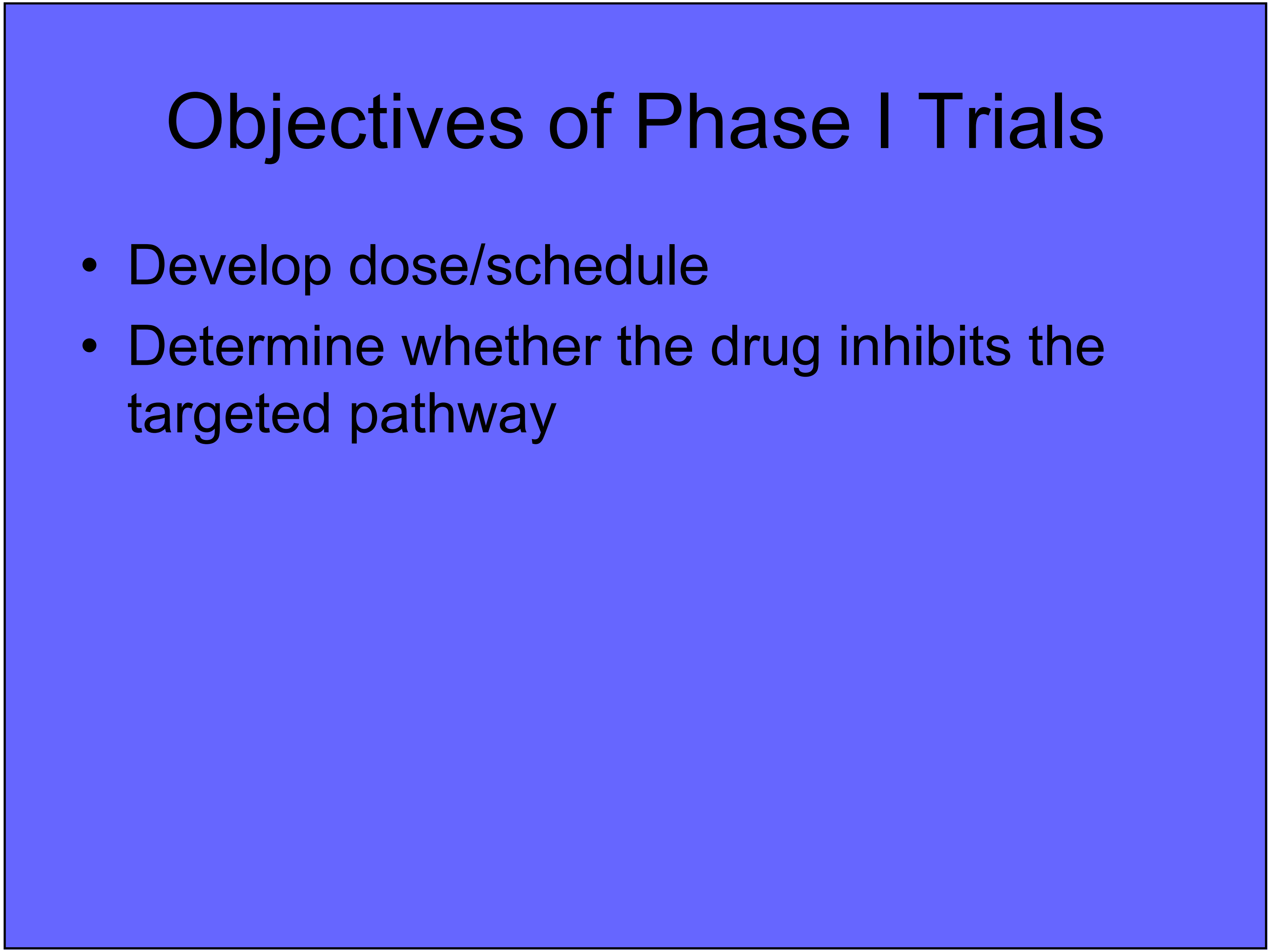
Issues in the Design of Phase I & II Trials for Molecularly Targeted Drugs

Richard Simon, D.Sc.

National Cancer Institute

http://brb.nci.nih.gov

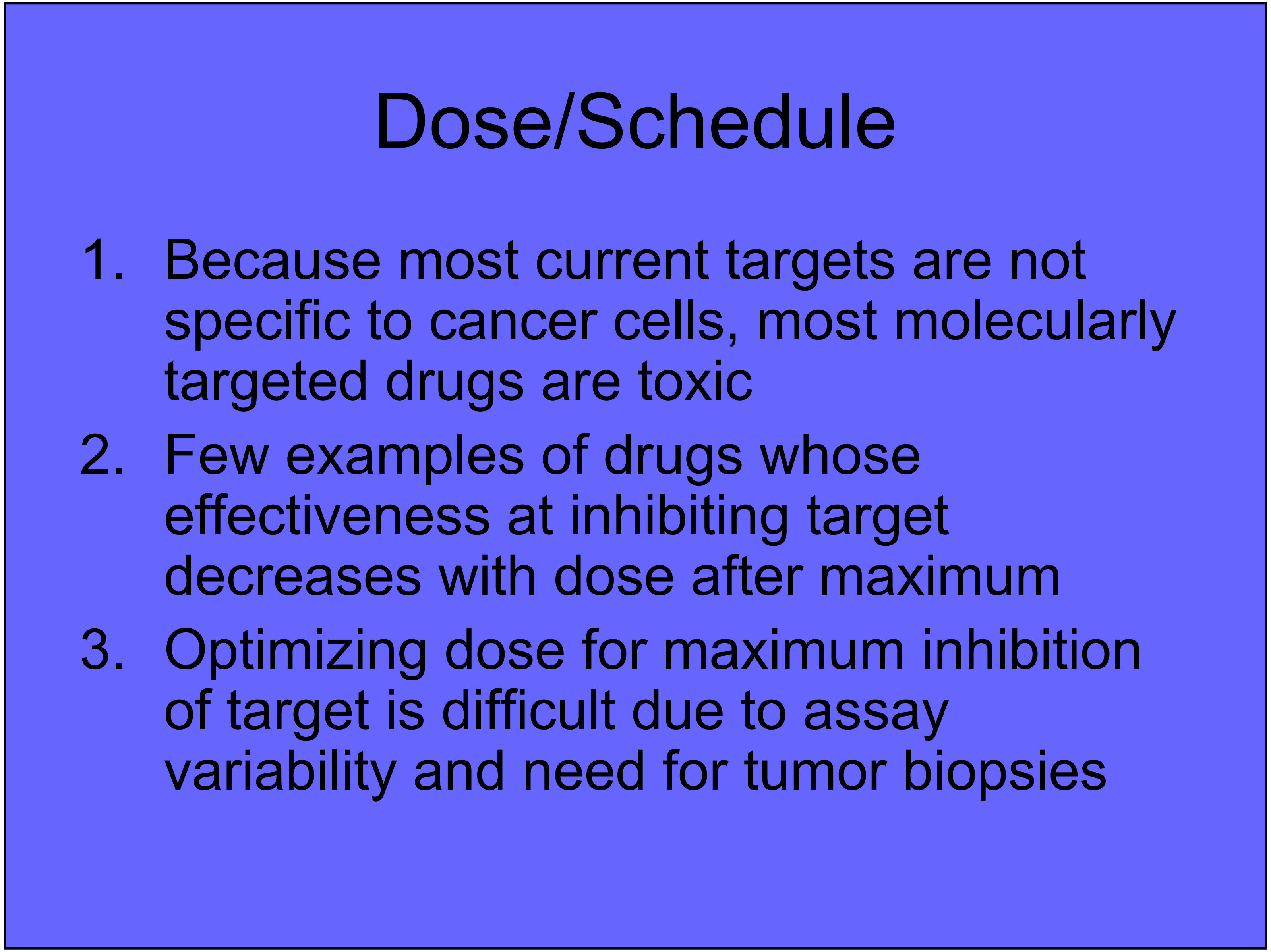


Objectives of Phase I Trials

- Develop dose/schedule
- Determine whether the drug inhibits the targeted pathway

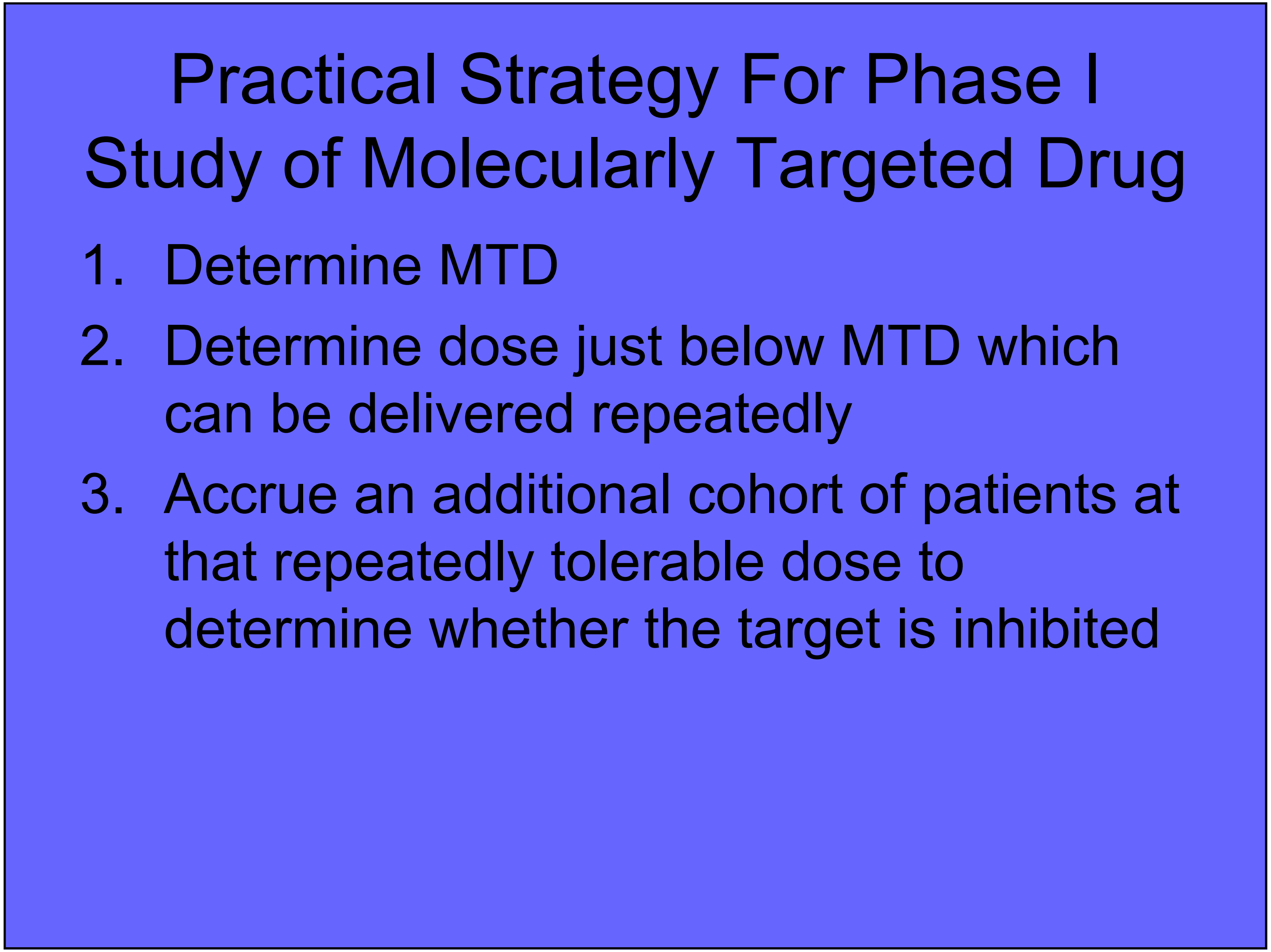
Dose/Schedule

- Ideal is to have a drug and target so specific for cancer cells that the drug can be delivered repeatedly at doses that completely shut down the de-regulated pathway without toxicity to normal cells



Dose/Schedule

1. Because most current targets are not specific to cancer cells, most molecularly targeted drugs are toxic
2. Few examples of drugs whose effectiveness at inhibiting target decreases with dose after maximum
3. Optimizing dose for maximum inhibition of target is difficult due to assay variability and need for tumor biopsies

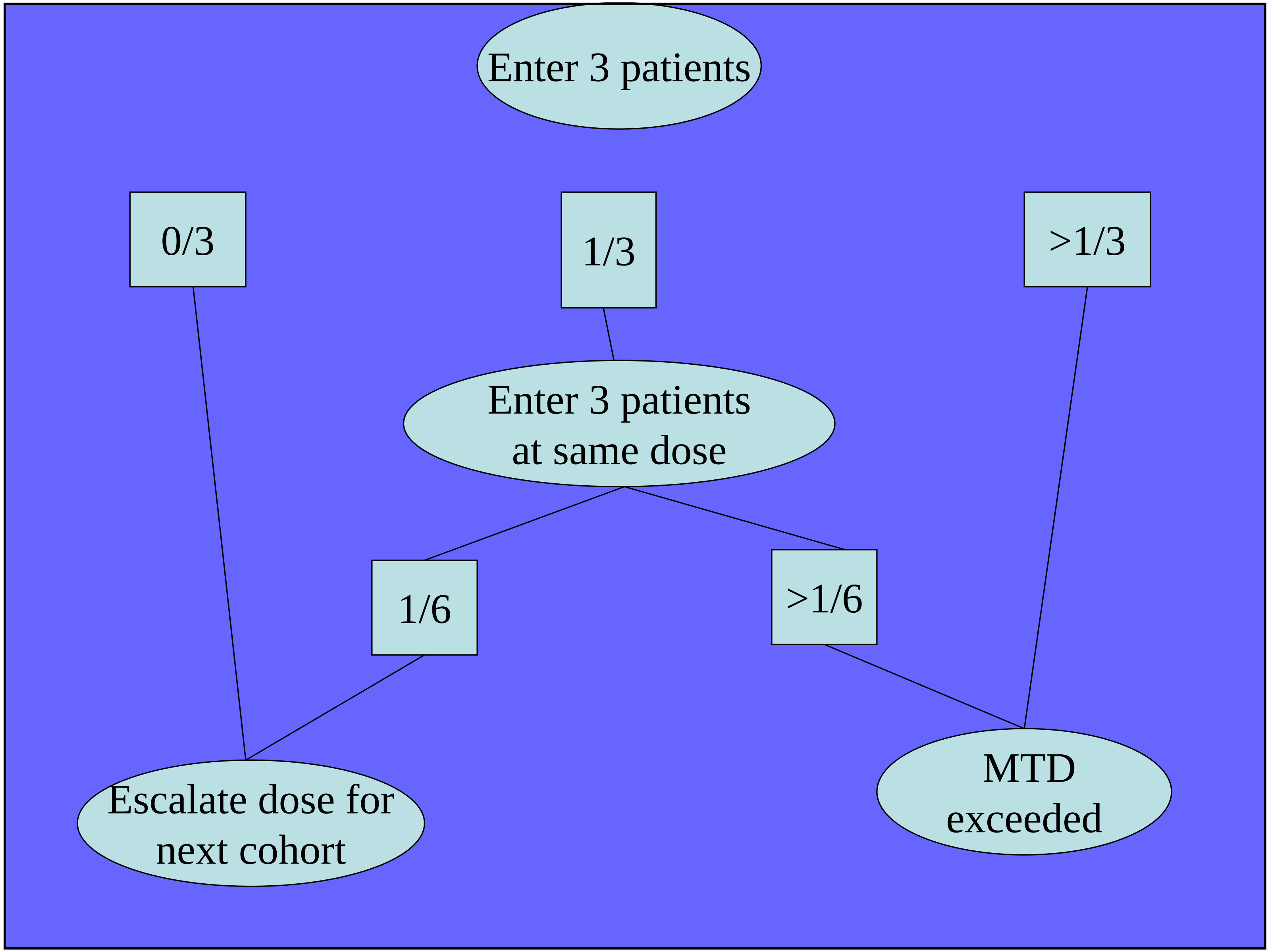


Practical Strategy For Phase I Study of Molecularly Targeted Drug

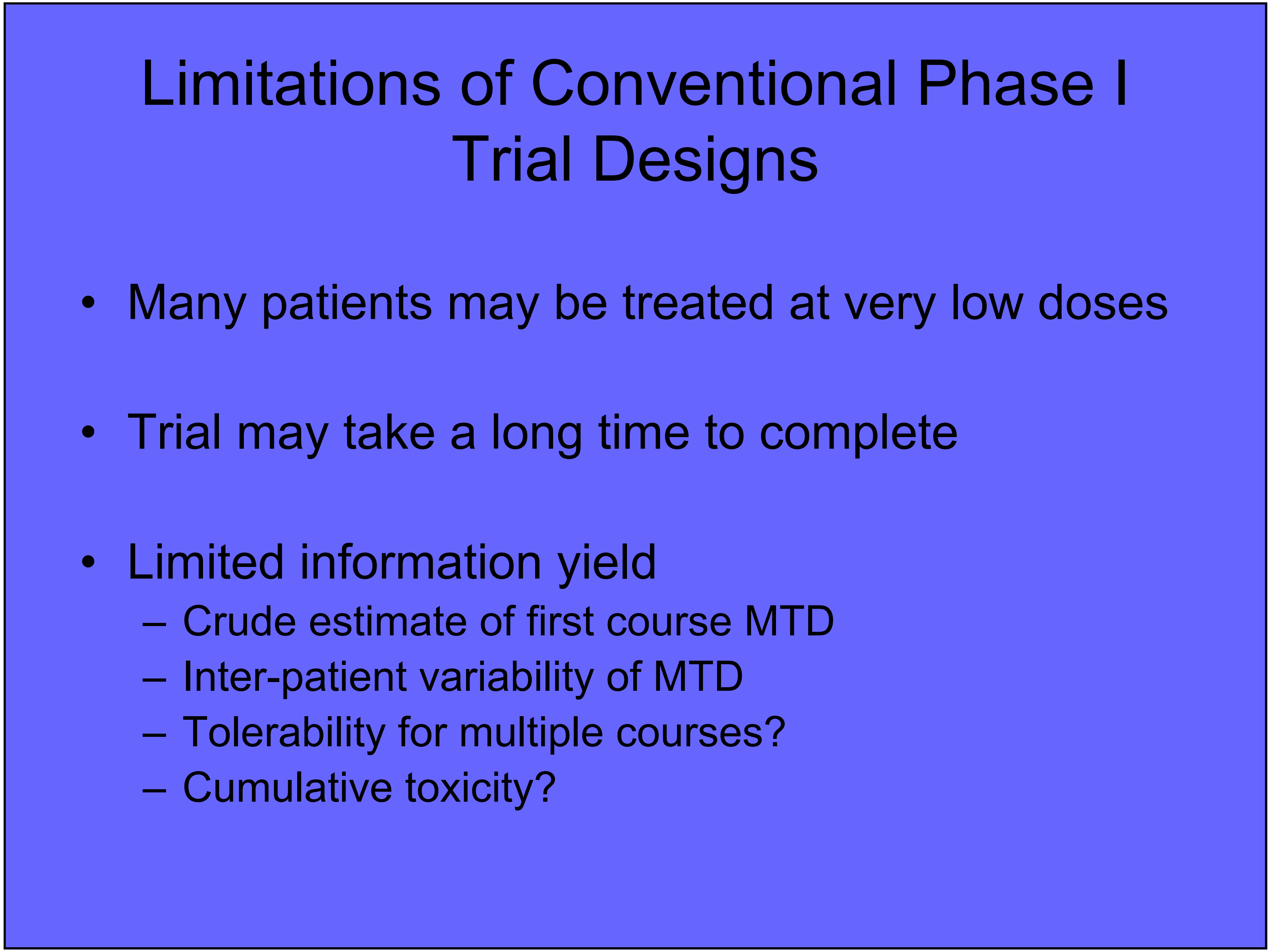
1. Determine MTD
2. Determine dose just below MTD which can be delivered repeatedly
3. Accrue an additional cohort of patients at that repeatedly tolerable dose to determine whether the target is inhibited

Conventional Phase I Designs

- Starting dose 1/10th LD_{10} in most sensitive species
- Modified Fibonacci dose steps
  - 100%, 67%, 50%, 40%, 33%, 33%, …
- Cohorts of 3-6 new patients per dose level
- Define MTD as highest dose with <33% DLT
- Use first course information only
- Use DLT vs non-DLT information
- No intra-patient dose escalation



Enter 3 patients
0/3
1/3
>1/3
Enter 3 patients
at same dose
1/6
>1/6
Escalate dose for
next cohort
MTD
exceeded



Limitations of Conventional Phase I Trial Designs

- Many patients may be treated at very low doses
- Trial may take a long time to complete
- Limited information yield
  - Crude estimate of first course MTD
  - Inter-patient variability of MTD
  - Tolerability for multiple courses?
  - Cumulative toxicity?

