculated by determining their earlier frequency
  - No confidence intervals and p-values!
  - Can look as often as you like!

- To evaluate the probability of a hypothesis:
  - Specify some prior probability
  - Update the information in light of new relevant data

- Bayesian probability is not interpreted in the same way as a p-value!!!
  - Higher Bayesian probability (closer to 1) means you have a better chance that it will actually occur
  - Pr(t>10) = 0.99 means you have a 99% probability of actually observing $t > 10$

Priors, Likelihoods, Posteriors… OH MY!

General philosophy: Start, Observe, and Combine!



The prior is the distribution we predict before the start of an experiment when we don't have any true data

The Likelihood is what we observe in our data… this can be over the whole trial, 1st half of the trial, etc…

The posterior takes information from both the prior distribution and the observed likelihood function to re-estimate a distribution based on all of the information we have at this point!

Bayesian Interim Monitoring in Phase II Trials

- Uses Bayesian predictive probabilities design that controls type I and II errors and allows for continuous monitoring to calculate the posterior distribution and the corresponding predicative probability (PP) as outlined in Lee and Liu, 2008.
- PP calculates the probability of concluding a positive result by the end of the trial based on the cumulative information in the current stage
- The protocol will be stopped early for any PP<0.05 or PP>0.95 for futility or superiority respectively

Lee JJ, Liu DD. Clinical Trials 2008: 5 (93)

Trial 1: Phase II trial of sorafenib + endocrine therapy in patients with advanced breast cancer

Outcome: Clinical response rate (CR+PR) n = 43 patients

H_0: $p \leq 0.10$ vs. H_a: $p \geq 0.25$

Response After 9 Patients	Prior Distribution (Response Rate)	Predictive Probability (After 43 Patients)	Lower Bound for futility (After 9 patients)	Lower Bound for futility (After 25 patients)	Lower Bound for futility (After 43 patients)
0/9	**Skeptical** Beta (0.1, 0.9)	0.005	<1/9	<3/25	< 8/43
0/9	**Neutral** Beta (0.25, 0.75)	0.016	<1/9	< 3/25	< 8/43
0/9	**Uniform** Beta (1, 1)	0.14	0/9	< 3/25	< 7/ 43

Stop trial early for futility!

Second Example: PFS at 7 months

Trial 2: BRE-43: H_0: $p \leq 0.25$ vs. H_a: $p \geq 0.50$

Progression -Free @ 7 mos	Prior Distribution (Response Rate)	Predictive Probability After 44 Patients	Lower Bound for futility (After 10 patients)	Lower Bound for futility (After 25 patients)	Lower Bound for futility (After 44 patients)
3/9	**Skeptical** Beta (0.25, 0.75)	0.46	< 2/9	< 7/25	< 15/44
3/9	**Neutral** Beta (0.5, 0.5)	0.52	< 2/9	< 7/25	< 15/44
3/9	**Uniform** Beta (1, 1)	0.56	< 2/9	< 7/25	< 15/44

Continue trial accrual!

Screening Design

Selection Design

Adaptive/Bayesian approaches

RANDOMIZED PHASE II DESIGNS

Randomized Phase II

- Why randomized?
  - Want to explore efficacy
  - Not willing to invest in phase III (yet)
  - Want some “control” or “prioritization”
  - Primarily two different kinds of randomized phase II studies
    - Phase II selection design (prioritization)
    - Phase II designs with reference control arm (control)

Randomization

- Tends to produce comparable study groups with respect to known and unknown risk factors
- Removes investigator bias in the allocation of participants
- Guarantees that statistical tests will have valid significance levels

Phase II Selection Design (prioritization)

- Two parallel one arm studies (classic case)
- Do not directly compare arms to each other in general
- Compare each to “null rate”
  - Why? To compare to each other, you’ll need a study at least two times as large
- “Pick the Winner” (Simon, 1985)
  - Appropriate to use when:
    - Selecting among NEW agents
    - Selecting among different schedules or doses
  - NOT appropriate when:
    - Trying to directly compare treatment efficacies (not powered)
    - Second arm is a true placebo control

Simon R, Wittes RE, Ellenberg SS. Cancer Treatment Reports 1985: 69(12)

Pick the winner (continued)

