New-ish Designs for Clinical Trials

Dave Harrington

Dana-Farber Cancer Institute

Harvard University

Acknowledgements/COI

- Examples contributed by
  - Cyrus Mehta, CYTEL
  - Lorenzo Trippa, Dana-Farber and Harvard
  - Mithat Gonen, MSKCC
- Conflicts
  - None to report

Outline

- Some general perspectives
- Bayesian adaptive randomization
- Basket trials
- Flexible/Adaptive Designs
- Reproducibility
- Data sharing

Scientific Issues driving new designs

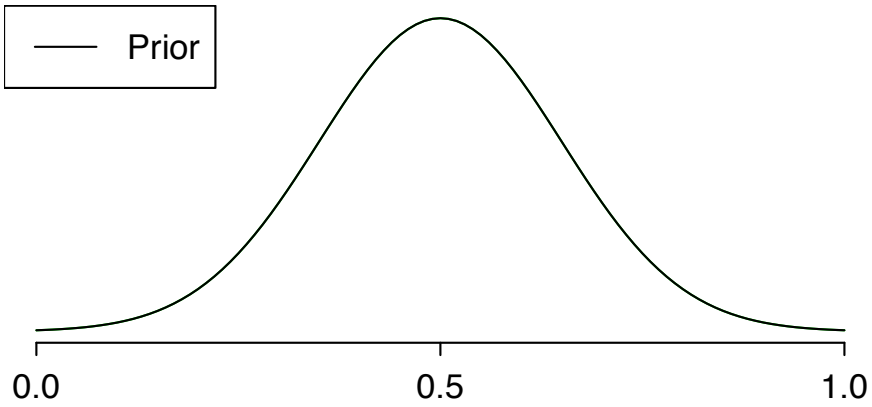
- A deeper understanding of biology for many diseases has led to an increase in the development of new drugs.
  - In many cases, these drugs may work better in patients with particular disease signatures.
  - Drugs may work across phenotypic subtypes of disease that have similar genetic or molecular signatures
- The design of phase III trials is as much art/intuition as science.
  - Too many phase III trials fail because of `poor' design choices – wrong patients, in appropriate target treatment effects, insufficient sample size.

Ecological Issues

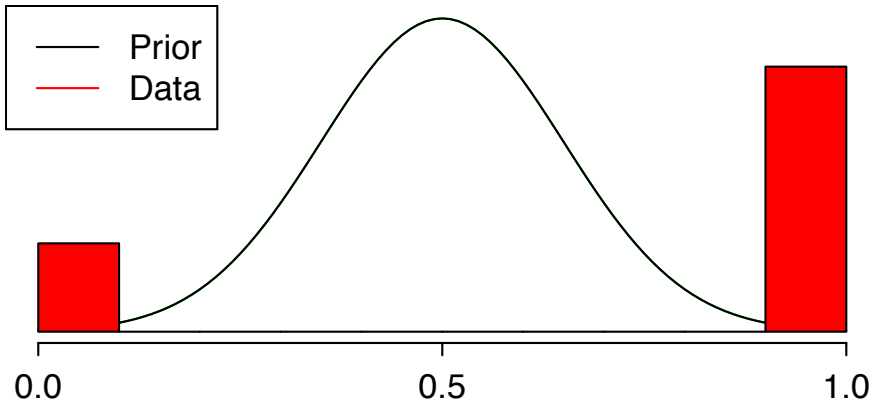
- Increased concern about reproducibility, especially in the results of post-hoc analyses
- Rapid development of new agents sometimes mean trial designers are working without sufficient information on
  - Subgroups
  - Likely treatment effects
  - Adherence

Bayesian Adaptive Randomization (BAR)