Accelerated titration designs for phase I clinical trials in oncology

R Simon, B Freidlin L Rubinstein et al.

JNCI 89:1138-47, 1997.

Cohort Escalation Options

1	Cohorts of 3 new patients per dose level with 40% dose increments. If 1 of 3 experience DLT in first course, expand to cohort of 6
2	Cohorts of 1 new patient per dose level. When first instance of first course DLT or second instance of first course grade 2 toxicity is observed, revert to design 1.
3	Same as design 2 except that double dose steps are used during accelerated stage.
4	Cohorts of 1 new patient per dose level and double dose steps. When first instance of *any course* DLT or second instance of *any course* grade 2 toxicity is observed, revert to design 1.

Within Patient Escalation Options

A	No within patient dose escalation
B	Escalate if grade 0-1 toxicity at previous course. De-escalate if grade 3+ toxicity at previous course. Do not assign dose at which 2 previous pts have experienced 3+ toxicity at that course or earlier.

Testing the 8 Designs

We fit the model to 20 phase I trials, relating to:

Flavone acetic acid (5)
Piroxantrone (2)
Chloroquinoxaline sulfonamide (2)
Pyrazoloacridine (1)
Cyclopentenylcytosine (1)
Fostriecin (2)
9-Aminocamptothecin (2)
Penclomedine (2)

For each trial, we performed 1000 simulations for each of the 8 designs, using the model.

We compiled the results to compare the performances of the 8 designs.

Model Relating Toxicity to Dose

$$Y_{ij} = \log\left(d_{ij} + \alpha D_{ij}\right) + \beta_i + \varepsilon_{ij}$$

d_{ij}=dose for i'th patient in course j

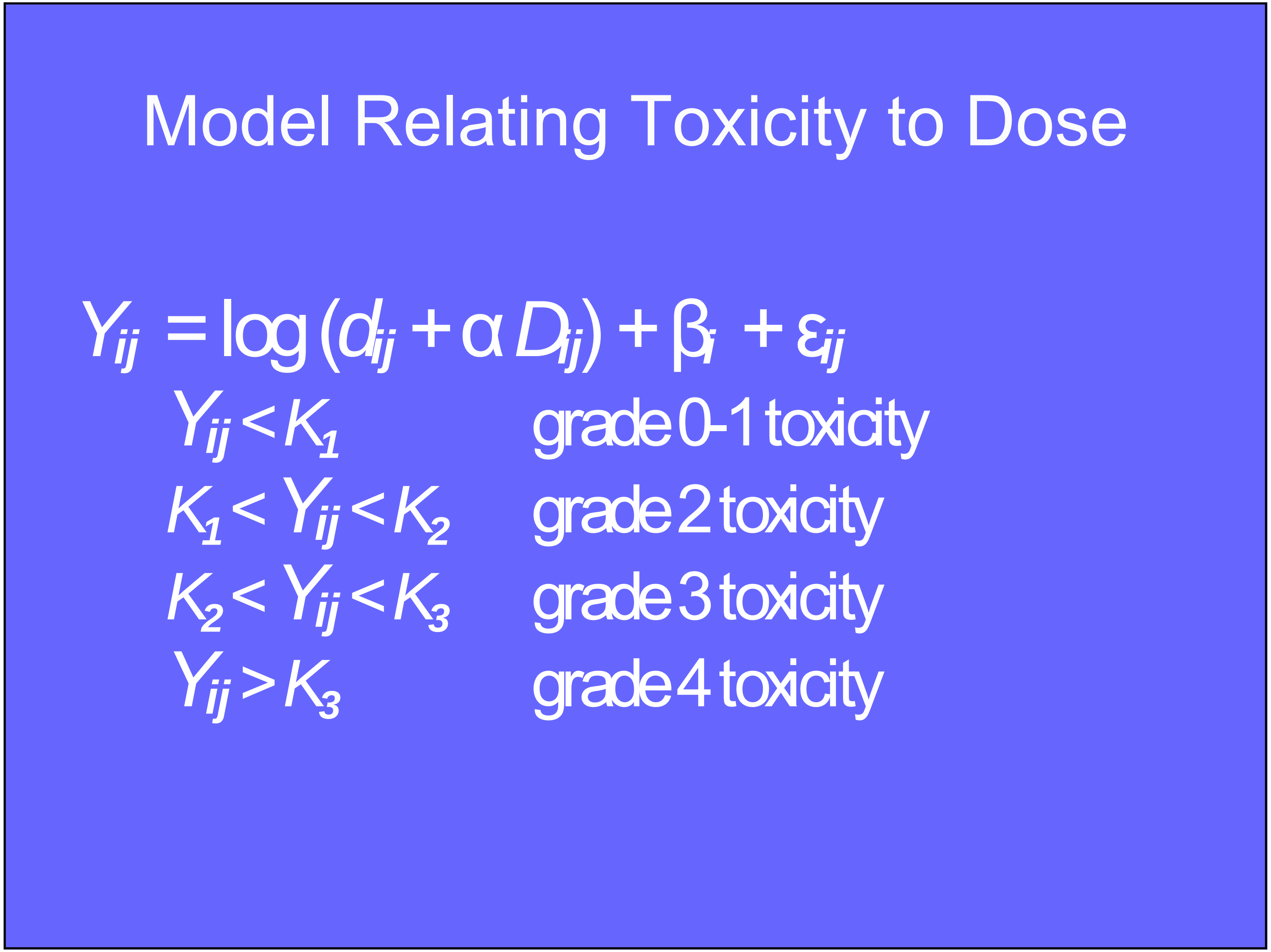
D_{ij}=cumulative dose up to course j for patient i

α=cumulative toxicity parameter

β_i=patient specific effect

$\sim N(0,\sigma_\beta^2)$

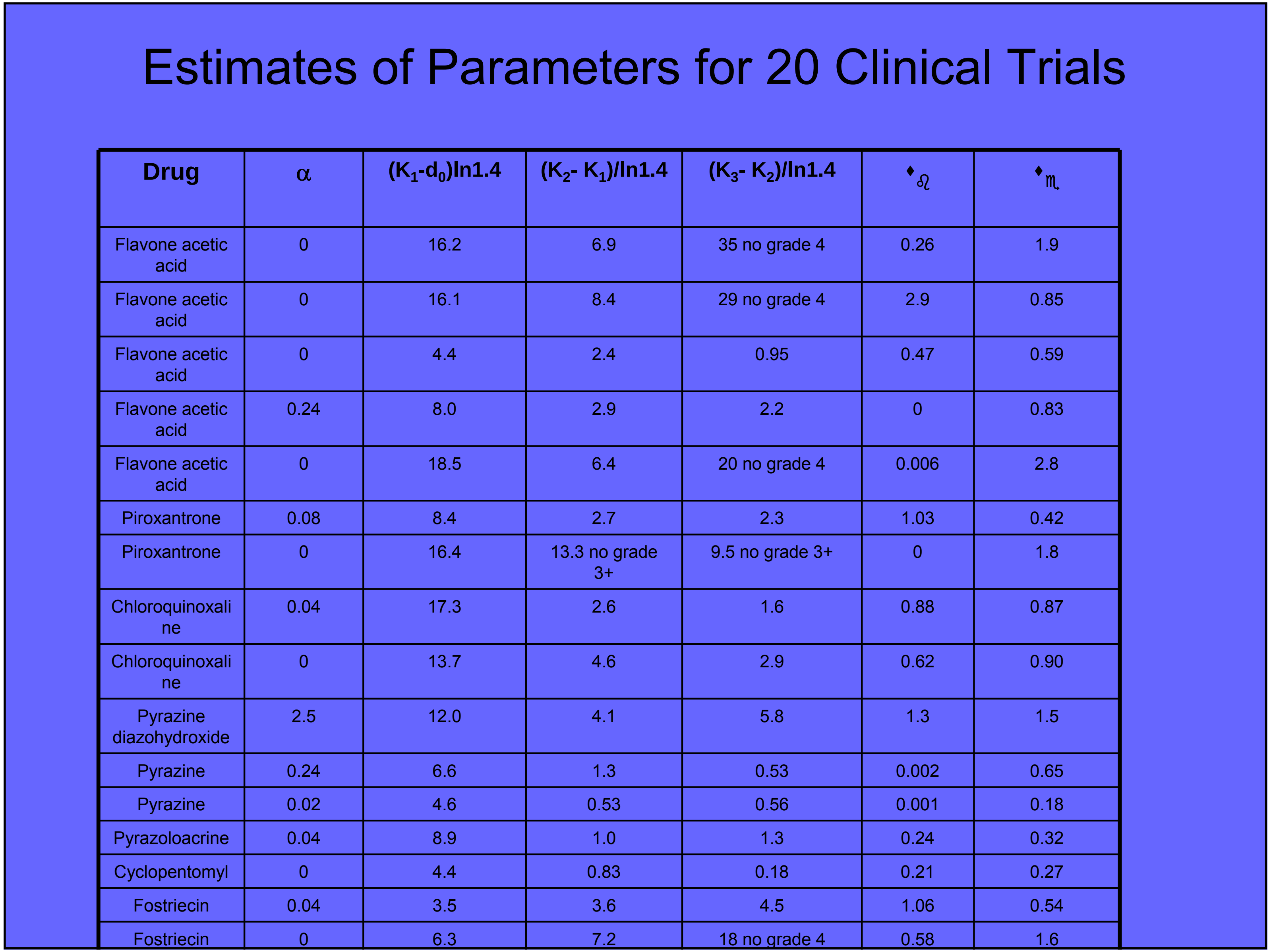
ε_{ij}=course specific random variation



Model Relating Toxicity to Dose

$$Y_{ij} = \log(d_{ij} + \alpha D_{ij}) + \beta_i + \varepsilon_{ij}$$

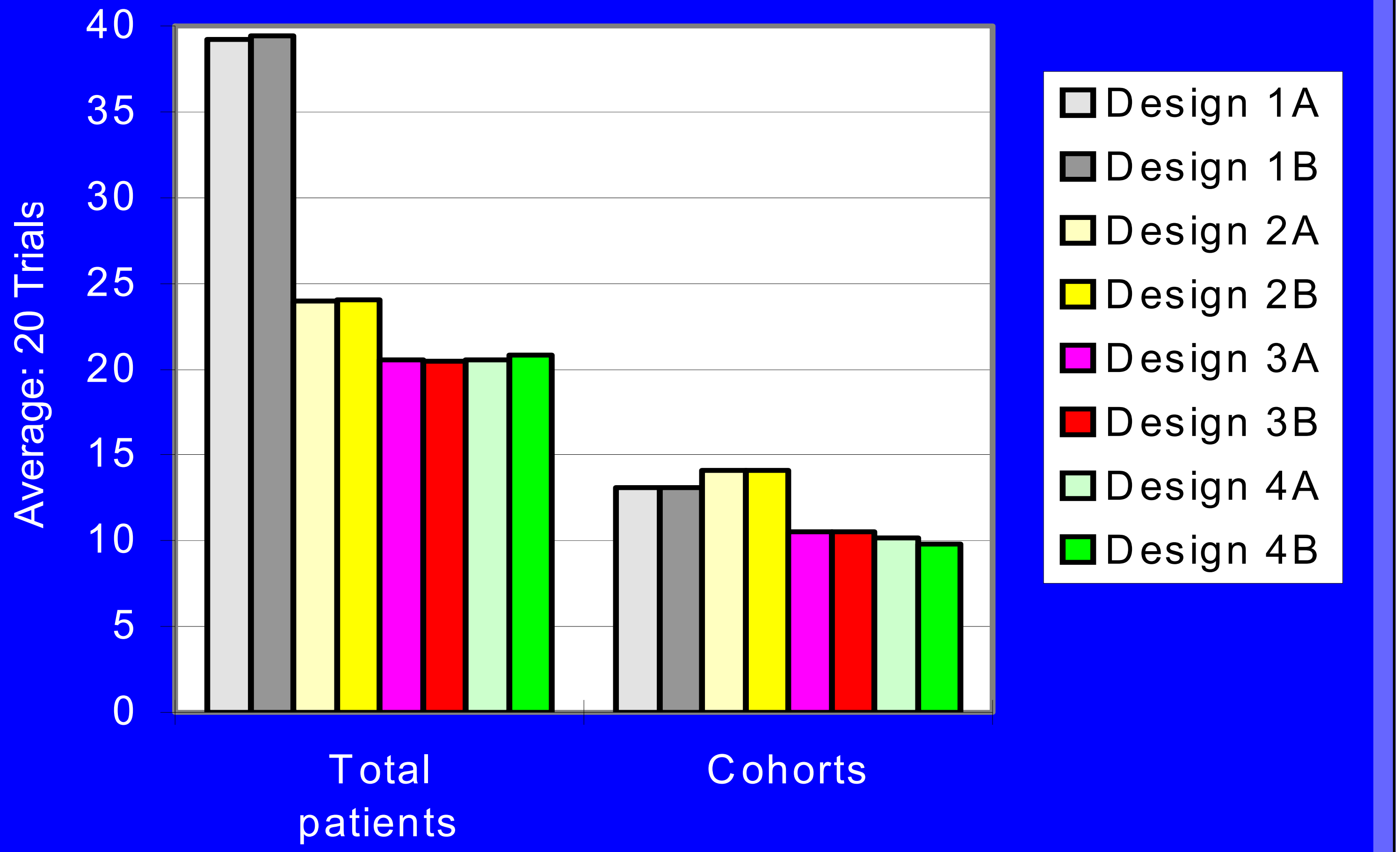
$Y_{ij} < K_1$	grade 0-1 toxicity
$K_1 < Y_{ij} < K_2$	grade 2 toxicity
$K_2 < Y_{ij} < K_3$	grade 3 toxicity
$Y_{ij} > K_3$	grade 4 toxicity



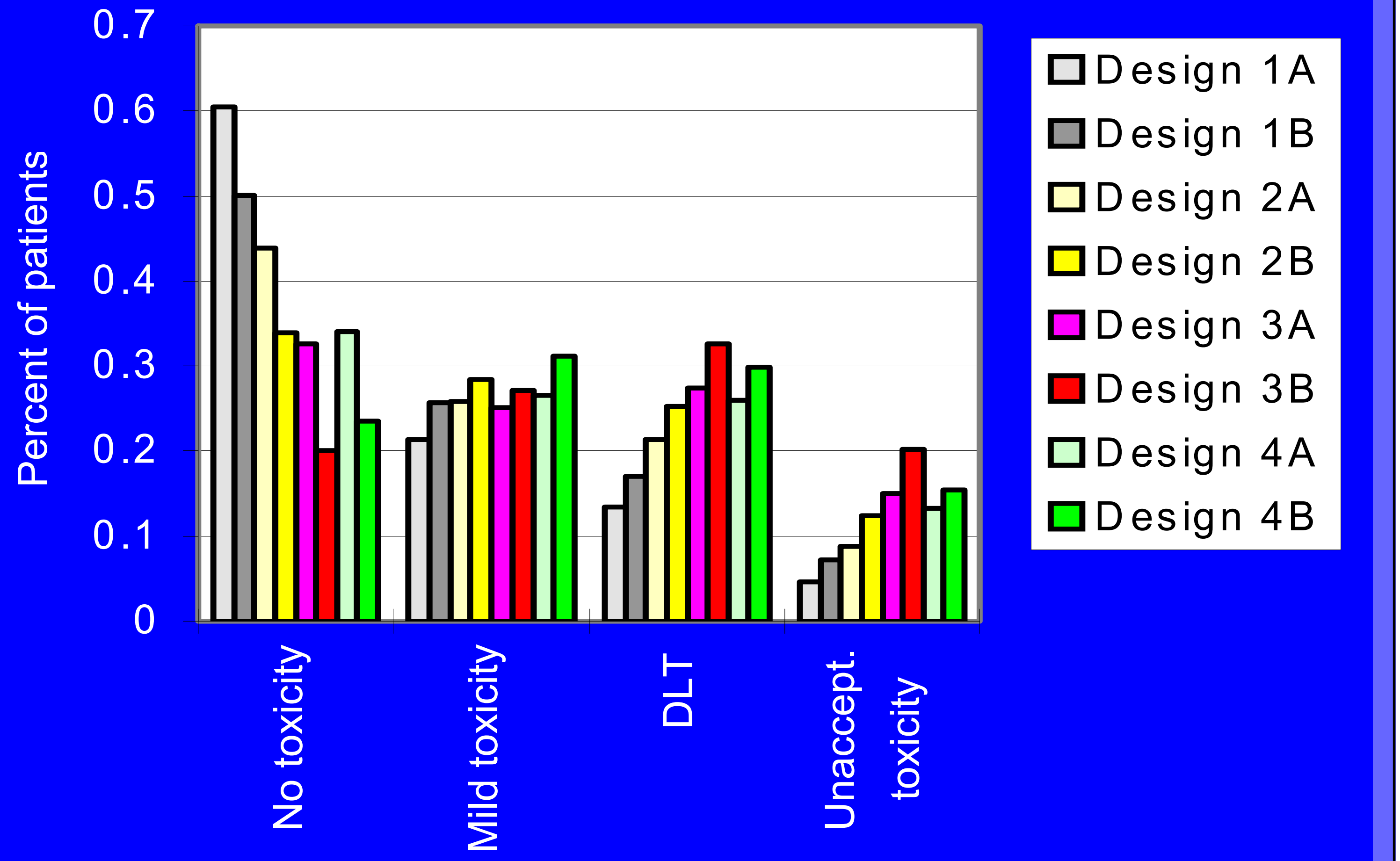
Estimates of Parameters for 20 Clinical Trials