- Uses 2+ Simon two-stage designs
  - Each arm is compared to a null rate
  - Must satisfy efficacy criteria of Simon design
  - Move the “winner” to phase III
  - Only have to pick winner if more than one arm shows efficacy
- Can be used when the goal is prioritizing which (if any) experimental regimen should move to phase III when no a priori information to favor one

Selection Design Considerations

- Goal is to narrow choice of agent and/or combination/schedule in Phase III
- Sample size based on selecting superior treatment with high probability (90%)
- Always picks a winner – even with nominal difference
- Beware of proper interpretation (No type I error control – False Positive)
- Software: R package ‘clinfun’ (pselect function)

Selection Design Table – Binary Endpoint

Sample size per treatment with 90% selection probability

Response Rates		n per group		
p1, …, pk-1	pk	K = 2	K=3	K = 4
10%	25%	21 (113)	31	37
20%	35%	29 (151)	44	52
30%	45%	35 (176)	52	62
40%	55%	37 (186)	55	67
50%	65%	36 (183)	54	65

n per group based on hypothesis testing of two proportions, 80% power, 5% alpha, two-sided test

Design A: Selection Design Table – Time to Event Endpoint

Sample size per treatment for exponential survival with 90% selection probability (time in years) (1 year accrual)

Median Time	Follow-up Time	HR	K = 2		K = 3	
			1.3	1.5	1.3	1.5
0.5	0.5		71	31	106	46
	1		59	26	88	38
0.75	0.5		89	40	133	59
	1		70	31	104	46
1	0.5		108	48	161	72
	1		82	36	122	54

Randomized Phase II designs with reference arm (screening)

- Relaxes assumption to allow for a comparison between control and new treatment arm
- Goal is preliminary, non-definitive randomized comparison.
  - Is no way considered definitive, and does not substitute for a phase III trial!
- 1-sided test, relax type I error rates

Rubinstein LV, Korn EL, Freidlin B, et al. JCO 2005: 23(28): 7199-7206

Screening Design Considerations

- Choice of parameter values must be made with great care
  - Large α reduces the screening ability of the study, as the rate of false-positives essentially nullifies the screening effect (little value in the phase II results)
- Large β runs the risk of terminating the study of a potentially useful regimen
- Overly optimistic Δ runs the risk of rejecting a regimen with a more limited, but still clinically significant, benefit

Screening Design Recommendations

- Requires a level of compromise
  - Single arm studies tend to require α and β be no more than 0.1
- Authors suggest α and β equal .20 and Δ equal 1.5 (or a target difference in response rate of 20%) as appropriate design parameters for consideration in phase II screening trials
  - Many argue one-sided alpha=0.20 is too large!

Sample size examples for screening trials

Table 1. Approximate required numbers of observed (total) treatment failures for screening trials with PFS endpoints, using the logrank test

Error rates	Hazard Ratios (Δ)			
	Δ = 1.3	Δ = 1.4	Δ = 1.5	Δ = 1.75
(α,β) = (10%,10%)	382	232	160	84
(α,β) = (10%,20%) or (20%,10%)	262	159	110	58
(α,β) = (20%,20%)	165	100	69	36

NOTE: Calculations were carried out using nQueryAdvisor 5.0 software (Statistical Solutions) based on methods given in Collett (26) with one-sided α.

Table 2. Approximate required numbers of total patients for screening trials with PFS rate (at a specified time) endpoints, using the binomial proportion test

Error rates	PFS Rates at a given time point (with equivalent hazard ratios Δ)			
	20% vs. 35% (1.53)	20% vs. 40% (1.76)	40% vs. 55% (1.53)	40% vs. 60% (1.79)
(α,β) = (10%,10%)	256	156	316	182
(α,β) = (10%,20%) or (20%,10%)	184	112	224	132
(α,β) = (20%,20%)	126	78	150	90

NOTE: Calculations were carried out using nQueryAdvisor 5.0 software (Statistical Solutions) based on methods given in Fleiss et al. (27) with one-sided α.