Drug	α	$(K_1\text{-}d_0)$ln1.4	$(K_2\text{- }K_1)$/ln1.4	$(K_3\text{- }K_2)$/ln1.4	♦♌	♦♏
Flavone acetic acid	0	16.2	6.9	35 no grade 4	0.26	1.9
Flavone acetic acid	0	16.1	8.4	29 no grade 4	2.9	0.85
Flavone acetic acid	0	4.4	2.4	0.95	0.47	0.59
Flavone acetic acid	0.24	8.0	2.9	2.2	0	0.83
Flavone acetic acid	0	18.5	6.4	20 no grade 4	0.006	2.8
Piroxantrone	0.08	8.4	2.7	2.3	1.03	0.42
Piroxantrone	0	16.4	13.3 no grade 3+	9.5 no grade 3+	0	1.8
Chloroquinoxaline	0.04	17.3	2.6	1.6	0.88	0.87
Chloroquinoxaline	0	13.7	4.6	2.9	0.62	0.90
Pyrazine diazohydroxide	2.5	12.0	4.1	5.8	1.3	1.5
Pyrazine	0.24	6.6	1.3	0.53	0.002	0.65
Pyrazine	0.02	4.6	0.53	0.56	0.001	0.18
Pyrazoloacrine	0.04	8.9	1.0	1.3	0.24	0.32
Cyclopentomyl	0	4.4	0.83	0.18	0.21	0.27
Fostriecin	0.04	3.5	3.6	4.5	1.06	0.54
Fostriecin	0	6.3	7.2	18 no grade 4	0.58	1.6

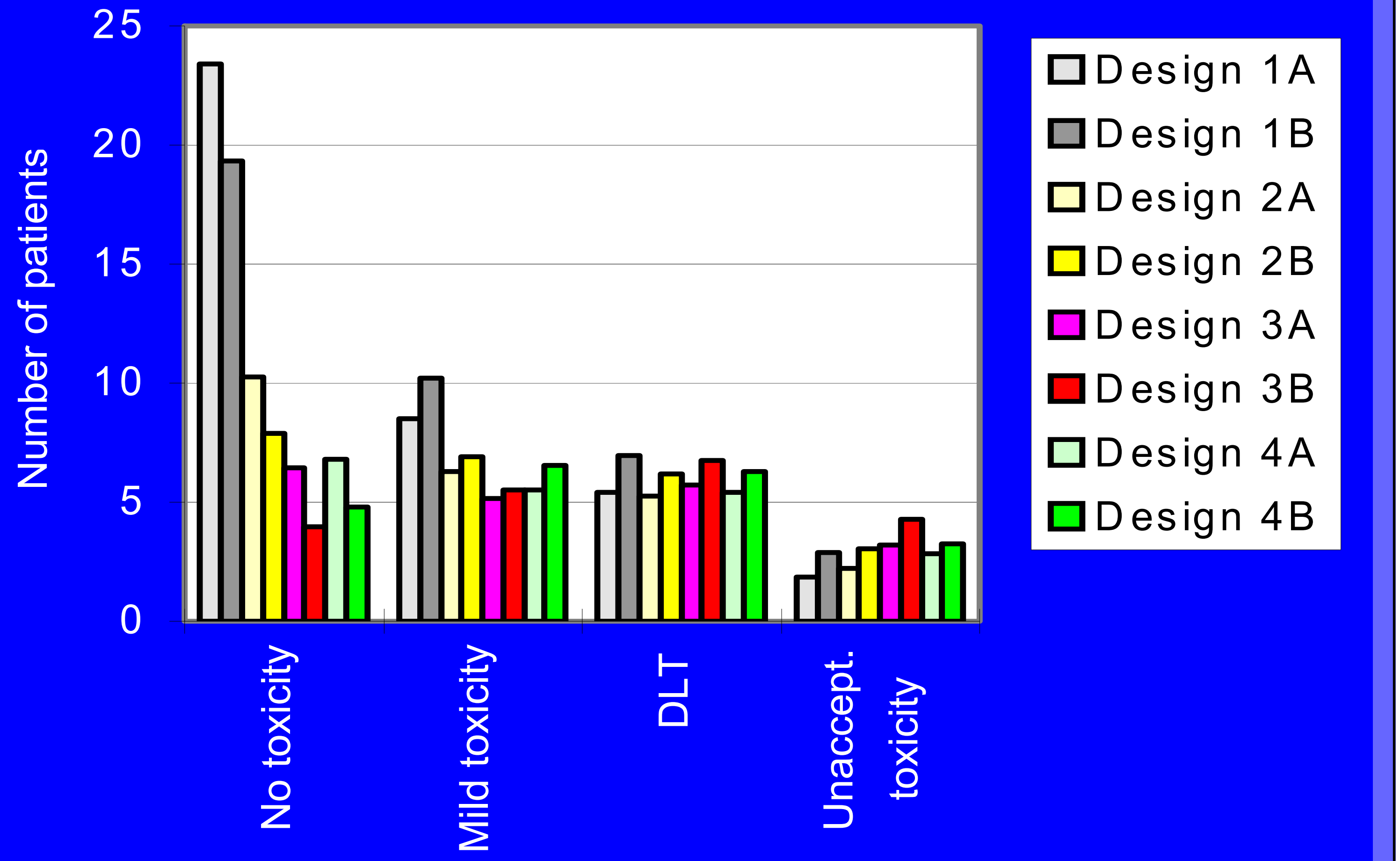
Patients and Cohorts Distribution for 8 Designs

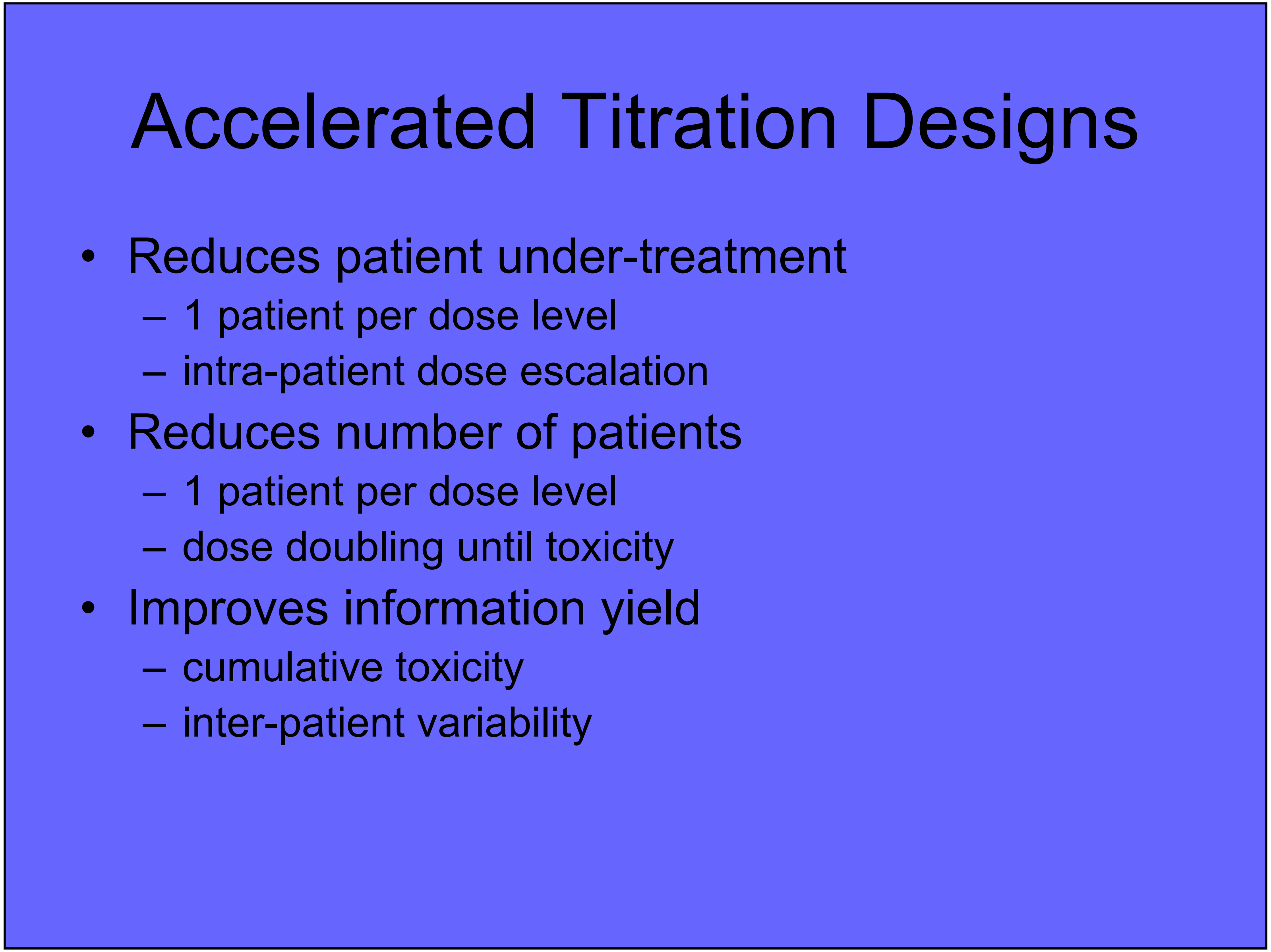


Toxicity Distribution for 8 Designs



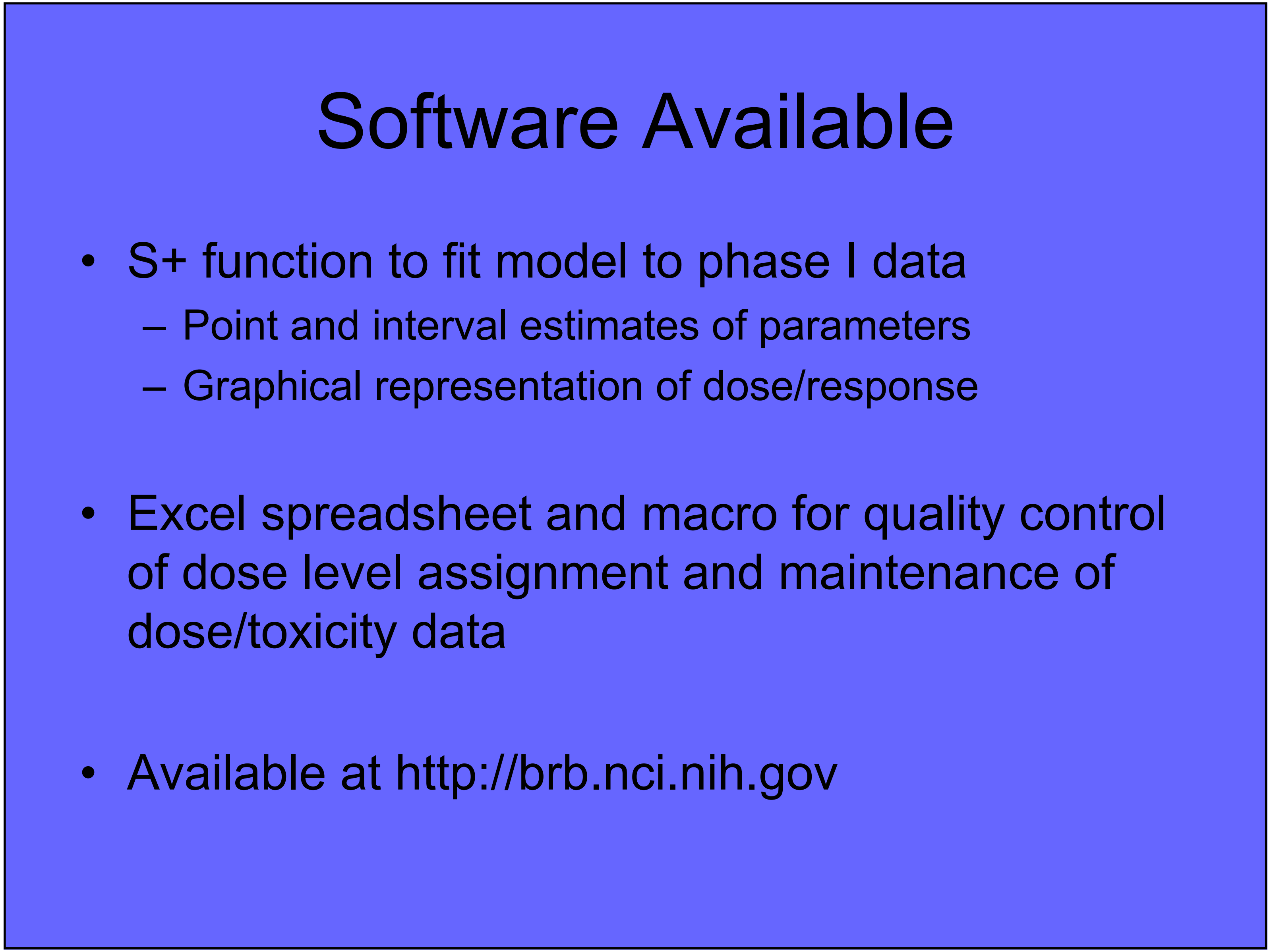
Toxicity Distribution for 8 Designs





Accelerated Titration Designs

- Reduces patient under-treatment
  - 1 patient per dose level
  - intra-patient dose escalation
- Reduces number of patients
  - 1 patient per dose level
  - dose doubling until toxicity
- Improves information yield
  - cumulative toxicity
  - inter-patient variability



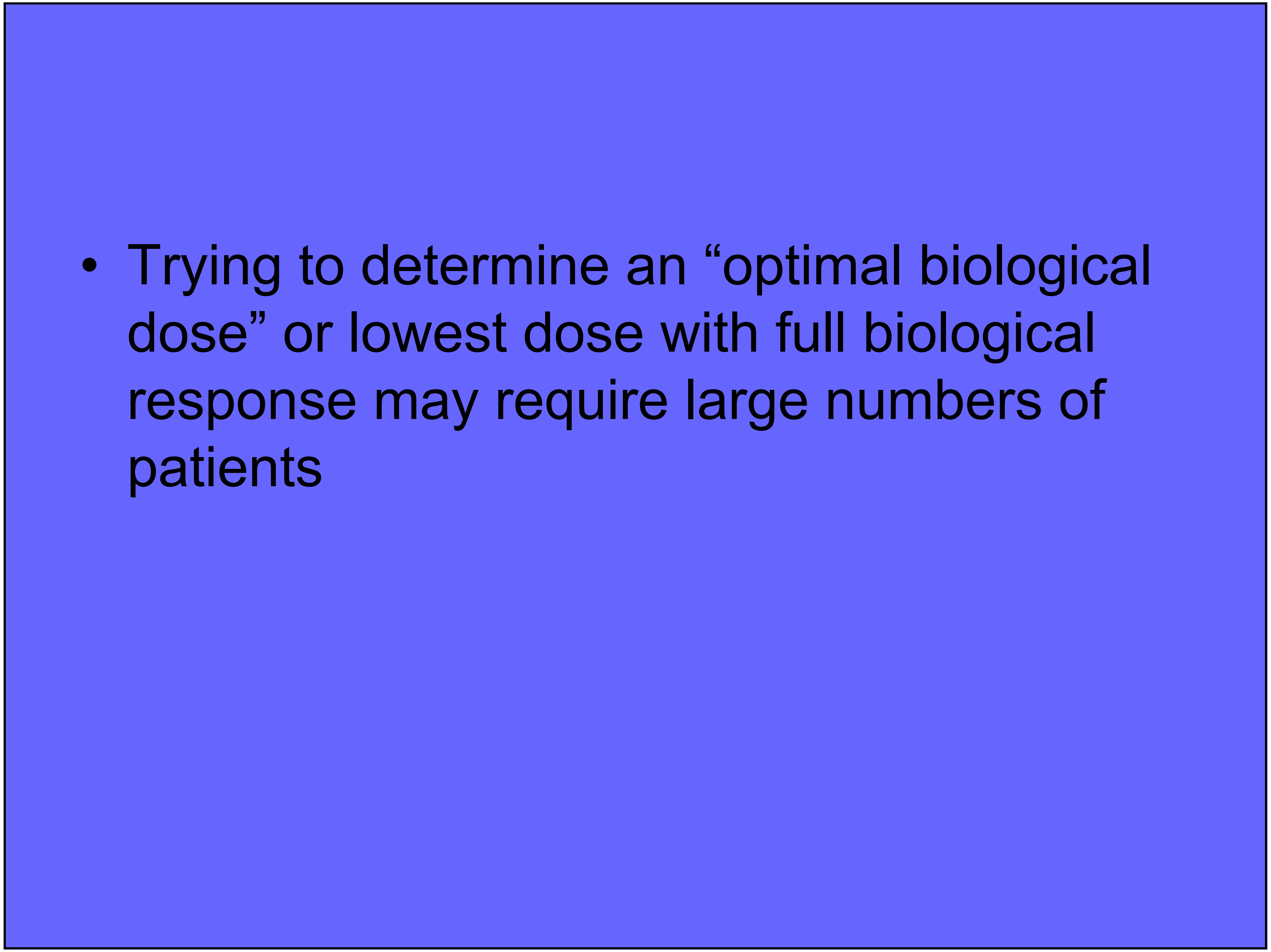
Software Available

- S+ function to fit model to phase I data
  - Point and interval estimates of parameters
  - Graphical representation of dose/response
- Excel spreadsheet and macro for quality control of dose level assignment and maintenance of dose/toxicity data
- Available at http://brb.nci.nih.gov

Korn et al Phase I Design for Finding Biologically Active Dose

Buolamwini & Adjei, Novel Anticancer Drug Protocols, Humana 2003

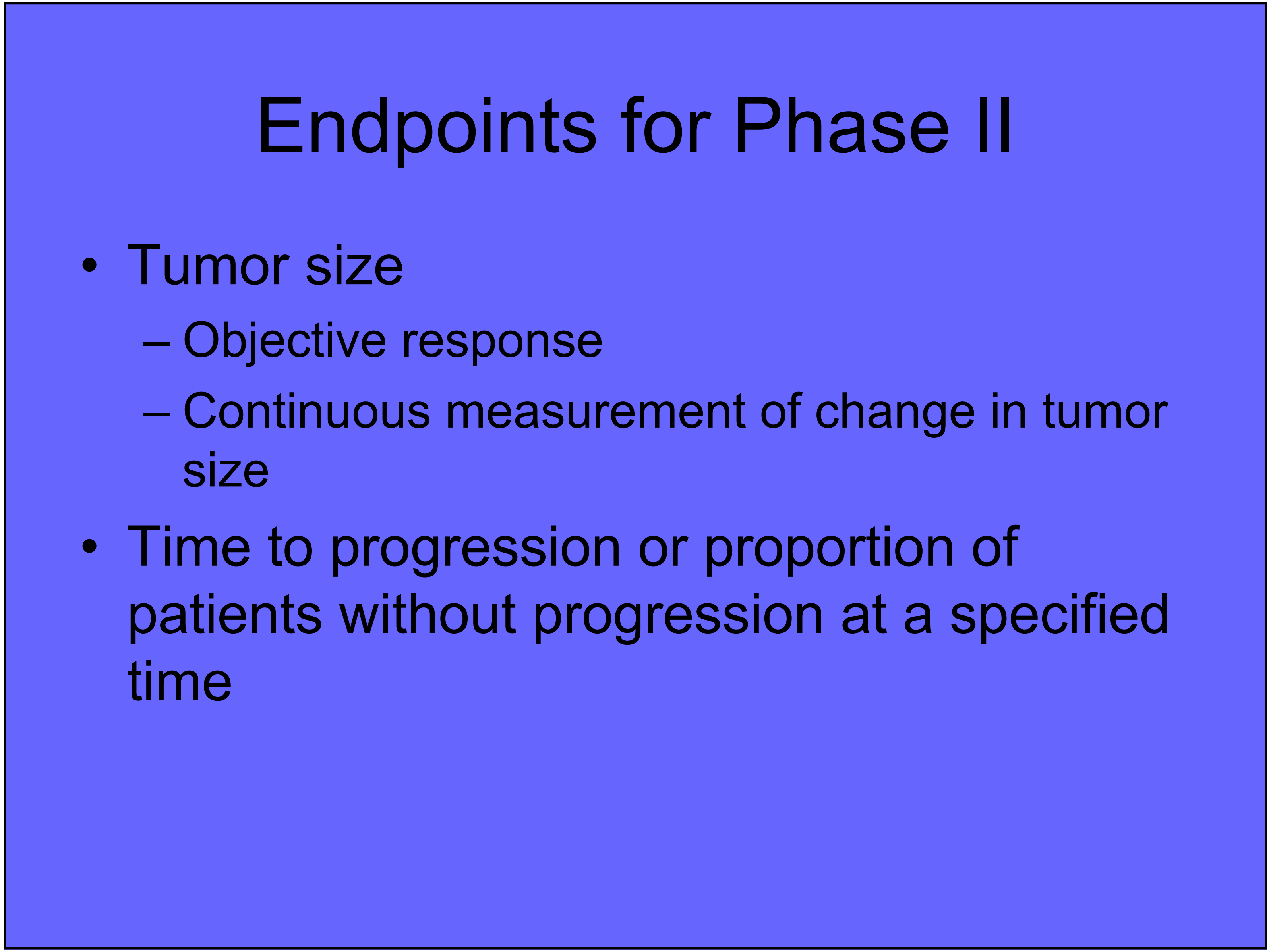
- Treat one patient per dose level until one biological response is seen
- After the first response, treat cohorts of 3-6 patients per dose
  - With 0-1 responses in 3 patients, escalate dose for next cohort
  - With 2-3 responses in 3 patients, expand cohort to 6 patients
  - With 5-6 responses, end
  - With <5 responses, escalate dose for next cohort