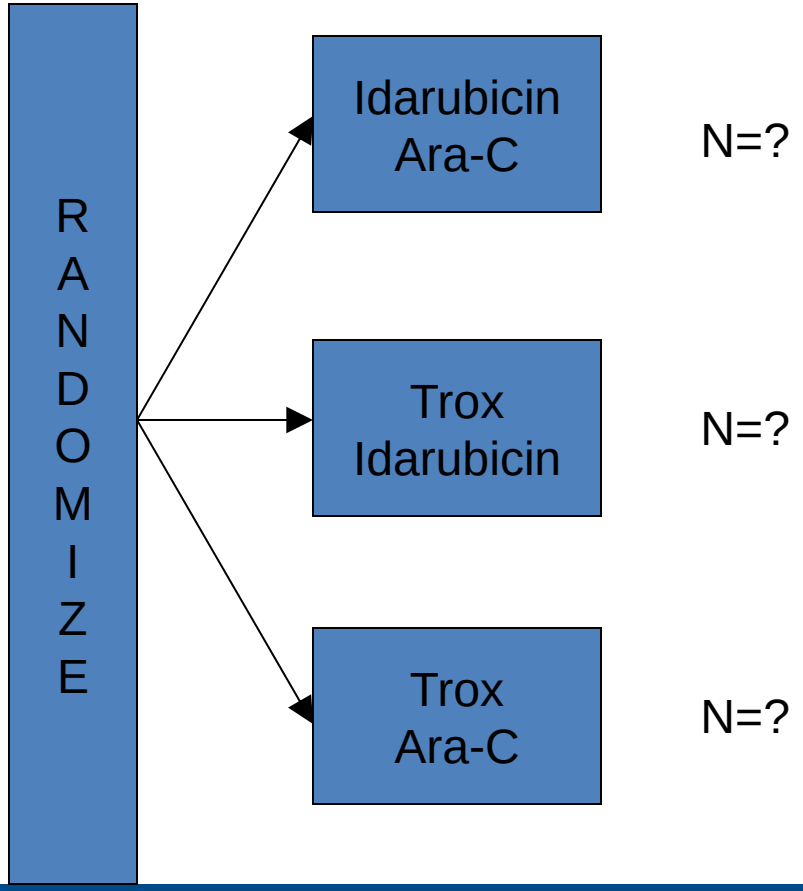
Randomized Phase II designs with reference arm (control)

- Includes reference arm to ensure that historical rate is “on target”
- Reference arm is not directly compared to experimental arm(s) (due to small N)
- Can see if failure (or success) is due to incomparability of patient populations

Adaptive Randomization Designs

- Randomization is “adapted” based on accumulated information
- Adaptive on Outcome (Bayesian)
  - Assign treatments according to accumulated information about best treatment. (Berry and Eick, 1995)
- Assign with higher probabilities to better therapies
- Regimens can be dropped if low probability of improved efficacy

Adaptive Designs



Adapt the randomization to learn while effectively treating patients on trial:

(1) Begin by randomizing with equal chance per arm

(2) Then, adjust probability of assignment to reflect the knowledge of the best treatment

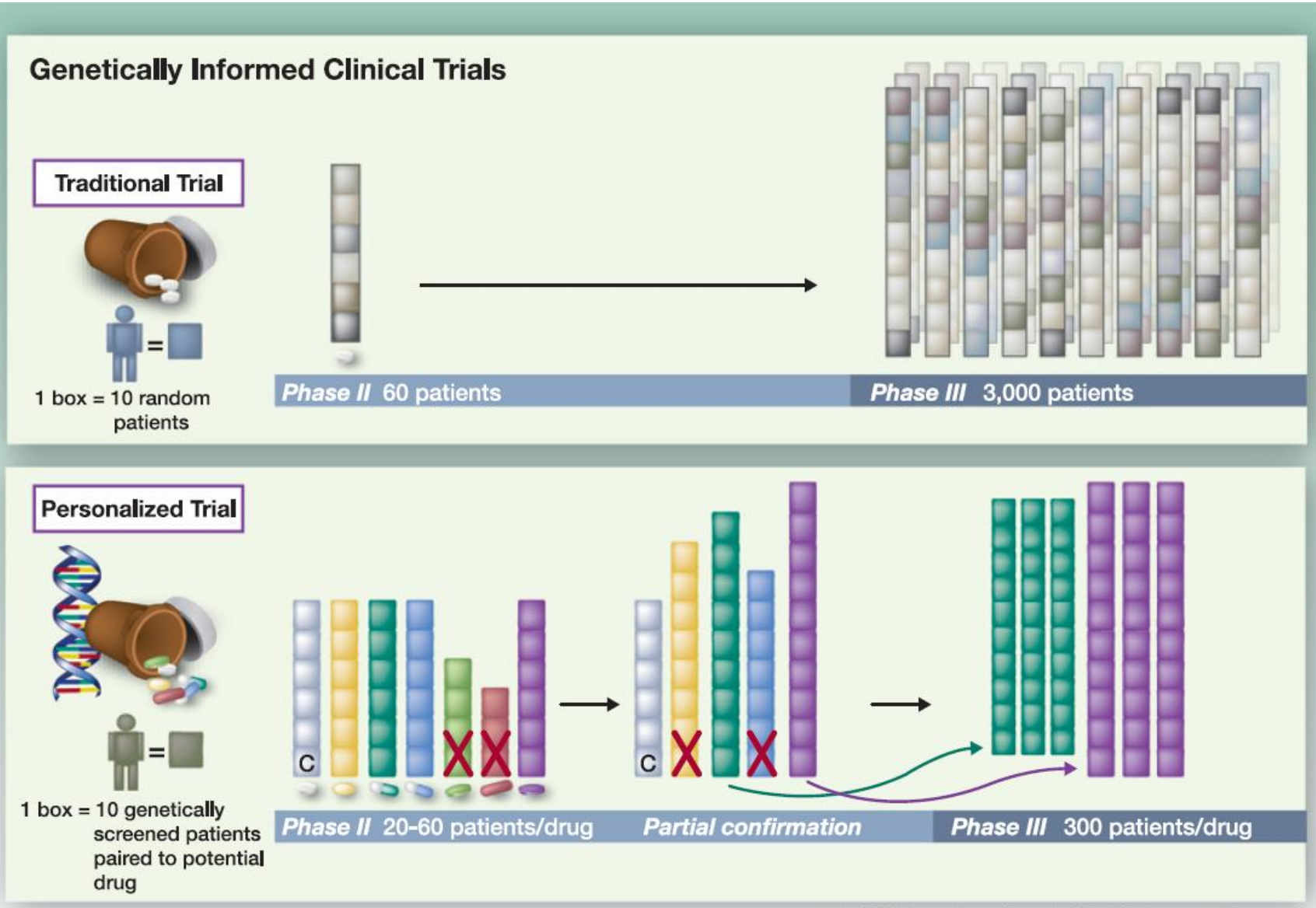
Adaptive Designs

- Begin assuming equally effective (1/3, 1/3, 1/3)
- May wait until a minimum number have been treated per arm
- Based on currently available (accumulated) data, randomize next patient (i.e., “weighted” randomization)
- Stopping rules: drop an arm when there is “strong” evidence that
  - It has low efficacy OR
  - It has lower efficacy than competing treatments

Biomarker - Integrated Approaches of Targeted Therapy for Lung Cancer Elimination (BATTLE)

- A phase II adaptive clinical trial studied four drugs for lung cancer patients based on molecular signatures in tumor biopsies.
- BATTLE's end point was disease control at eight weeks.
- BATTLE employed an adaptive randomization approach (Bayesian based method).
- The model leads to greater use of successful drugs and minimization or dropping of those less successful.

Clin Cancer Res; 18(3) February 1, 2012



Issues with innovative designs

- Statistically intensive
  - "buy your statistician a beer (or bourbon)"
    - Yu Shyr
  - Probably cannot be used "off-the-shelf"
  - require specialized software
- Need to be validated
  - do they behave as promised?
  - are they 'robust' (i.e., do they work when incorrect assumptions are made)?

SEAMLESS PHASE II/III DESIGNS

Why seamless phase II/III designs?

1. No lead time between the learning (phase II) and confirmatory phases (phase III)
2. Data collected at the learning phase is combined for final analysis.
3. Impact on overall alpha level and power:
   - Traditional phase II then phase III: $\alpha = \alpha_{II}\alpha_{III}$, $power = power_{II}power_{III}$
   - Seamless phase II/III: $\alpha = \alpha_{III}$, $power = power_{III}$
   - Alpha and power for seamless are: $1/\alpha_{II}$ and $1/power_{II}$ larger