- Trying to determine an “optimal biological dose” or lowest dose with full biological response may require large numbers of patients

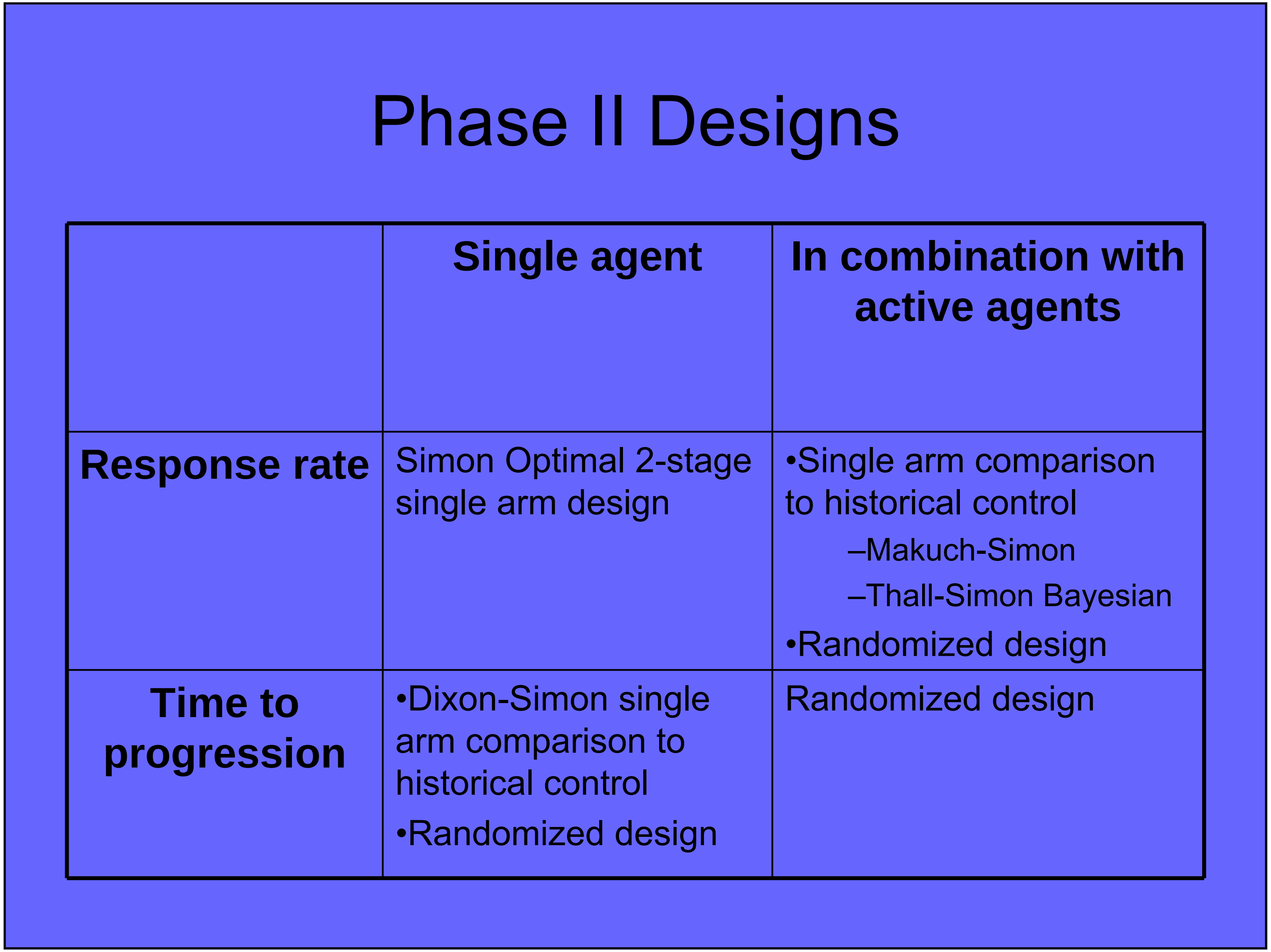
Objectives of Phase II Trials of Targeted Agents

- Determine whether there is a population of patients for whom the drug demonstrates sufficient anti-tumor activity to warrant a phase III trial
- Optimize the regimen in which the drug will be used in the phase III trial
- Optimize the target population for the phase III trial
- Develop predictive biomarker for identifying target population and a robust test for use in the phase III trial



Endpoints for Phase II

- Tumor size
  - – Objective response
  - – Continuous measurement of change in tumor size
- Time to progression or proportion of patients without progression at a specified time



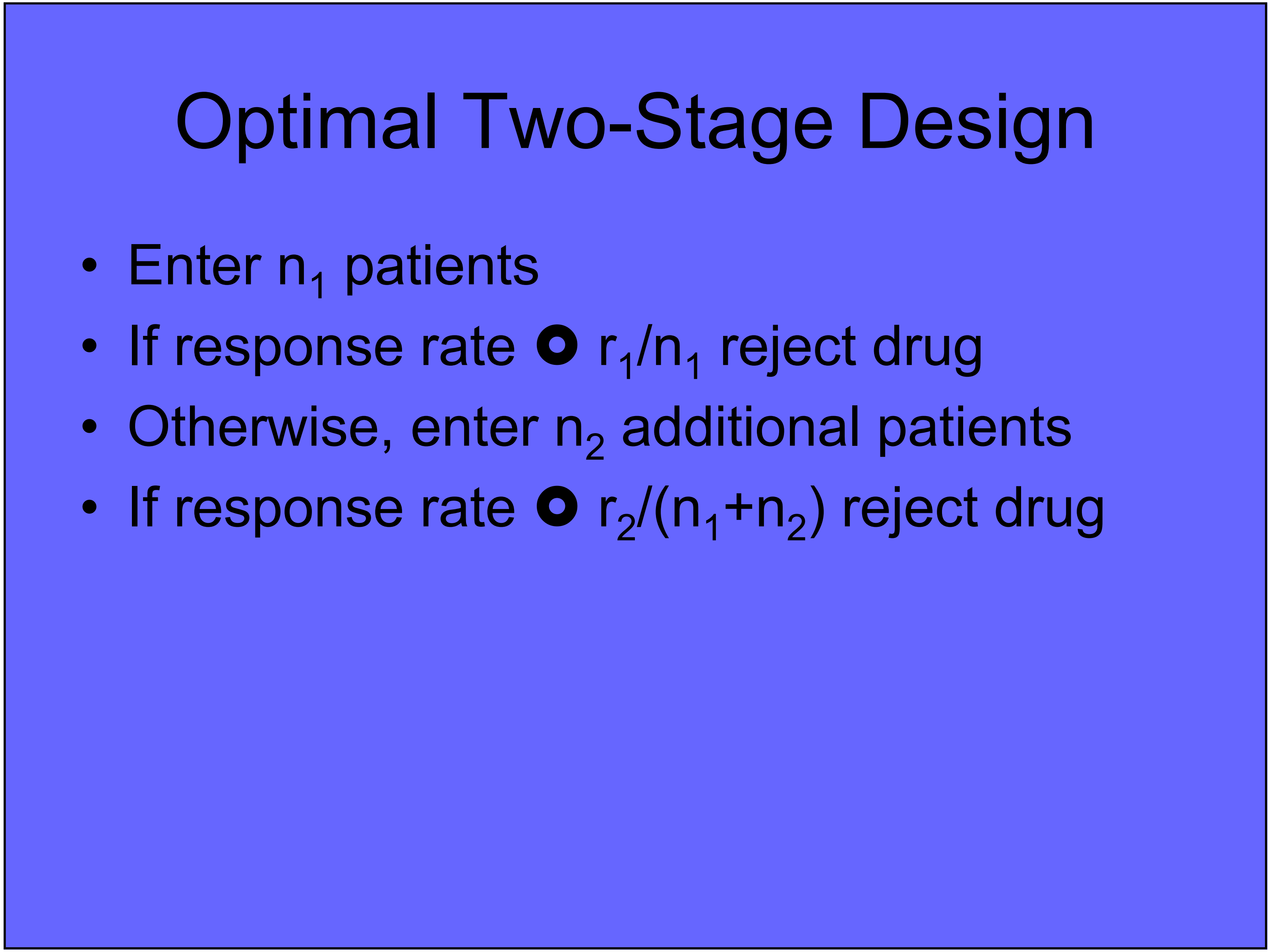
Phase II Designs

	Single agent	In combination with active agents
Response rate	Simon Optimal 2-stage single arm design	•Single arm comparison to historical control –Makuch-Simon –Thall-Simon Bayesian •Randomized design
Time to progression	•Dixon-Simon single arm comparison to historical control •Randomized design	Randomized design

Optimal two-stage Designs

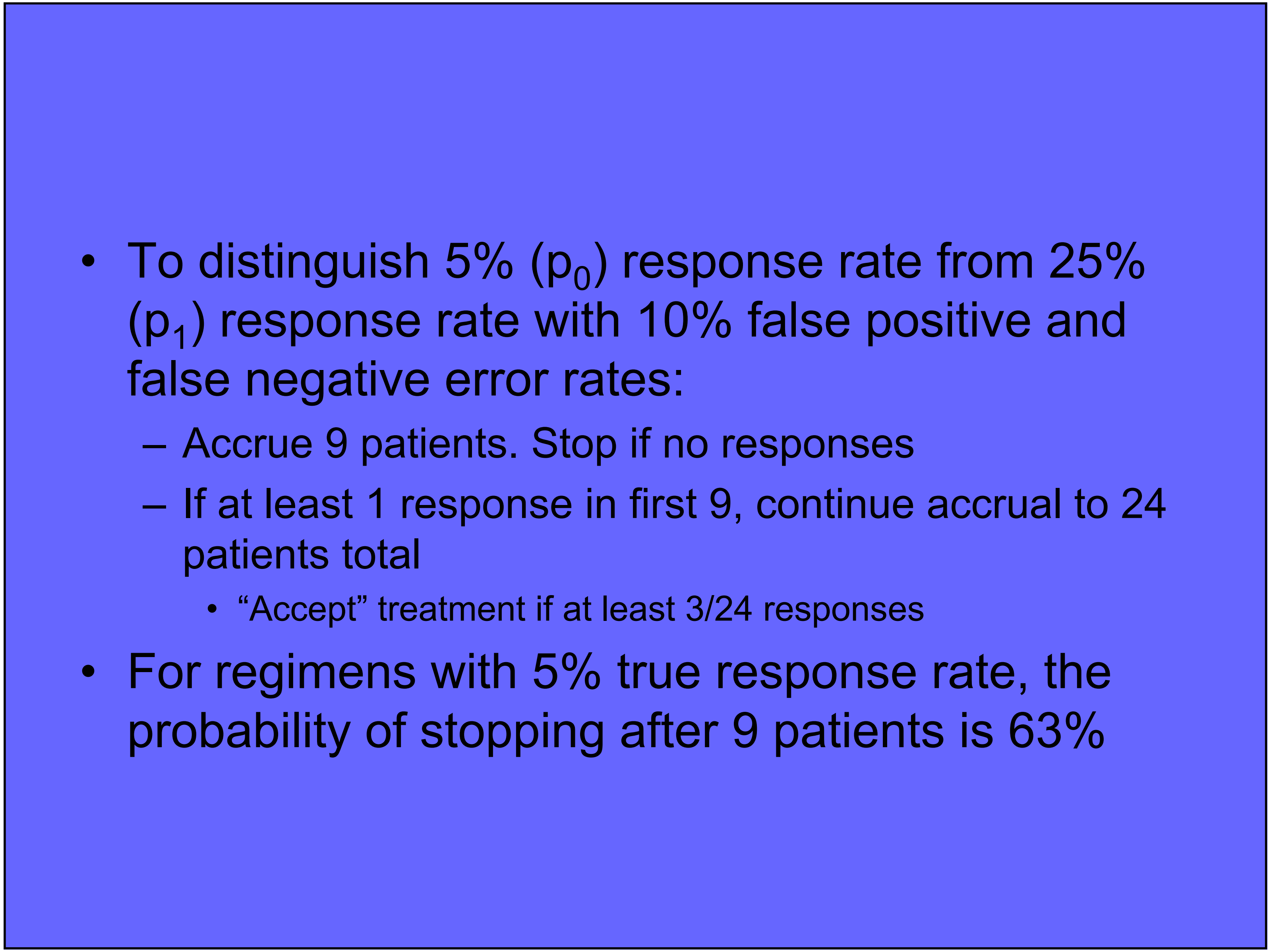
R Simon, Controlled Clinical Trials 10:1-10,1989

- Tests hypothesis $p \leq p_0$ against alternative $p \geq p_1$
- p_0 represents inadequate activity
- p_1 represents activity of interest

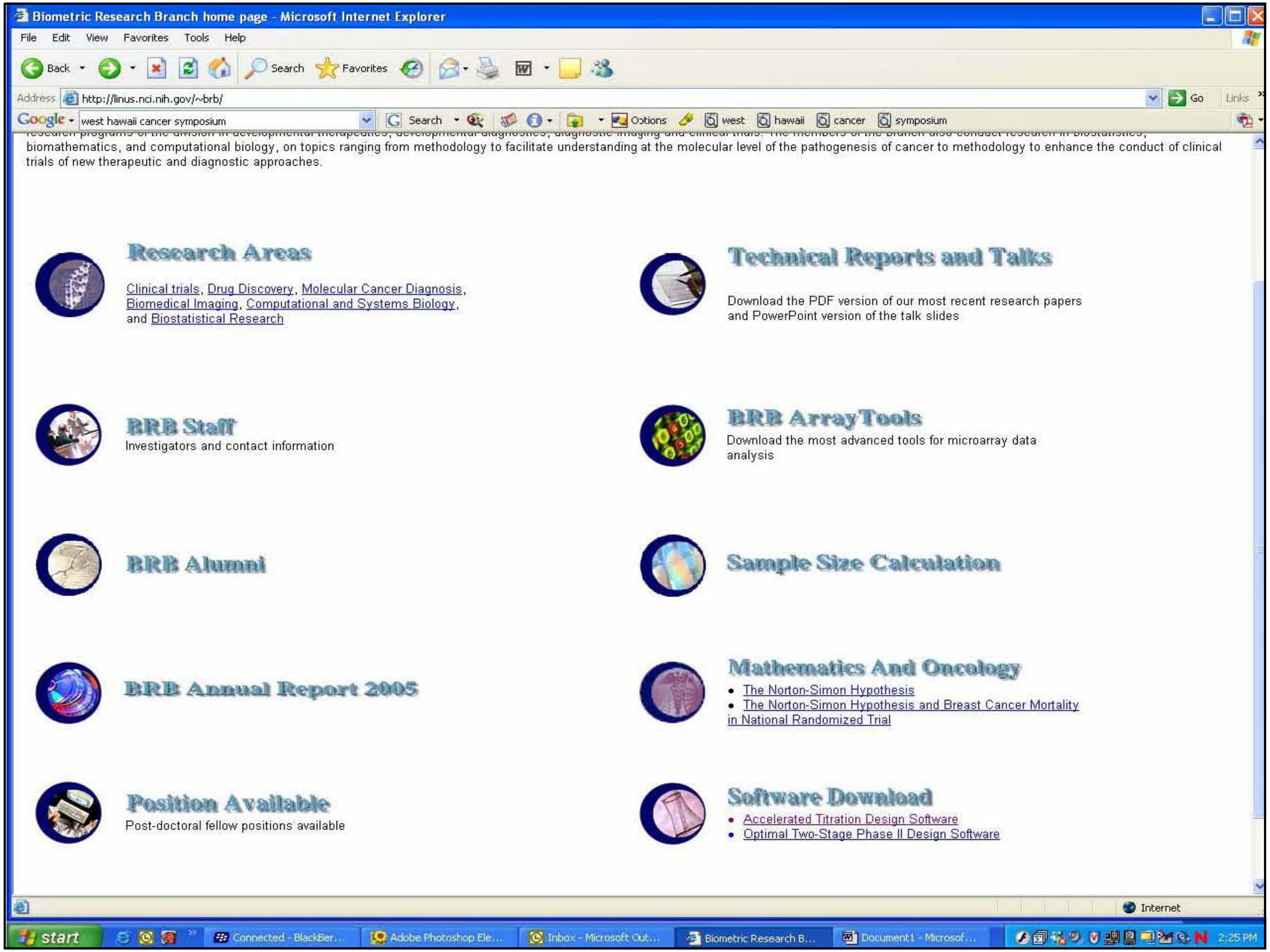


Optimal Two-Stage Design

- Enter n_1 patients
- If response rate ⦿ r_1/n_1 reject drug
- Otherwise, enter n_2 additional patients
- If response rate ⦿ $r_2/(n_1+n_2)$ reject drug



- To distinguish 5% (p_0) response rate from 25% (p_1) response rate with 10% false positive and false negative error rates:
    - – Accrue 9 patients. Stop if no responses
    - – If at least 1 response in first 9, continue accrual to 24 patients total
        - "Accept" treatment if at least 3/24 responses
- For regimens with 5% true response rate, the probability of stopping after 9 patients is 63%



research programs of the division in developmental therapeutics, developmental diagnostics, diagnostic imaging and clinical trials. The members of the branch also conduct research in biostatistics, biomathematics, and computational biology, on topics ranging from methodology to facilitate understanding at the molecular level of the pathogenesis of cancer to methodology to enhance the conduct of clinical trials of new therapeutic and diagnostic approaches.