Paper by Hunsberger, Zhao, Simon

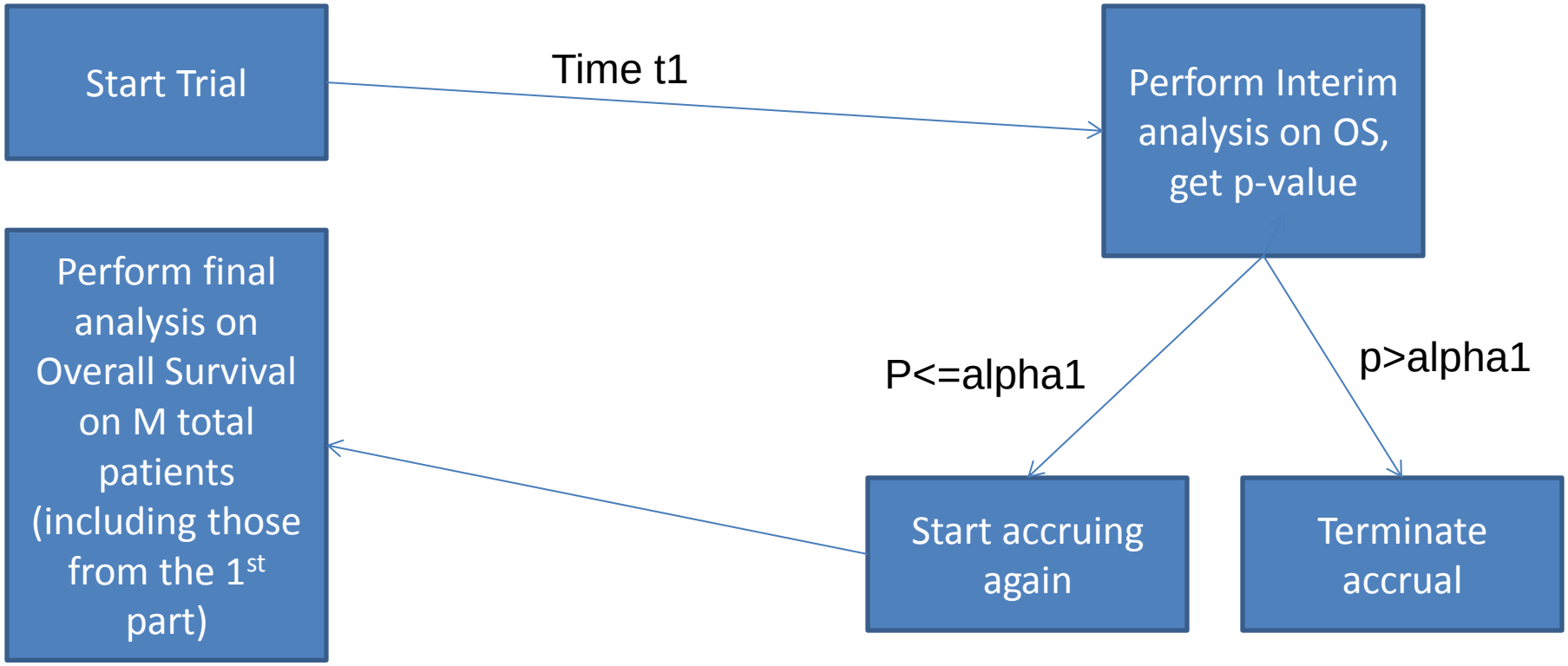
- Compared 3 different study designs:

1. Single arm phase II study design using historical controls
2. Randomized phase II study
   - PFS is primary endpoint, 90% power, one-sided 0.1 alpha level test
   - Ran simulations based off a pancreatic cancer study with median PFS equal to 3 months
   - Wanted to see an improvement to 4.5 months (hazard ratio = 1.5)

Hunsberger S, Zhao Y, Simon R. A Comparison of Phase II Study Strategies. *Clin Cancer Res* 2009;15(19):5950-5955.

Skipped Phase II (Phase III with futility analysis)

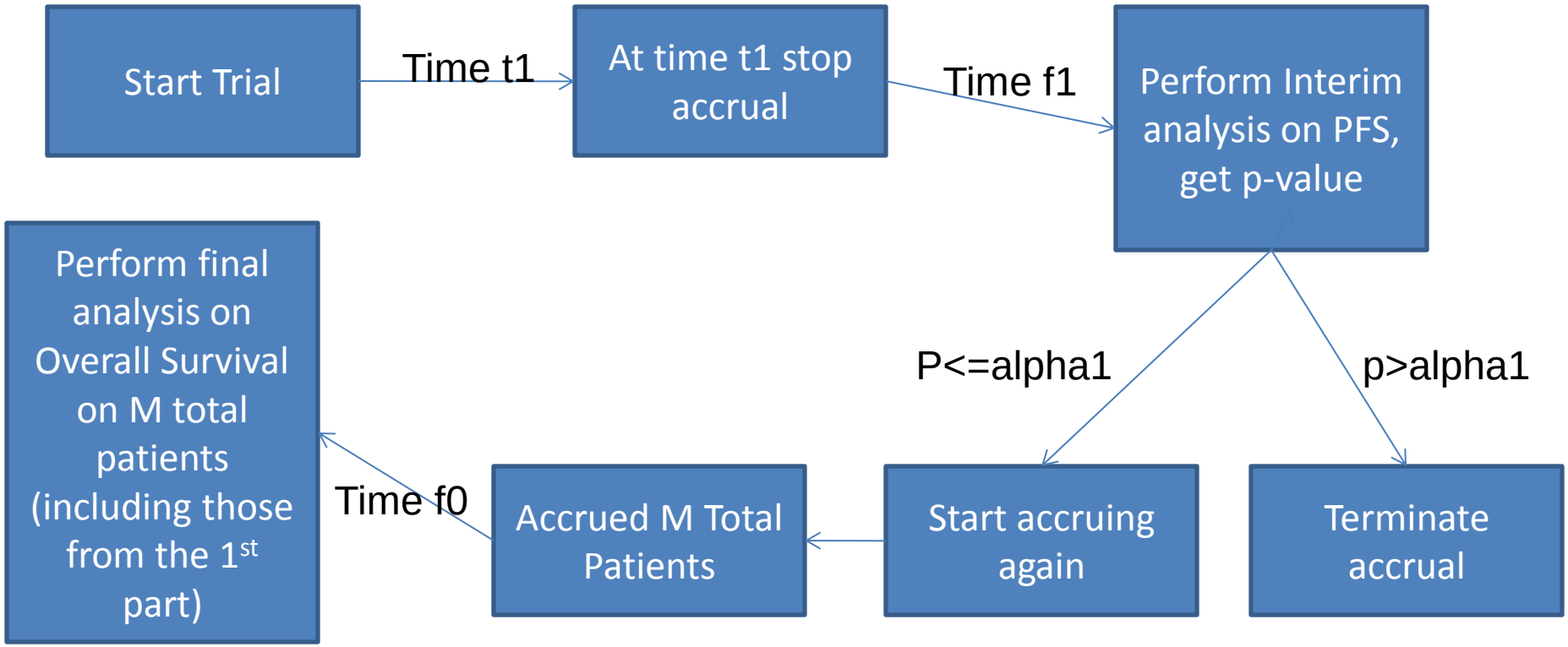
Must specify t1 for time of interim analysis and threshold to stop trial at interim (alpha1)



Determine t1 and alpha1 so that 81% overall power (randomized phase II with 90% for PFS followed by randomized phase III study with 90% power for OS)

Integrated Phase II/III Design

Must specify t1 for to stop initial accrual, f1 (minimum time for first follow-up), threshold to stop trial at interim (alpha1), M patients for total trial, f0 (minimum overall follow-up time)



Determine t1 and alpha1 so that 81% overall power (randomized phase II with 90% for PFS followed by randomized phase III study with 90% power for OS)

Simulation Results

E[N] is expected total sample size
E[T] is expected total study time (in months)

Global Null: No treatment effect on PFS or OS
Global Alternative: Treatment effect on both PFS and OS

Accrual per month	Hazard Ratios		Medians		Designs		Global Null		Global Alternative	
	Y1	OS	Y1	OS		α_1	E[N]	E[T]	E[N]	E[T]
30	1.5	1.3	3	6	Futility based on OS	0.2	513	17.1	719	24.0
						0.5	504	16.8	699	23.3
					Sequence of studies	0.1	325	14.44	935	39.6
					Integrated II/III ($f_1 = 0$)	0.1	350	11.7	713	23.8
						0.2	**336**	**11.2**	**704**	**23.5**
15	1.5	1.3	3	12	Futility based on OS	0.2	524	34.9	725	48.3
						0.5	500	33.3	670	46.6
					Sequence of studies	0.1	297	24.0	906	74.2
					Integrated II/III ($f_1 = 0$)	0.1	**292**	**19.4**	708	47.2
						0.2	302	20.2	705	47.0
15	1.3	1.2	3	6	Futility based on OS	0.2	765	51.0	1253	83.6
						0.5	819	54.6	1210	80.7
					Sequence of studies	0.1	629	45.4	1702	121.9
					Integrated II/III ($f_1 = 0$)	0.1	586	39.1	1249	83.3
						0.2	**547**	**36.5**	1227	81.8

Software

- There is correlation between PFS and OS in the integrated II/III design because we are using the same people in both parts
  - Hard to calculate by hand!