Research Areas

Clinical trials, Drug Discovery, Molecular Cancer Diagnosis, Biomedical Imaging, Computational and Systems Biology, and Biostatistical Research

Technical Reports and Talks

Download the PDF version of our most recent research papers and PowerPoint version of the talk slides

BRB Staff

Investigators and contact information

BRB ArrayTools

Download the most advanced tools for microarray data analysis

BRB Alumni

Sample Size Calculation

BRB Annual Report 2005

Mathematics And Oncology

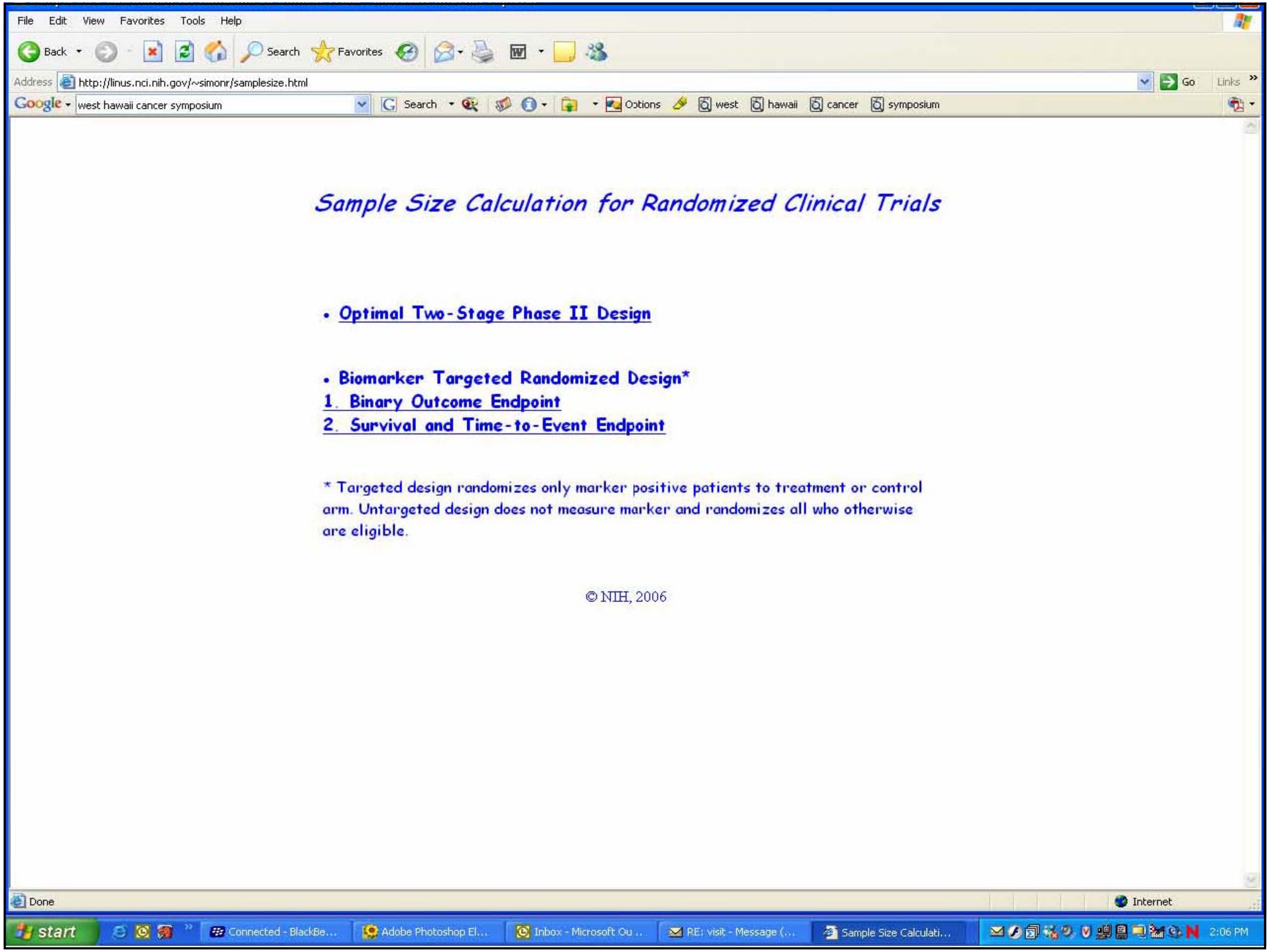
- The Norton-Simon Hypothesis
- The Norton-Simon Hypothesis and Breast Cancer Mortality in National Randomized Trial

Position Available

Post-doctoral fellow positions available

Software Download

- Accelerated Titration Design Software
- Optimal Two-Stage Phase II Design Software



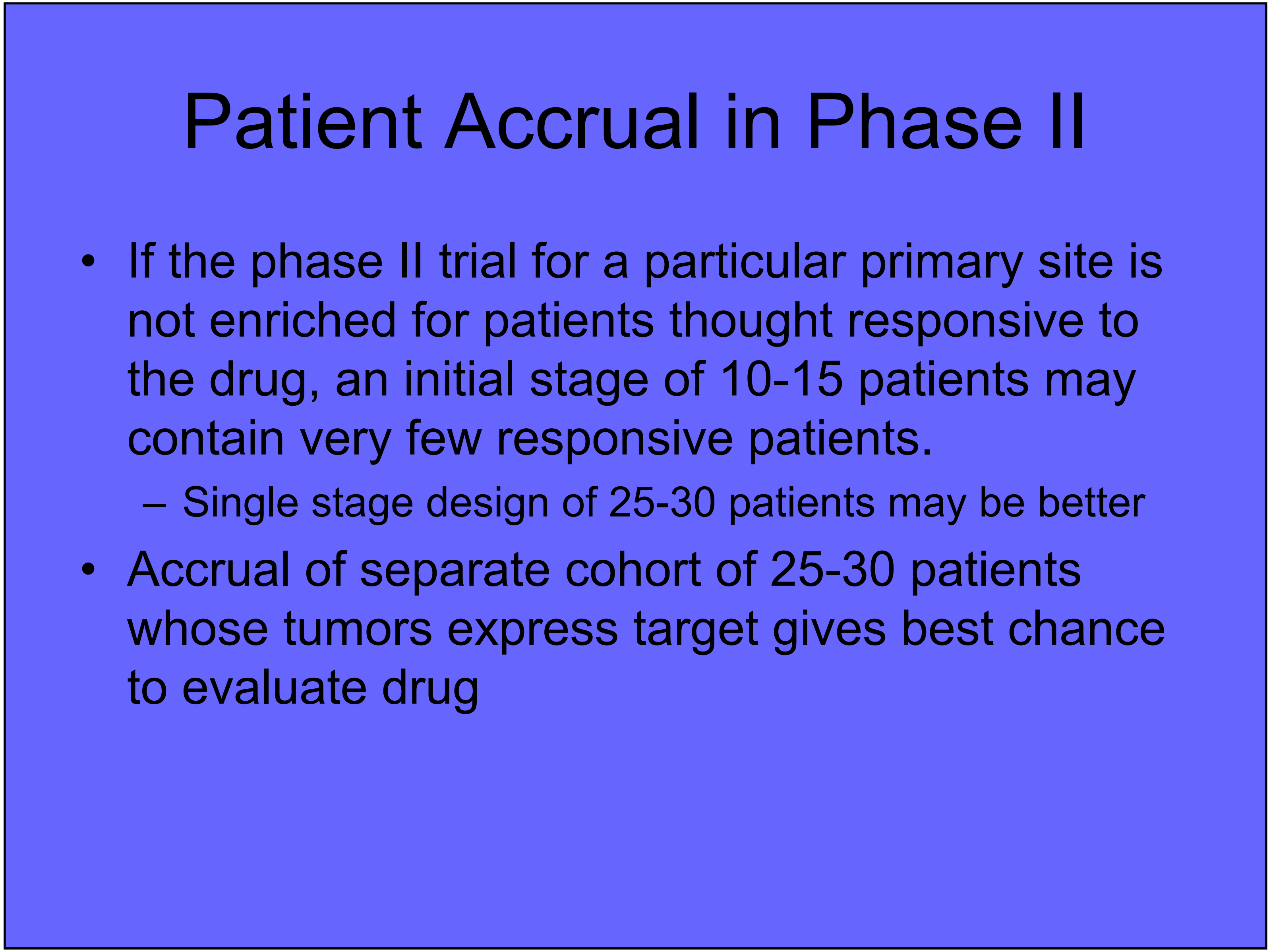
Sample Size Calculation for Randomized Clinical Trials

- Optimal Two-Stage Phase II Design

- Biomarker Targeted Randomized Design*

1. Binary Outcome Endpoint
2. Survival and Time-to-Event Endpoint

* Targeted design randomizes only marker positive patients to treatment or control arm. Untargeted design does not measure marker and randomizes all who otherwise are eligible.



Patient Accrual in Phase II

- If the phase II trial for a particular primary site is not enriched for patients thought responsive to the drug, an initial stage of 10-15 patients may contain very few responsive patients.
  - Single stage design of 25-30 patients may be better
- Accrual of separate cohort of 25-30 patients whose tumors express target gives best chance to evaluate drug

Non-randomized Phase II Designs of Combinations

- Difficult to interpret
  - Activity compared to what?
  - How accurately is outcome for control group known based on past data
  - Is there a comparable group of past patients receiving the control regimen
  - How comparable is response assessment and follow-up evaluation for historical control group?

Limitations in Using Optimal Two-Stage Designs in Evaluating a New Drug with Active Agents

- For a new drug in combination with active agents, p_0 represents the response probability of the active agents without the new drug in the same type of patients being selected for the phase II study of the combination regimen

- The effectiveness of the single arm design is limited by the availability of a large number of comparable patients who have been treated with the active agents alone

- For combination regimens, unless p_0 is based on a large number of patients, the methods of Makuch-Simon or Bayesian Thall-Simon designs should be used instead of the optimal two-stage design.

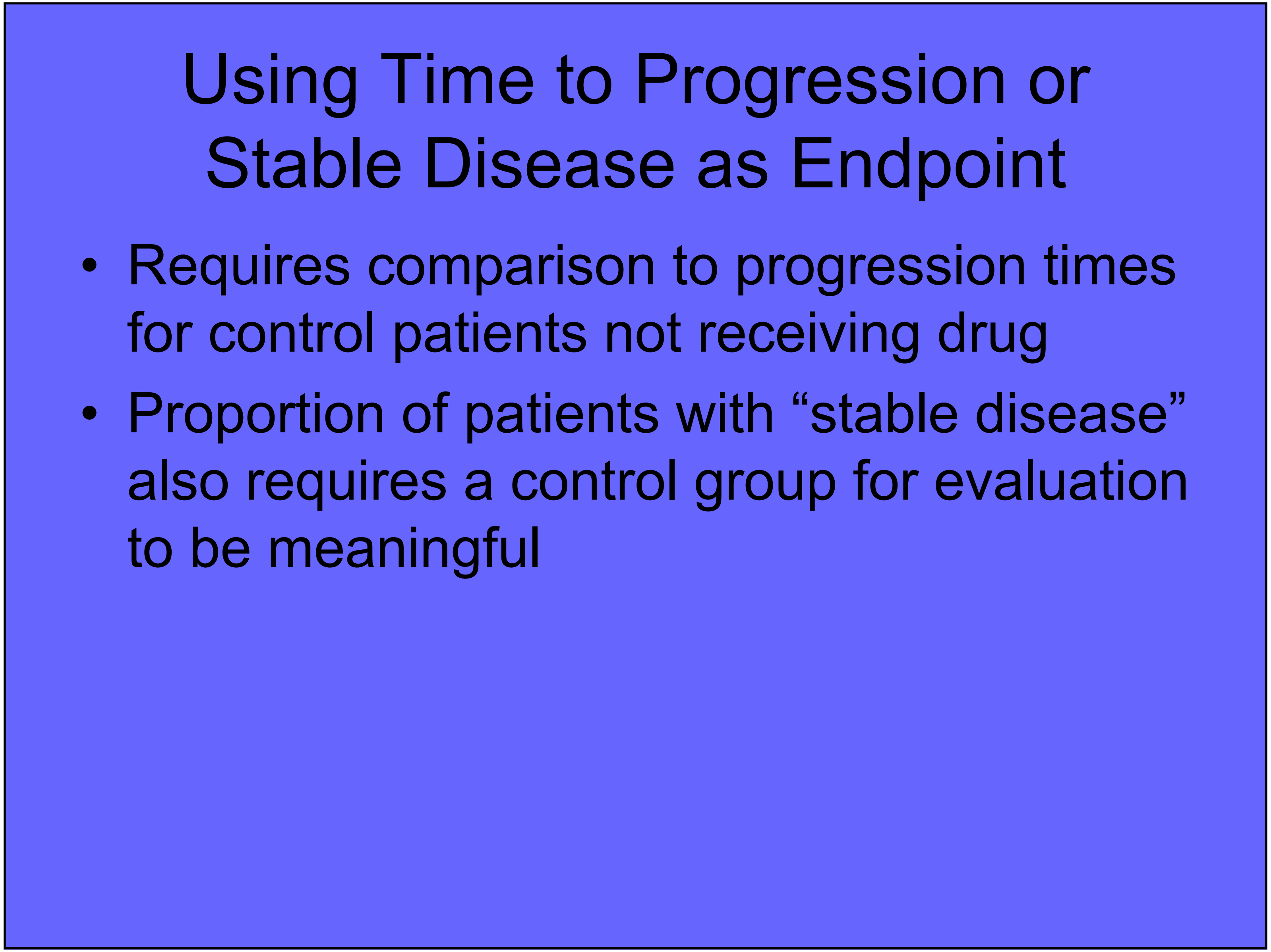
- The Makuch-Simon and Thall-Simon designs require individual patient data for historical controls. This increases focus on comparability and they take into account the actual number of historical controls and the resulting uncertainty in p_0

Thall-Simon Bayesian Single Arm Phase II Designs Using a Specific Set of Historical Control Patients

- Thall, PF, and Simon, R. Incorporating historical control data in planning phase II clinical trials. Stat. in Med. 9:215-228, 1990.
- Thall, P F and Simon R. A Bayesian approach to establishing sample size and monitoring criteria for phase II clinical trials. Controlled Clinical Trials 15:463-481, 1994.
- Thall, PF, Simon R. and Estey E. Bayesian designs for Clinical trials with multiple outcomes.Statistics in Medicine 14:357-379, 1995.
- Thall PF, Simon R, Estey E: A new statistical strategy for monitoring safety and efficacy in single-arm clinical trials. Journal of Clinical Oncology 14:296-303, 1996.

Sample Size Planning for Single Arm Phase II Studies With Historical Control

- Makuch, RW, and Simon, RM.: Sample size considerations for non-randomized comparative studies. J. Chron. Dis. 33: 175-181, 1980.
- Dixon, DO, and Simon, R. Sample size considerations for studies comparing survival curves using historical controls. J. Clin. Epidemiology 41: 1209-1214, 1988.



Using Time to Progression or Stable Disease as Endpoint

- Requires comparison to progression times for control patients not receiving drug
- Proportion of patients with “stable disease” also requires a control group for evaluation to be meaningful

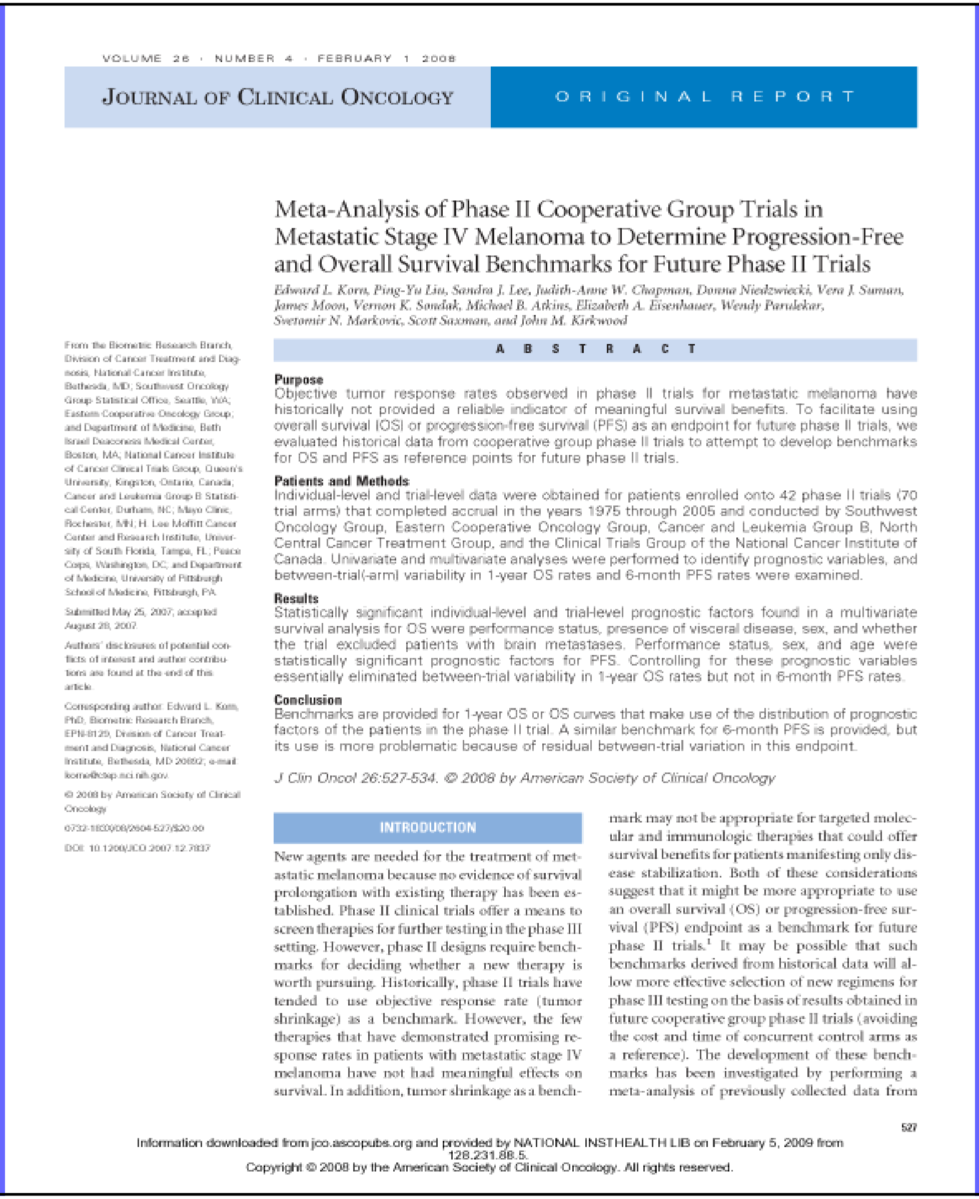
Time to Progression Endpoint

- It is difficult to reliably evaluate time to progression endpoint without a randomized control group
- With historical controls, specific controls should be used for whom comparability of prognosis and surveillance for progression can be established

Number of Patients on Experimental Treatment to have 80% Power for Detecting 15% Absolute Increase (α=.05) in Proportion Without Progression at T months vs Historical Controls

(from Makuch & Simon)

Number of Historical Controls	90% Progression at T in Controls	80% Progression at T In Controls
20	>1000	>1000
30	223	>1000
40	108	285
50	80	167
75	58	101
100	50	83
200	42	65



VOLUME 26 · NUMBER 4 · FEBRUARY 1 2008

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Meta-Analysis of Phase II Cooperative Group Trials in Metastatic Stage IV Melanoma to Determine Progression-Free and Overall Survival Benchmarks for Future Phase II Trials

Edward L. Korn, Ping-Yu Liu, Sandra J. Lee, Judith-Anne W. Chapman, Donna Niedzwiecki, Vera J. Suman, James Moon, Vernon K. Sondak, Michael B. Atkins, Elizabeth A. Eisenhauer, Wendy Parulekar, Svetomir N. Markovic, Scott Saxman, and John M. Kirkwood

From the Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD; Southwest Oncology Group Statistical Office, Seattle, WA; Eastern Cooperative Oncology Group; and Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA; National Cancer Institute of Cancer Clinical Trials Group, Queen's University, Kingston, Ontario, Canada; Cancer and Leukemia Group B Statistical Center, Durham, NC; Mayo Clinic, Rochester, MN; H. Lee Moffitt Cancer Center and Research Institute, University of South Florida, Tampa, FL; Peace Corps, Washington, DC; and Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA.