- But there is a great website to help you with this!

  http://linus.nci.nih.gov/brb/samplesize/ip23study1.html

General Concepts

- Overall Type I error control. Determine alpha spent and stopping boundaries (α1, β1, α2) at each stage of testing.
- Two-stage Designs $T_k = p_k \quad k = 1, 2 \quad (MIP)$
- Test statistics at kth stage:

Overall type − I error control: $\alpha = \alpha_1 + \alpha_2(\beta_1 - \alpha_1)$

MIP

$T_k = \prod_{i=1}^{k} p_i \quad k = 1, 2 \ (MPP)$ where p_i = stagewise p − value

Overall type − I error control: $\alpha = \alpha_1 + \alpha_2 \ln \frac{\beta_1}{\alpha_1}$

(MPP)

General Concepts

- Stopping Boundary:

$$\begin{cases} \text{Stop for efficacy if} & T_k \le \alpha_k \\ \text{Stop for futility if} & T_k > \beta_k \\ \text{Continue with adaptations if} & \alpha_k < T_k \le \beta_k \end{cases}$$

Diagram for Seamless Design with Multiple Arms

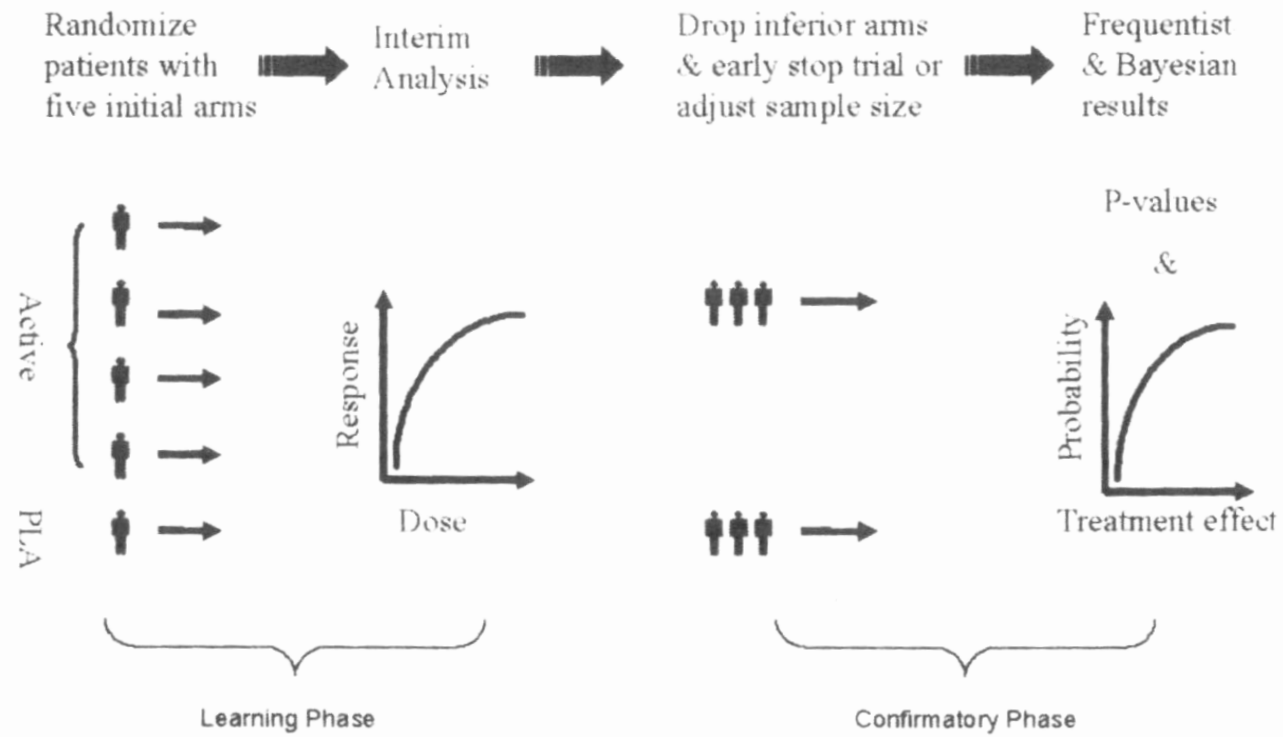


Figure 11.1: Seamless Design

Design 1: Strong control of alpha – allows comparison of which specific treatment is effective

Steps:

1. Stage 1: H_{0i}, i-1,2,…,m_1 null hypotheses for m_1 comparisons across M treatment groups.
2. $p1_i$, i=1,2,…,m_1 are raw p-values. Bonferroni adjusted p-values = $adj(p1_i)=m_1 p1_i$
3. Decision at stage 1:

$$\begin{cases} adj\ \ p1_i < \alpha_1 & \text{Reject } H_{0i} \text{ (efficacy)} \\ adj\ \ p1_i > \beta_1 & \text{Accept } H_{0i} \text{ (futility)} \\ \text{Otherwise} & \text{Continue to Stage 2} \end{cases}$$

4. Stage 2: Choose m2 comparisons based on stage 1. Do step 2 above.
5. Decision at stage 2: if adj(p1) + adj(p2) < α_2, reject null, otherwise don't reject H_0.

Summary

- Think about why/whether a multi-arm trial is needed
  - Very useful when there is lack of historical data for comparison
  - Phase II randomized is NOT a short-cut to avoid a larger more definitive trial
  - Adaptive designs can be very efficient for selection, but require more maintenance

References (1)

- Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012 Mar 28;483(7391):531-3.
- Berry DA, Herbst RS, Rubin EH. Reports from the 2010 Clinical and Translational Cancer Research Think Tank meeting: design strategies for personalized therapy trials. Clin Cancer Res. 2012 Feb 1;18(3):638-44.
- Berry DA, Eick SG. Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. Stat Med. 1995 Feb 15;14(3):231-46.
- Blume, JD. Likelihood Methods for Measuring Statistical Evidence, Stat Med. 2002 (21), 2563-2599.
- Ellenberg SS, Eisenberger MA. An efficient design for phase III studies of combination chemotherapies. Cancer Treat Rep. 1985 Oct;69(10):1147-54.
- Ensign LG, Gehan EA, Kamen DS, Thall PF. , An optimal three-stage design for phase II clinical trials. Stat Med. 1994 Sep 15;13(17):1727-36.
- Gehan EA. The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. J Chronic Dis. 1961 Apr;13:346-53.

References (2)

- Halpern J, Brown BW Jr. Sequential treatment allocation procedures in clinical trials--with particular attention to the analysis of results for the biased coin design. Stat Med. 1986 May-Jun;5(3):211-29.
- Huang B, et al. Application of recent Bayesian interim monitoring strategies in Phase II screening trials in breast cancer. SCT conference, 2010.
- Hunsberger S, Zhao Y, Simon R. A Comparison of Phase II Study Strategies. *Clin Cancer Res* 2009;15(19):5950-5955.
- Schaid DJ, Ingle JN, Wieand S, Ahmann DL.A design for phase II testing of anticancer agents within a phase III clinical trial. Control Clin Trials. 1988 Jun;9(2):107-18.
- Scher HI, Heller G. Picking the winners in a sea of plenty. Clin Cancer Res. 2002 Feb;8(2):400-4.
- Simon R, Wittes RE, Ellenberg SS. Randomized phase II clinical trials. Cancer Treat Rep. 1985 Dec;69(12):1375-81.
- Simon R. Optimal two-stage designs for phase II clinical trials. Control Clin Trials. 1989 Mar;10(1):1-10.
- Storer BE. A sequential phase II/III trial for binary outcomes. Stat Med. 1990 Mar;9(3):229-35.
- Thall PF, Cheng SC. Treatment comparisons based on two-dimensional safety and efficacy alternatives in oncology trials. Biometrics. 1999 Sep;55(3):746-53.
- Ye F, Shyr Y. Balanced two-stage designs for phase II clinical trials.Clin Trials. 2007;4(5):514-24.