Submitted May 25, 2007; accepted August 28, 2007.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Edward L. Korn, PhD, Biometric Research Branch, EPN-8129, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD 20892; e-mail: korne@ctep.nci.nih.gov.

0732-183X/08/2604-527/$20.00

DOI: 10.1200/JCO.2007.12.7837

ABSTRACT

Purpose
Objective tumor response rates observed in phase II trials for metastatic melanoma have historically not provided a reliable indicator of meaningful survival benefits. To facilitate using overall survival (OS) or progression-free survival (PFS) as an endpoint for future phase II trials, we evaluated historical data from cooperative group phase II trials to attempt to develop benchmarks for OS and PFS as reference points for future phase II trials.

Patients and Methods
Individual-level and trial-level data were obtained for patients enrolled onto 42 phase II trials (70 trial arms) that completed accrual in the years 1975 through 2005 and conducted by Southwest Oncology Group, Eastern Cooperative Oncology Group, Cancer and Leukemia Group B, North Central Cancer Treatment Group, and the Clinical Trials Group of the National Cancer Institute of Canada. Univariate and multivariate analyses were performed to identify prognostic variables, and between-trial(-arm) variability in 1-year OS rates and 6-month PFS rates were examined.

Results
Statistically significant individual-level and trial-level prognostic factors found in a multivariate survival analysis for OS were performance status, presence of visceral disease, sex, and whether the trial excluded patients with brain metastases. Performance status, sex, and age were statistically significant prognostic factors for PFS. Controlling for these prognostic variables essentially eliminated between-trial variability in 1-year OS rates but not in 6-month PFS rates.

Conclusion
Benchmarks are provided for 1-year OS or OS curves that make use of the distribution of prognostic factors of the patients in the phase II trial. A similar benchmark for 6-month PFS is provided, but its use is more problematic because of residual between-trial variation in this endpoint.

J Clin Oncol 26:527-534. © 2008 by American Society of Clinical Oncology

INTRODUCTION

New agents are needed for the treatment of metastatic melanoma because no evidence of survival prolongation with existing therapy has been established. Phase II clinical trials offer a means to screen therapies for further testing in the phase III setting. However, phase II designs require benchmarks for deciding whether a new therapy is worth pursuing. Historically, phase II trials have tended to use objective response rate (tumor shrinkage) as a benchmark. However, the few therapies that have demonstrated promising response rates in patients with metastatic stage IV melanoma have not had meaningful effects on survival. In addition, tumor shrinkage as a benchmark may not be appropriate for targeted molecular and immunologic therapies that could offer survival benefits for patients manifesting only disease stabilization. Both of these considerations suggest that it might be more appropriate to use an overall survival (OS) or progression-free survival (PFS) endpoint as a benchmark for future phase II trials.[1] It may be possible that such benchmarks derived from historical data will allow more effective selection of new regimens for phase III testing on the basis of results obtained in future cooperative group phase II trials (avoiding the cost and time of concurrent control arms as a reference). The development of these benchmarks has been investigated by performing a meta-analysis of previously collected data from

Issues in the Design of Phase I & II Trials for
Molecularly Targeted Drugs

Richard Simon, D.Sc.
Biometric Research Branch
Division of Cancer Treatment and Diagnosis, National Cancer Institute

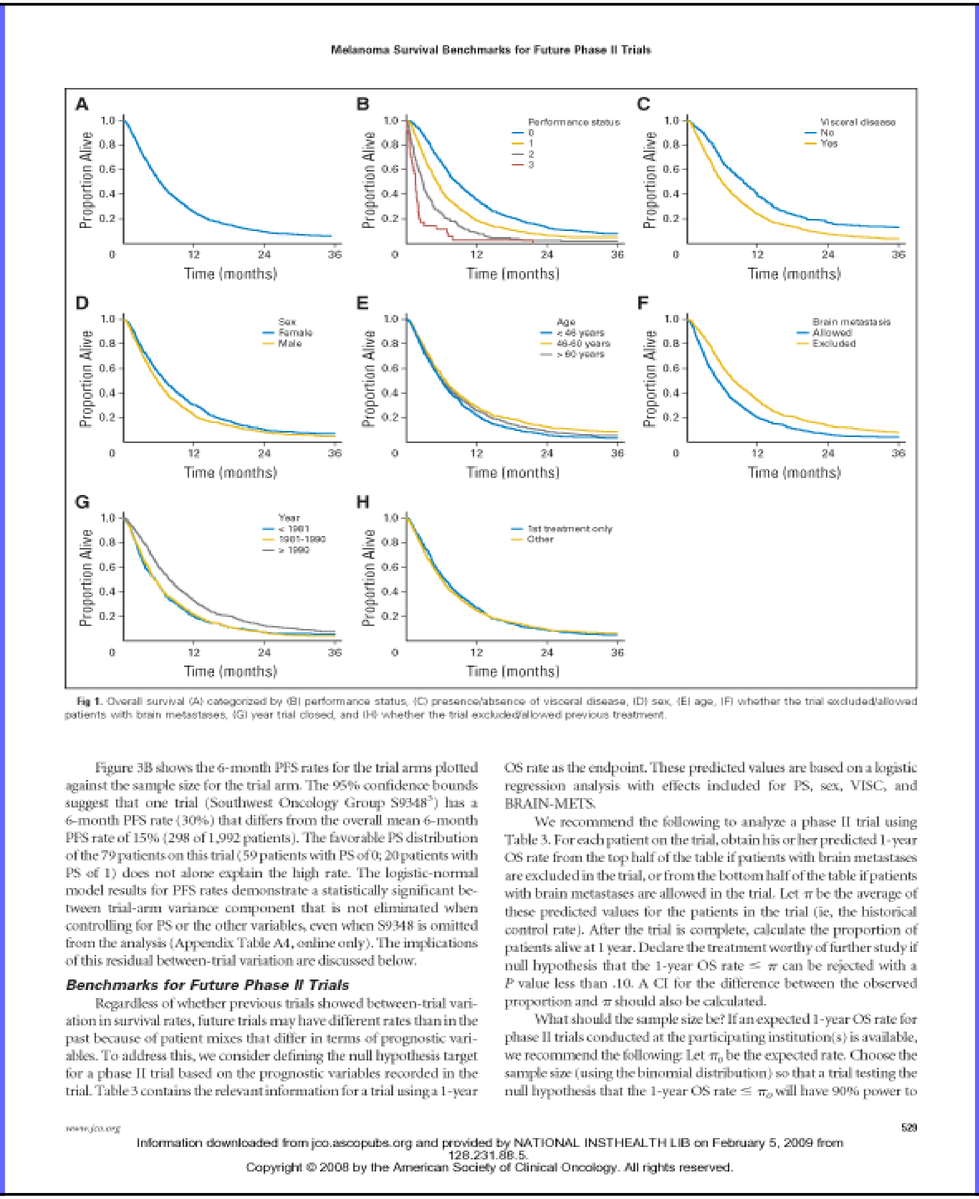


Fig 1. Overall survival (A) categorized by (B) performance status, (C) presence/absence of visceral disease, (D) sex, (E) age, (F) whether the trial excluded/allowed patients with brain metastases, (G) year trial closed, and (H) whether the trial excluded/allowed previous treatment.

Figure 3B shows the 6-month PFS rates for the trial arms plotted against the sample size for the trial arm. The 95% confidence bounds suggest that one trial (Southwest Oncology Group S9348[2]) has a 6-month PFS rate (30%) that differs from the overall mean 6-month PFS rate of 15% (298 of 1,992 patients). The favorable PS distribution of the 79 patients on this trial (59 patients with PS of 0; 20 patients with PS of 1) does not alone explain the high rate. The logistic-normal model results for PFS rates demonstrate a statistically significant between trial-arm variance component that is not eliminated when controlling for PS or the other variables, even when S9348 is omitted from the analysis (Appendix Table A4, online only). The implications of this residual between-trial variation are discussed below.

Benchmarks for Future Phase II Trials

Regardless of whether previous trials showed between-trial variation in survival rates, future trials may have different rates than in the past because of patient mixes that differ in terms of prognostic variables. To address this, we consider defining the null hypothesis target for a phase II trial based on the prognostic variables recorded in the trial. Table 3 contains the relevant information for a trial using a 1-year OS rate as the endpoint. These predicted values are based on a logistic regression analysis with effects included for PS, sex, VISC, and BRAIN-METS.

We recommend the following to analyze a phase II trial using Table 3. For each patient on the trial, obtain his or her predicted 1-year OS rate from the top half of the table if patients with brain metastases are excluded in the trial, or from the bottom half of the table if patients with brain metastases are allowed in the trial. Let π be the average of these predicted values for the patients in the trial (ie, the historical control rate). After the trial is complete, calculate the proportion of patients alive at 1 year. Declare the treatment worthy of further study if null hypothesis that the 1-year OS rate $\leq \pi$ can be rejected with a P value less than .10. A CI for the difference between the observed proportion and π should also be calculated.

What should the sample size be? If an expected 1-year OS rate for phase II trials conducted at the participating institution(s) is available, we recommend the following: Let π_0 be the expected rate. Choose the sample size (using the binomial distribution) so that a trial testing the null hypothesis that the 1-year OS rate $\leq \pi_0$ will have 90% power to

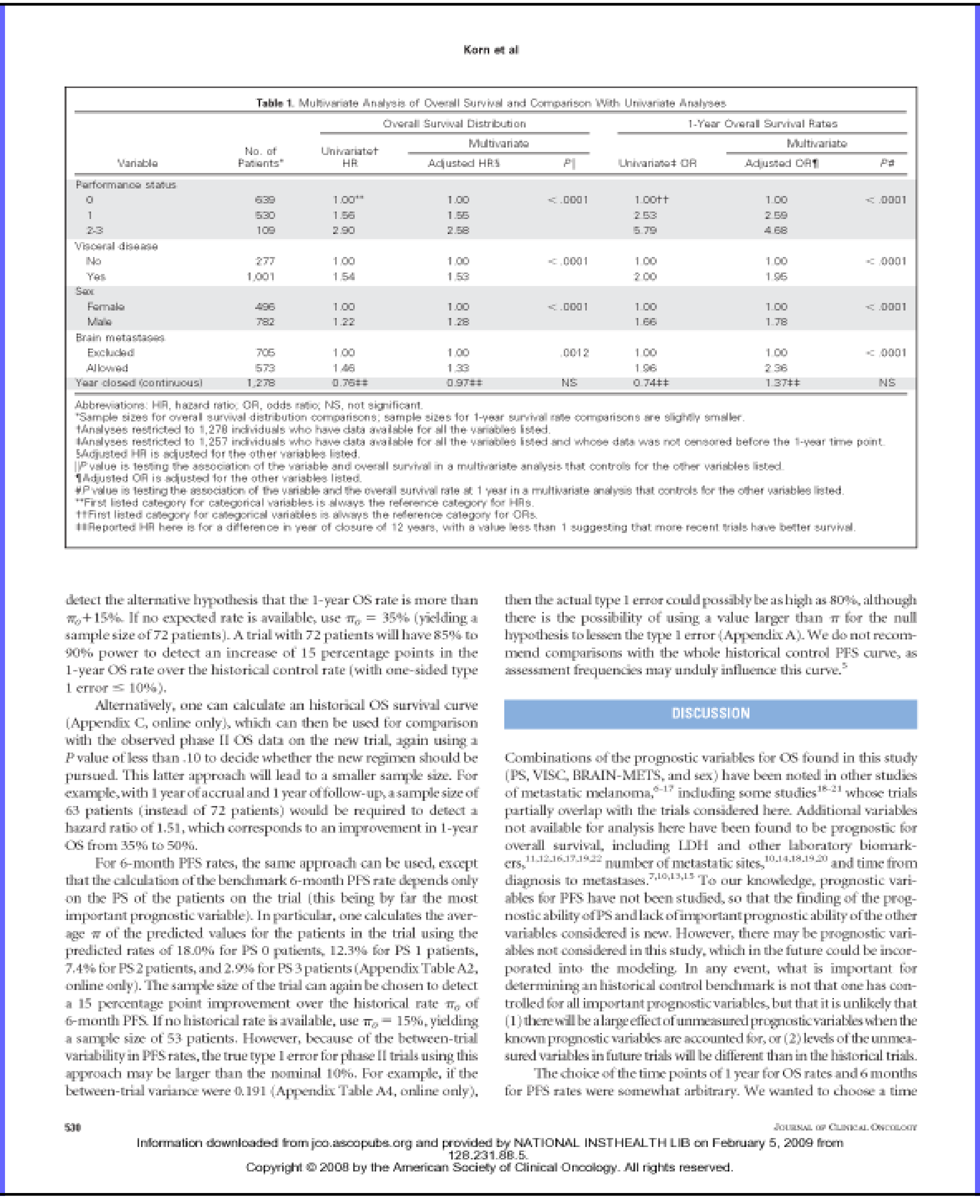


Table 1. Multivariate Analysis of Overall Survival and Comparison With Univariate Analyses

		Overall Survival Distribution			1-Year Overall Survival Rates		
			Multivariate			Multivariate	
Variable	No. of Patients*	Univariate† HR	Adjusted HR§	P‖	Univariate‡ OR	Adjusted OR¶	P#
Performance status							
0	639	1.00**	1.00	< .0001	1.00††	1.00	< .0001
1	530	1.56	1.55		2.53	2.59	
2-3	109	2.90	2.58		5.79	4.68	
Visceral disease							
No	277	1.00	1.00	< .0001	1.00	1.00	< .0001
Yes	1,001	1.54	1.53		2.00	1.95	
Sex							
Female	496	1.00	1.00	< .0001	1.00	1.00	< .0001
Male	782	1.22	1.28		1.66	1.78	
Brain metastases							
Excluded	705	1.00	1.00	.0012	1.00	1.00	< .0001
Allowed	573	1.46	1.33		1.96	2.36	
Year closed (continuous)	1,278	0.76‡‡	0.97‡‡	NS	0.74‡‡	1.37‡‡	NS

Abbreviations: HR, hazard ratio; OR, odds ratio; NS, not significant.
*Sample sizes for overall survival distribution comparisons; sample sizes for 1-year survival rate comparisons are slightly smaller.
†Analyses restricted to 1,278 individuals who have data available for all the variables listed.
‡Analyses restricted to 1,257 individuals who have data available for all the variables listed and whose data was not censored before the 1-year time point.
§Adjusted HR is adjusted for the other variables listed.
‖P value is testing the association of the variable and overall survival in a multivariate analysis that controls for the other variables listed.
¶Adjusted OR is adjusted for the other variables listed.
#P value is testing the association of the variable and the overall survival rate at 1 year in a multivariate analysis that controls for the other variables listed.
**First listed category for categorical variables is always the reference category for HRs.
††First listed category for categorical variables is always the reference category for ORs.
‡‡Reported HR here is for a difference in year of closure of 12 years, with a value less than 1 suggesting that more recent trials have better survival.

detect the alternative hypothesis that the 1-year OS rate is more than π_0+15%. If no expected rate is available, use π_0 = 35% (yielding a sample size of 72 patients). A trial with 72 patients will have 85% to 90% power to detect an increase of 15 percentage points in the 1-year OS rate over the historical control rate (with one-sided type 1 error ≤ 10%).

Alternatively, one can calculate an historical OS survival curve (Appendix C, online only), which can then be used for comparison with the observed phase II OS data on the new trial, again using a *P* value of less than .10 to decide whether the new regimen should be pursued. This latter approach will lead to a smaller sample size. For example, with 1 year of accrual and 1 year of follow-up, a sample size of 63 patients (instead of 72 patients) would be required to detect a hazard ratio of 1.51, which corresponds to an improvement in 1-year OS from 35% to 50%.

For 6-month PFS rates, the same approach can be used, except that the calculation of the benchmark 6-month PFS rate depends only on the PS of the patients on the trial (this being by far the most important prognostic variable). In particular, one calculates the average π of the predicted values for the patients in the trial using the predicted rates of 18.0% for PS 0 patients, 12.3% for PS 1 patients, 7.4% for PS 2 patients, and 2.9% for PS 3 patients (Appendix Table A2, online only). The sample size of the trial can again be chosen to detect a 15 percentage point improvement over the historical rate π_0 of 6-month PFS. If no historical rate is available, use π_0 = 15%, yielding a sample size of 53 patients. However, because of the between-trial variability in PFS rates, the true type 1 error for phase II trials using this approach may be larger than the nominal 10%. For example, if the between-trial variance were 0.191 (Appendix Table A4, online only), then the actual type 1 error could possibly be as high as 80%, although there is the possibility of using a value larger than π for the null hypothesis to lessen the type 1 error (Appendix A). We do not recommend comparisons with the whole historical control PFS curve, as assessment frequencies may unduly influence this curve.[5]

DISCUSSION

Combinations of the prognostic variables for OS found in this study (PS, VISC, BRAIN-METS, and sex) have been noted in other studies of metastatic melanoma,[6-17] including some studies[18-21] whose trials partially overlap with the trials considered here. Additional variables not available for analysis here have been found to be prognostic for overall survival, including LDH and other laboratory biomarkers,[11,12,16,17,19,22] number of metastatic sites,[10,14,18,19,20] and time from diagnosis to metastases.[7,10,13,15] To our knowledge, prognostic variables for PFS have not been studied, so that the finding of the prognostic ability of PS and lack of important prognostic ability of the other variables considered is new. However, there may be prognostic variables not considered in this study, which in the future could be incorporated into the modeling. In any event, what is important for determining an historical control benchmark is not that one has controlled for all important prognostic variables, but that it is unlikely that (1) there will be a large effect of unmeasured prognostic variables when the known prognostic variables are accounted for, or (2) levels of the unmeasured variables in future trials will be different than in the historical trials.

The choice of the time points of 1 year for OS rates and 6 months for PFS rates were somewhat arbitrary. We wanted to choose a time

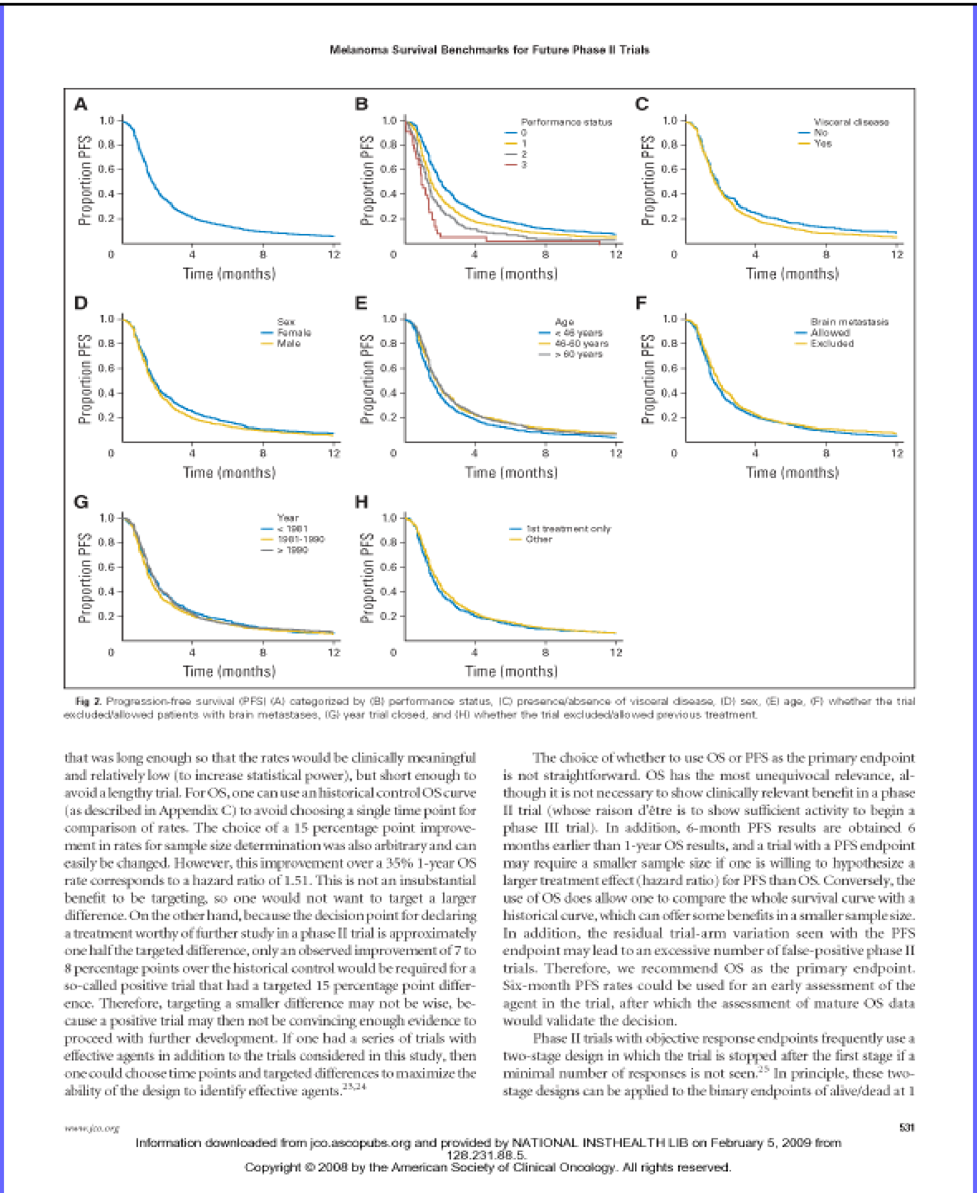
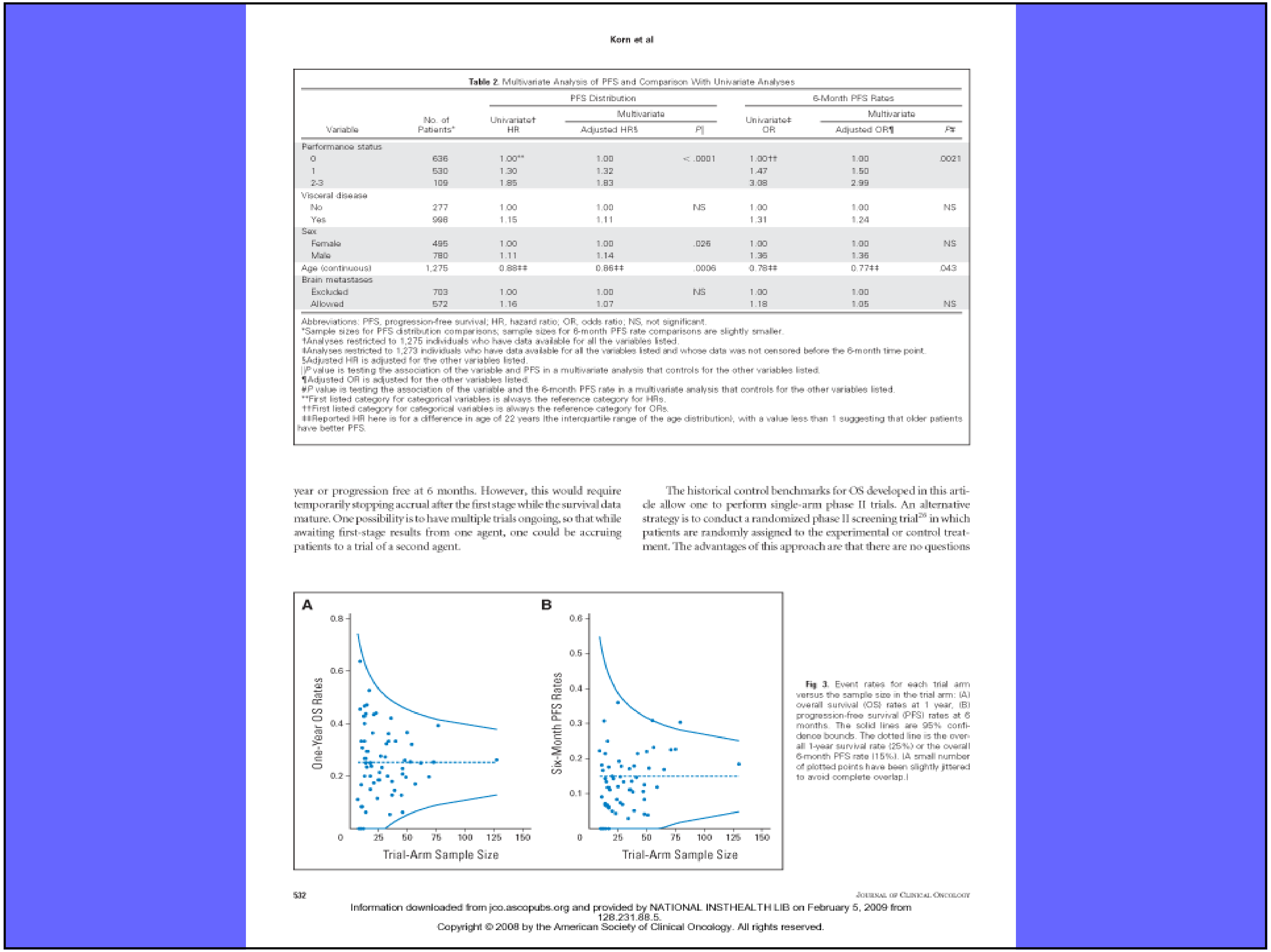


Fig 2. Progression-free survival (PFS) (A) categorized by (B) performance status, (C) presence/absence of visceral disease, (D) sex, (E) age, (F) whether the trial excluded/allowed patients with brain metastases, (G) year trial closed, and (H) whether the trial excluded/allowed previous treatment.

that was long enough so that the rates would be clinically meaningful and relatively low (to increase statistical power), but short enough to avoid a lengthy trial. For OS, one can use an historical control OS curve (as described in Appendix C) to avoid choosing a single time point for comparison of rates. The choice of a 15 percentage point improvement in rates for sample size determination was also arbitrary and can easily be changed. However, this improvement over a 35% 1-year OS rate corresponds to a hazard ratio of 1.51. This is not an insubstantial benefit to be targeting, so one would not want to target a larger difference. On the other hand, because the decision point for declaring a treatment worthy of further study in a phase II trial is approximately one half the targeted difference, only an observed improvement of 7 to 8 percentage points over the historical control would be required for a so-called positive trial that had a targeted 15 percentage point difference. Therefore, targeting a smaller difference may not be wise, because a positive trial may then not be convincing enough evidence to proceed with further development. If one had a series of trials with effective agents in addition to the trials considered in this study, then one could choose time points and targeted differences to maximize the ability of the design to identify effective agents.[23,24]

The choice of whether to use OS or PFS as the primary endpoint is not straightforward. OS has the most unequivocal relevance, although it is not necessary to show clinically relevant benefit in a phase II trial (whose raison d'être is to show sufficient activity to begin a phase III trial). In addition, 6-month PFS results are obtained 6 months earlier than 1-year OS results, and a trial with a PFS endpoint may require a smaller sample size if one is willing to hypothesize a larger treatment effect (hazard ratio) for PFS than OS. Conversely, the use of OS does allow one to compare the whole survival curve with a historical curve, which can offer some benefits in a smaller sample size. In addition, the residual trial-arm variation seen with the PFS endpoint may lead to an excessive number of false-positive phase II trials. Therefore, we recommend OS as the primary endpoint. Six-month PFS rates could be used for an early assessment of the agent in the trial, after which the assessment of mature OS data would validate the decision.

Phase II trials with objective response endpoints frequently use a two-stage design in which the trial is stopped after the first stage if a minimal number of responses is not seen.[25] In principle, these two-stage designs can be applied to the binary endpoints of alive/dead at 1



Korn et al

Table 2. Multivariate Analysis of PFS and Comparison With Univariate Analyses

		PFS Distribution			6-Month PFS Rates		
			Multivariate			Multivariate	
Variable	No. of Patients*	Univariate† HR	Adjusted HR§	P‖	Univariate‡ OR	Adjusted OR¶	P#
Performance status							
0	636	1.00**	1.00	< .0001	1.00††	1.00	.0021
1	530	1.30	1.32		1.47	1.50	
2-3	109	1.85	1.83		3.08	2.99	
Visceral disease							
No	277	1.00	1.00	NS	1.00	1.00	NS
Yes	998	1.15	1.11		1.31	1.24	
Sex							
Female	495	1.00	1.00	.026	1.00	1.00	NS
Male	780	1.11	1.14		1.36	1.36	
Age (continuous)	1,275	0.88‡‡	0.86‡‡	.0006	0.78‡‡	0.77‡‡	.043
Brain metastases							
Excluded	703	1.00	1.00	NS	1.00	1.00	
Allowed	572	1.16	1.07		1.18	1.05	NS

Abbreviations: PFS, progression-free survival; HR, hazard ratio; OR, odds ratio; NS, not significant.
*Sample sizes for PFS distribution comparisons; sample sizes for 6-month PFS rate comparisons are slightly smaller.
†Analyses restricted to 1,275 individuals who have data available for all the variables listed.
‡Analyses restricted to 1,273 individuals who have data available for all the variables listed and whose data was not censored before the 6-month time point.
§Adjusted HR is adjusted for the other variables listed.
‖P value is testing the association of the variable and PFS in a multivariate analysis that controls for the other variables listed.
¶Adjusted OR is adjusted for the other variables listed.
#P value is testing the association of the variable and the 6-month PFS rate in a multivariate analysis that controls for the other variables listed.
**First listed category for categorical variables is always the reference category for HRs.
††First listed category for categorical variables is always the reference category for ORs.
‡‡Reported HR here is for a difference in age of 22 years (the interquartile range of the age distribution), with a value less than 1 suggesting that older patients have better PFS.

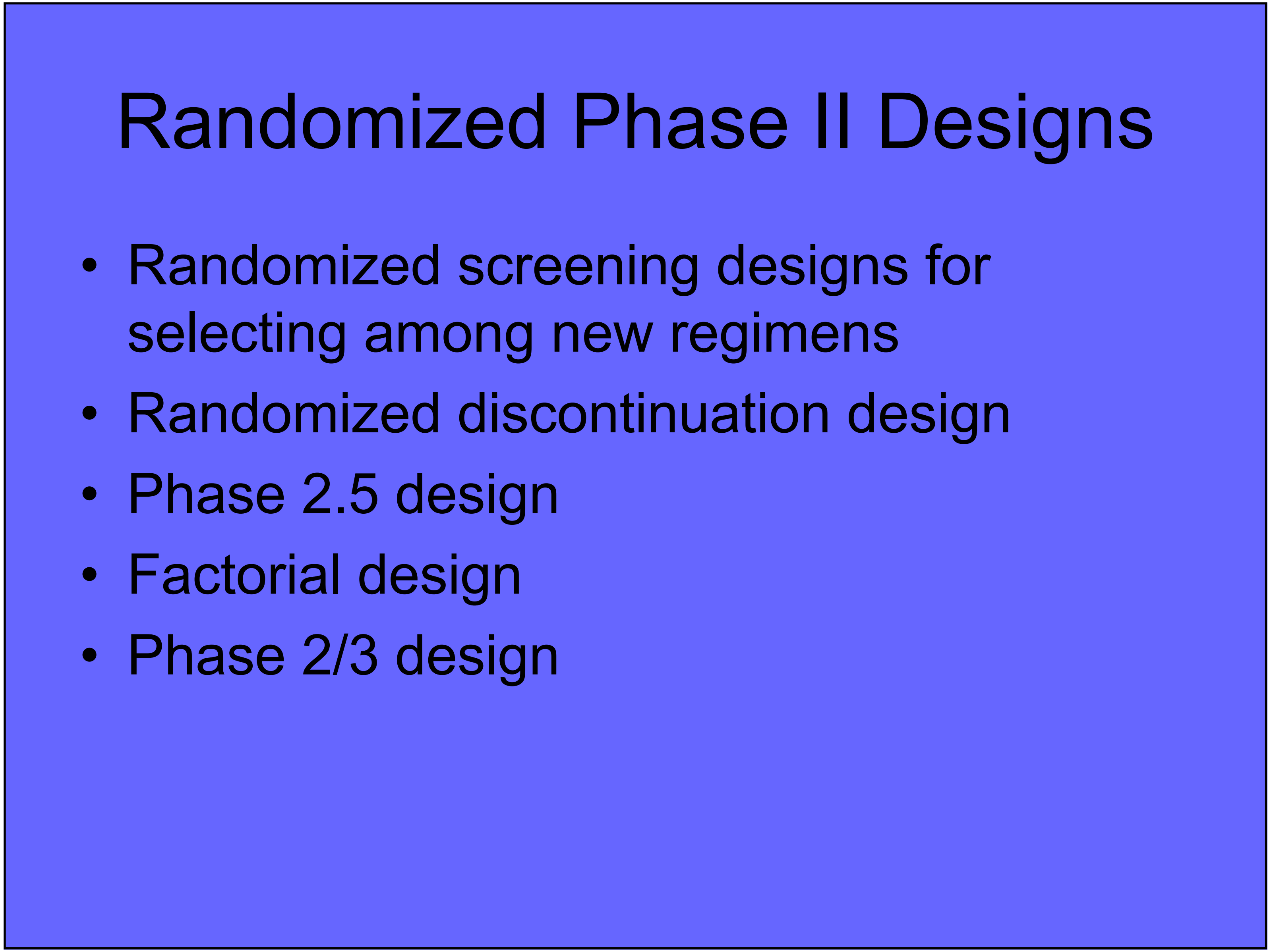
year or progression free at 6 months. However, this would require temporarily stopping accrual after the first stage while the survival data mature. One possibility is to have multiple trials ongoing, so that while awaiting first-stage results from one agent, one could be accruing patients to a trial of a second agent.

The historical control benchmarks for OS developed in this article allow one to perform single-arm phase II trials. An alternative strategy is to conduct a randomized phase II screening trial[28] in which patients are randomly assigned to the experimental or control treatment. The advantages of this approach are that there are no questions

Fig 3. Event rates for each trial arm versus the sample size in the trial arm: (A) overall survival (OS) rates at 1 year, (B) progression-free survival (PFS) rates at 6 months. The solid lines are 95% confidence bounds. The dotted line is the overall 1-year survival rate (25%) or the overall 6-month PFS rate (15%). (A small number of plotted points have been slightly jittered to avoid complete overlap.)

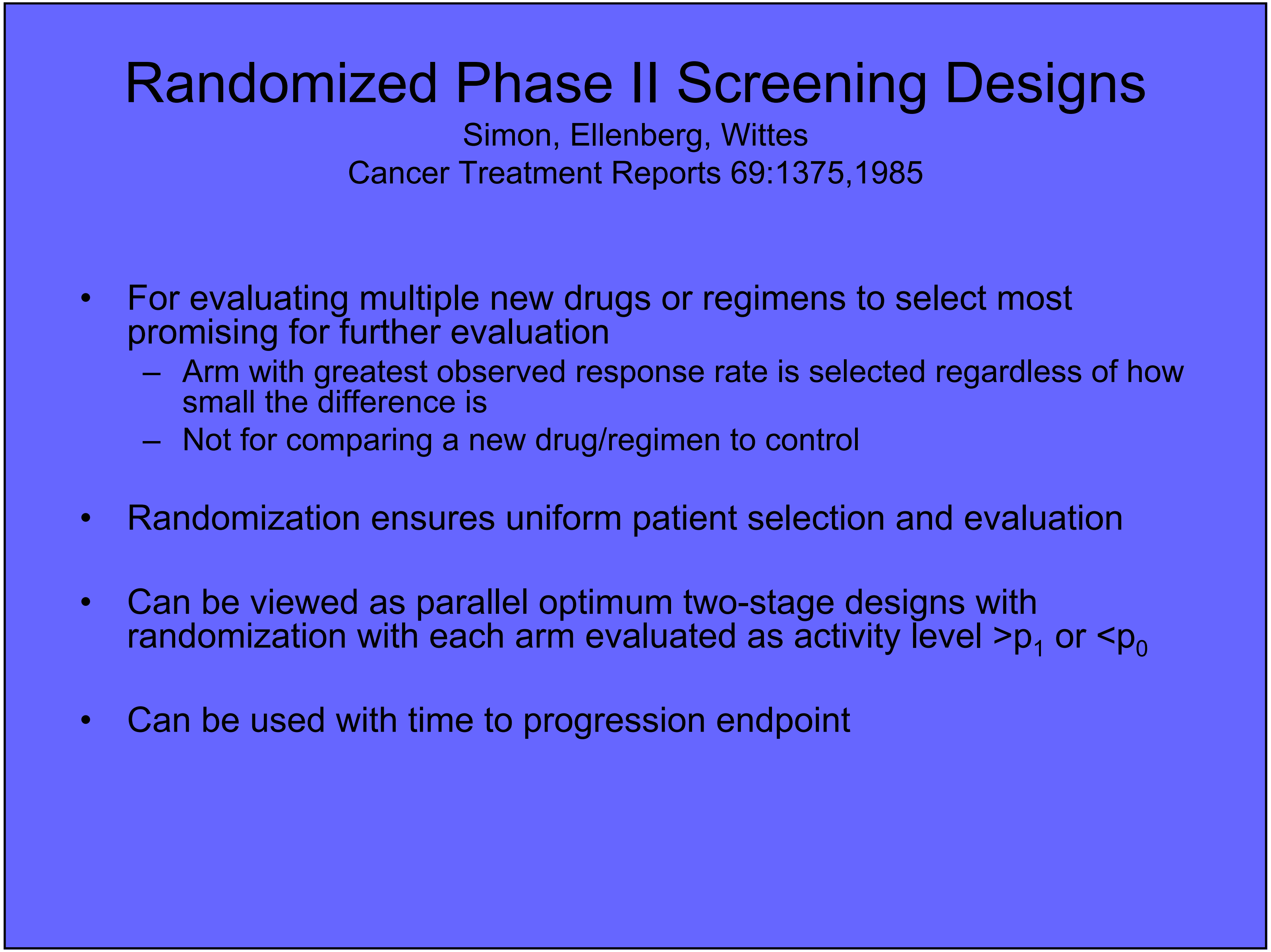
532 Journal of Clinical Oncology

Information downloaded from jco.ascopubs.org and provided by NATIONAL INSTHEALTH LIB on February 5, 2009 from 128.231.88.5.



Randomized Phase II Designs

- Randomized screening designs for selecting among new regimens
- Randomized discontinuation design
- Phase 2.5 design
- Factorial design
- Phase 2/3 design



Randomized Phase II Screening Designs

Simon, Ellenberg, Wittes
Cancer Treatment Reports 69:1375,1985

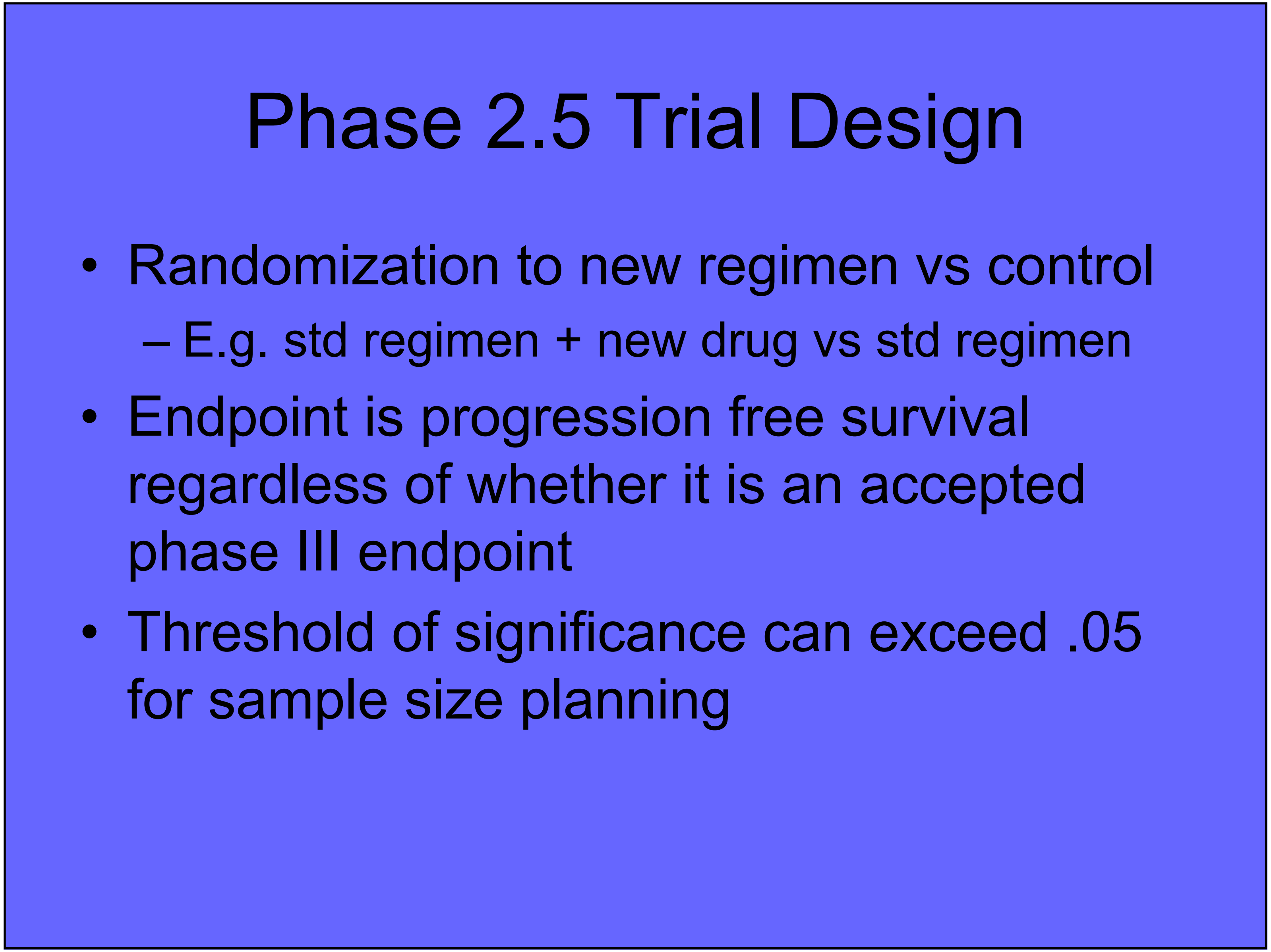
- For evaluating multiple new drugs or regimens to select most promising for further evaluation
  - Arm with greatest observed response rate is selected regardless of how small the difference is
  - Not for comparing a new drug/regimen to control
- Randomization ensures uniform patient selection and evaluation
- Can be viewed as parallel optimum two-stage designs with randomization with each arm evaluated as activity level $>p_1$ or $<p_0$
- Can be used with time to progression endpoint

Patients per Arm for 2-arm Randomized Selection Design Assures Correct Selection When True Response Probabilites Differ by 10%

Response Probability of Inferior Rx	85% Probability of Correct Selection	90% Probability of Correct Selection
5%	20	29
10%	28	42
20%	41	62
40%	54	82

Phase 2.5 Trial Design for Comparing New Regimen to Control Using PFS Endpoint

- Simon R et al. Clinical trial designs for the early clinical development of therapeutic cancer vaccines. Journal of Clinical Oncology 19:1848-54, 2001
- Korn EL et al. Clinical trial designs for cytostatic agents: Are new approaches needed? Journal of Clinical Oncology 19:265-272, 2001



Phase 2.5 Trial Design

- Randomization to new regimen vs control
  - E.g. std regimen + new drug vs std regimen
- Endpoint is progression free survival regardless of whether it is an accepted phase III endpoint
- Threshold of significance can exceed .05 for sample size planning

Number of Events Required for Randomized Trial With Time to Event Endpoint

$$E = 2\left(\frac{k_\alpha + k_\beta}{\ln\left(\delta\right)}\right)^2$$

δ=hazard ratio or ratio of medians

For α=0.05, β=0.20, δ=1.67 (40% reduction in hazard),
E=47 events are required
For α=0.10, 35 events

For α=0.05, β=0.20, δ=1.5, (33% reduction in hazard),
E=75 events are required
For α=0.10, 55 events

Total Sample Size
Randomized Phase 2.5
2 years accrual, 1.5 years followup

Improvement in median PFS	Hazard Ratio	α=.05	α=.10	α=.20
4 → 6 months	1.5	216	168	116
6 → 9 months	1.5	228	176	120
4 → 8 months	2	76	60	40
6→12 months	2	84	64	44

Randomized Discontinuation Design RDD

Ratain et al.

- The RDD starts all patients on the drug
- Patients with early progression go off study
- Patients with objective response continue on the drug
- Patients with stable disease are randomized to continue the drug or stop the drug
- PFS from time of randomization is the endpoint

Randomize Discontinuation Design

- B Freidlin & R Simon. Evaluation of the randomized disconuation design, Journal of Clinical Oncology 23, 2005
- In some cases the RDD is less efficient than a standard up-front randomized phase II design
- RDD may be more efficient than an up-front randomized phase II design if the following conditions hold:
  - Disease is rapidly progressive
  - Most tumors are resistant to the drug
  - Continuous multi-course treatment is required for sensitive tumors
  - No predictive biomarker available for sensitive tumors

Randomized Discontinuation Design (RDD)

- The RDD requires a large sample size
- The RDD is not a phase III trial because it does not establish the clinical utility of administering the drug to the patient compared to not administering it using a phase III endpoint

Phase II/III Design

- Randomized trial comparing regimen containing new drug to control regimen
- Perform interim futility analysis comparing treatments using PFS (progression-free survival) endpoint
- If $p_{pfs}<p^*$ then continue trial to evaluate phase III endpoint
- Otherwise, terminate the trial and consider the new treatment ineffective

Use of Partial Surrogate Endpoints in Integrated Phase II/III Designs

Sally Hunsberger, Yingdong Zhao, and Richard Simon

From the Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute Bethesda MD

Address reprint requests to Sally Hunsberger, PhD, Biometric Research Branch 6130 Executive Blvd, EPN-8120, MSC 7434 National Cancer Institute, Bethesda MD, 20892; phone 301-402-0637; fax 301-4020560; e-mail: sallyh@ctep.nci.nih.gov

ABSTRACT

The traditional oncology drug development paradigm of single arm phase II studies followed by a randomized phase III study has limitations for modern oncology drug development. Interpretation of single arm phase II study results is difficult when a new drug is used in combination with other agents and when progression free survival is used as the endpoint rather than tumor shrinkage. Randomized phase II studies are more informative for these objectives but increase both the number of patients and time required to determine the value of a new experimental agent. In this paper, we compare an integrated phase II/III study design to other study designs to determine the most efficient drug development path in terms of number of patients and length of time to conclusion of drug efficacy on overall survival.

1

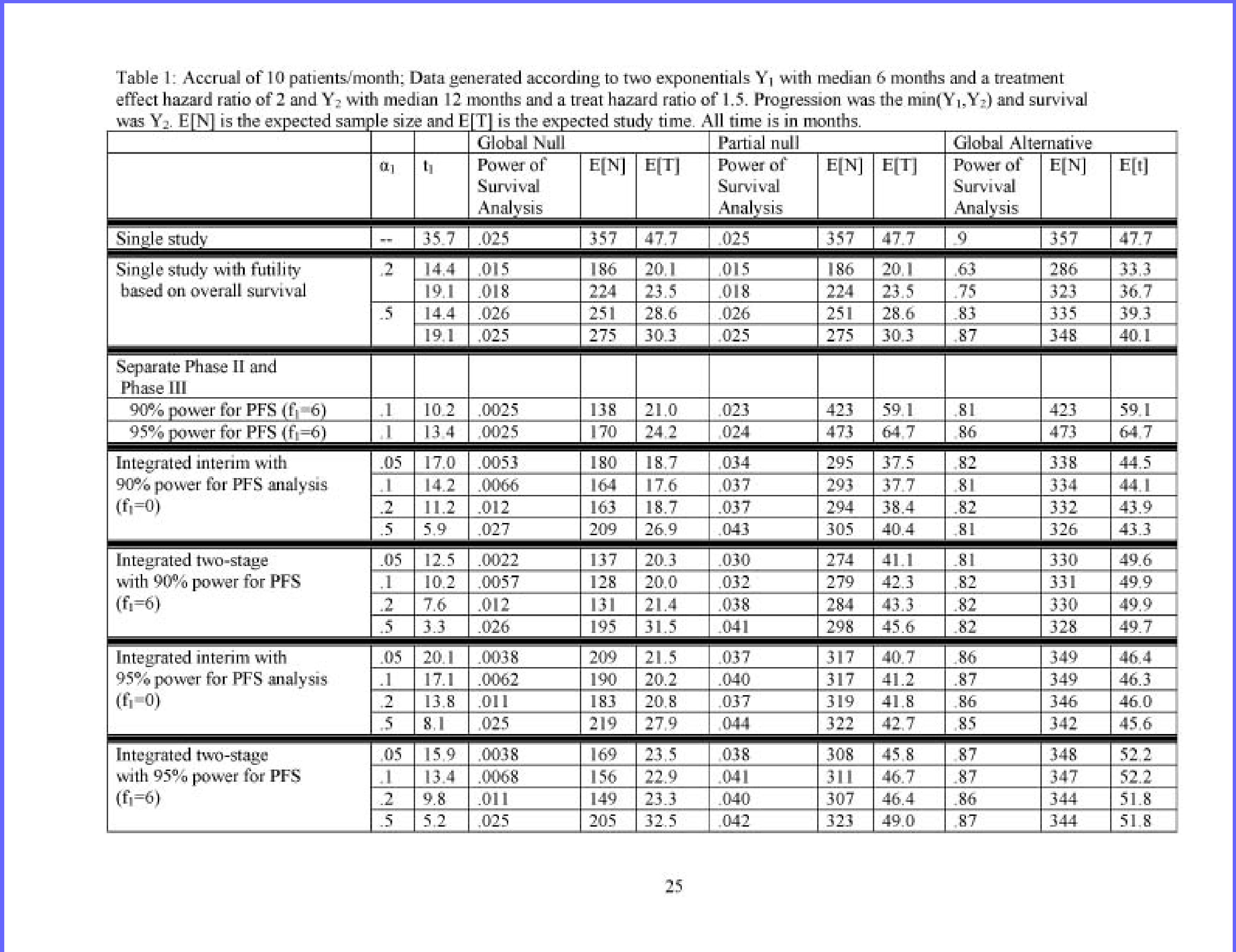
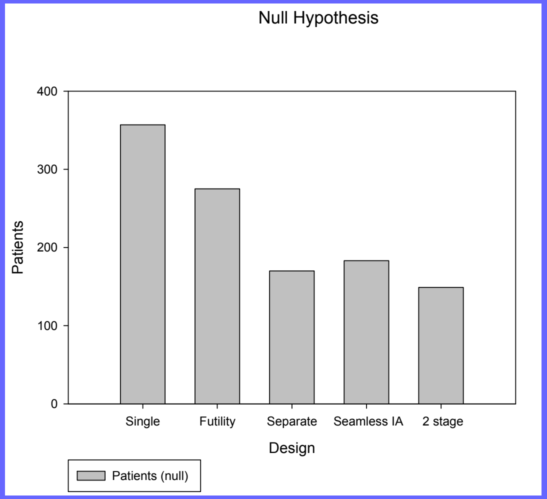
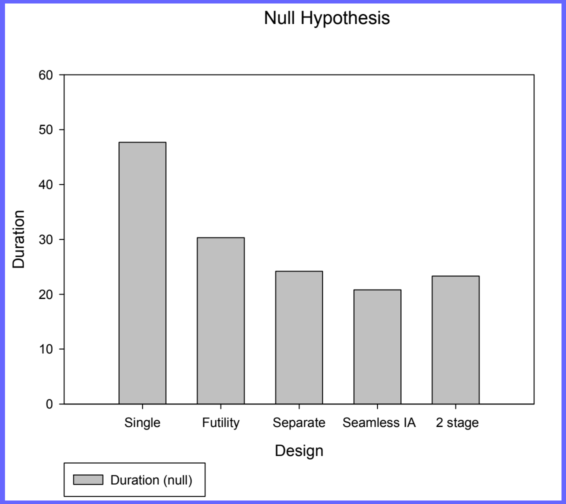
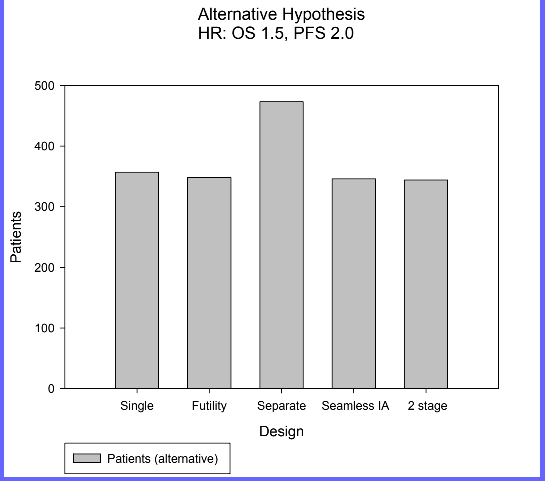


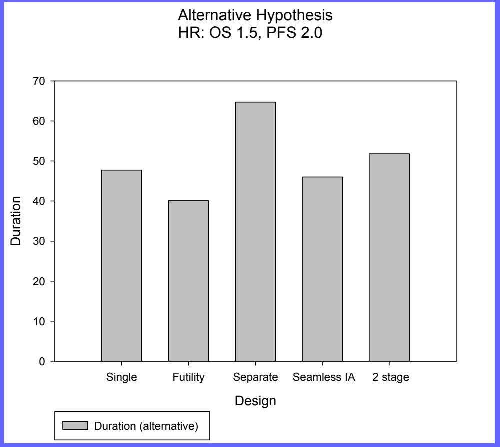
Table 1: Accrual of 10 patients/month; Data generated according to two exponentials Y_1 with median 6 months and a treatment effect hazard ratio of 2 and Y_2 with median 12 months and a treat hazard ratio of 1.5. Progression was the min(Y_1,Y_2) and survival was Y_2. E[N] is the expected sample size and E[T] is the expected study time. All time is in months.

			Global Null			Partial null			Global Alternative		
	α_1	t_1	Power of Survival Analysis	E[N]	E[T]	Power of Survival Analysis	E[N]	E[T]	Power of Survival Analysis	E[N]	E[t]
Single study	--	35.7	.025	357	47.7	.025	357	47.7	.9	357	47.7
Single study with futility based on overall survival	.2	14.4	.015	186	20.1	.015	186	20.1	.63	286	33.3
		19.1	.018	224	23.5	.018	224	23.5	.75	323	36.7
	.5	14.4	.026	251	28.6	.026	251	28.6	.83	335	39.3
		19.1	.025	275	30.3	.025	275	30.3	.87	348	40.1
Separate Phase II and Phase III											
90% power for PFS (f_1=6)	.1	10.2	.0025	138	21.0	.023	423	59.1	.81	423	59.1
95% power for PFS (f_1=6)	.1	13.4	.0025	170	24.2	.024	473	64.7	.86	473	64.7
Integrated interim with 90% power for PFS analysis (f_1=0)	.05	17.0	.0053	180	18.7	.034	295	37.5	.82	338	44.5
	.1	14.2	.0066	164	17.6	.037	293	37.7	.81	334	44.1
	.2	11.2	.012	163	18.7	.037	294	38.4	.82	332	43.9
	.5	5.9	.027	209	26.9	.043	305	40.4	.81	326	43.3
Integrated two-stage with 90% power for PFS (f_1=6)	.05	12.5	.0022	137	20.3	.030	274	41.1	.81	330	49.6
	.1	10.2	.0057	128	20.0	.032	279	42.3	.82	331	49.9
	.2	7.6	.012	131	21.4	.038	284	43.3	.82	330	49.9
	.5	3.3	.026	195	31.5	.041	298	45.6	.82	328	49.7
Integrated interim with 95% power for PFS analysis (f_1=0)	.05	20.1	.0038	209	21.5	.037	317	40.7	.86	349	46.4
	.1	17.1	.0062	190	20.2	.040	317	41.2	.87	349	46.3
	.2	13.8	.011	183	20.8	.037	319	41.8	.86	346	46.0
	.5	8.1	.025	219	27.9	.044	322	42.7	.85	342	45.6
Integrated two-stage with 95% power for PFS (f_1=6)	.05	15.9	.0038	169	23.5	.038	308	45.8	.87	348	52.2
	.1	13.4	.0068	156	22.9	.041	311	46.7	.87	347	52.2
	.2	9.8	.011	149	23.3	.040	307	46.4	.86	344	51.8
	.5	5.2	.025	205	32.5	.042	323	49.0	.87	344	51.8


Null Hypothesis
Patients
400
300
200
100
0
Single
Futility
Separate
Seamless IA
2 stage
Design
Patients (null)


Null Hypothesis
Duration
60
50
40
30
20
10
0
Single
Futility
Separate
Seamless IA
2 stage
Design
Duration (null)


Alternative Hypothesis
HR: OS 1.5, PFS 2.0
Patients
0
100
200
300
400
500
Single
Futility
Separate
Seamless IA
2 stage
Design
Patients (alternative)


Alternative Hypothesis
HR: OS 1.5, PFS 2.0
Duration
0
10
20
30
40
50
60
70
Single
Futility
Separate
Seamless IA
2 stage
Design
Duration (alternative)

Acknowledgements

- Michaele Christian
- Dennis Dixon
- Susan Ellenberg
- Boris Freidlin
- Sally Hunsberger
- Robert Makuch
- Larry Rubinstein
- Peter Thall
- Robert Wittes
- Yingdong Zhao