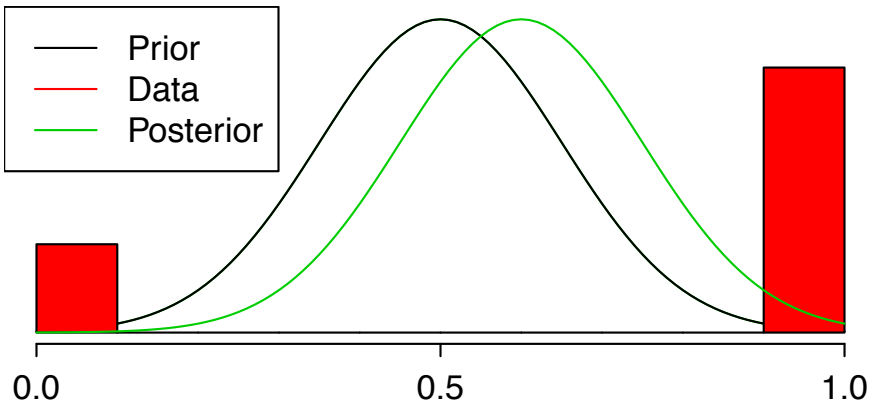
- Simple idea
  - Use accumulating information to adjust randomization fraction
- Complicated execution
  - Modeling assumptions
  - Analysis is complex
  - Logistic support is essential, since data are continually updated



Prior assumption about distribution of response probabilities

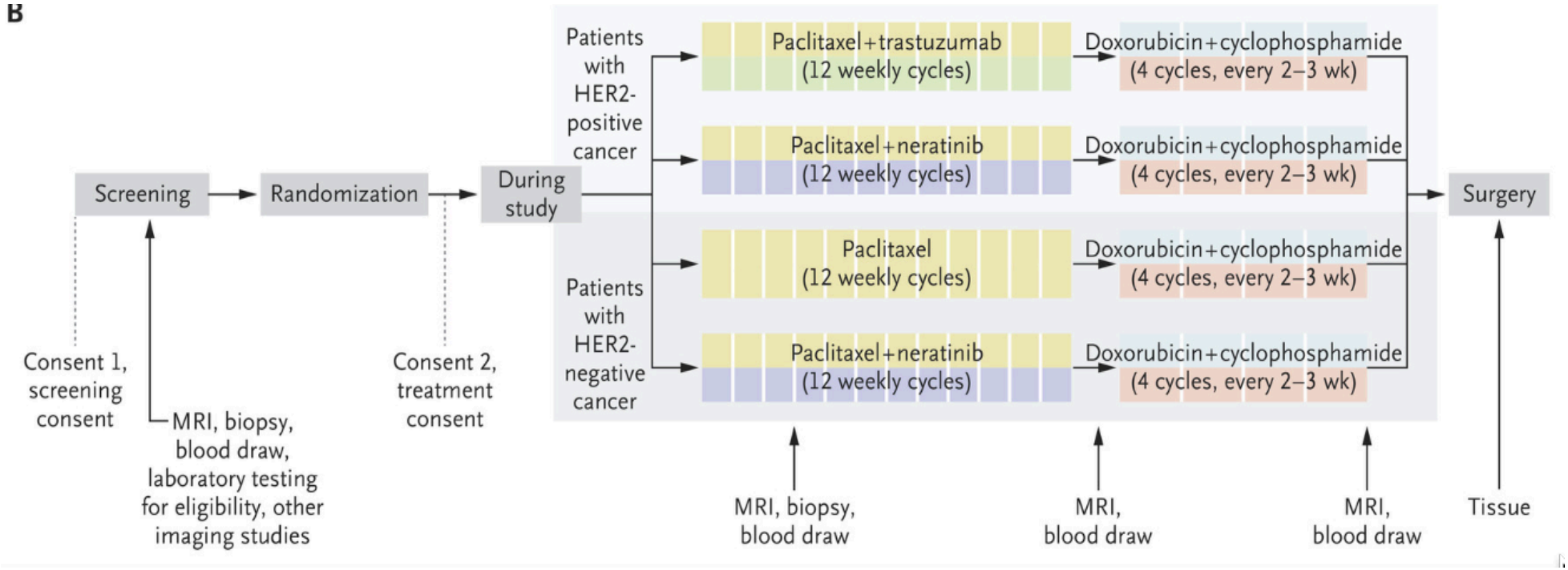


Data during trial on response (x = 1) and non-response (x = 0)

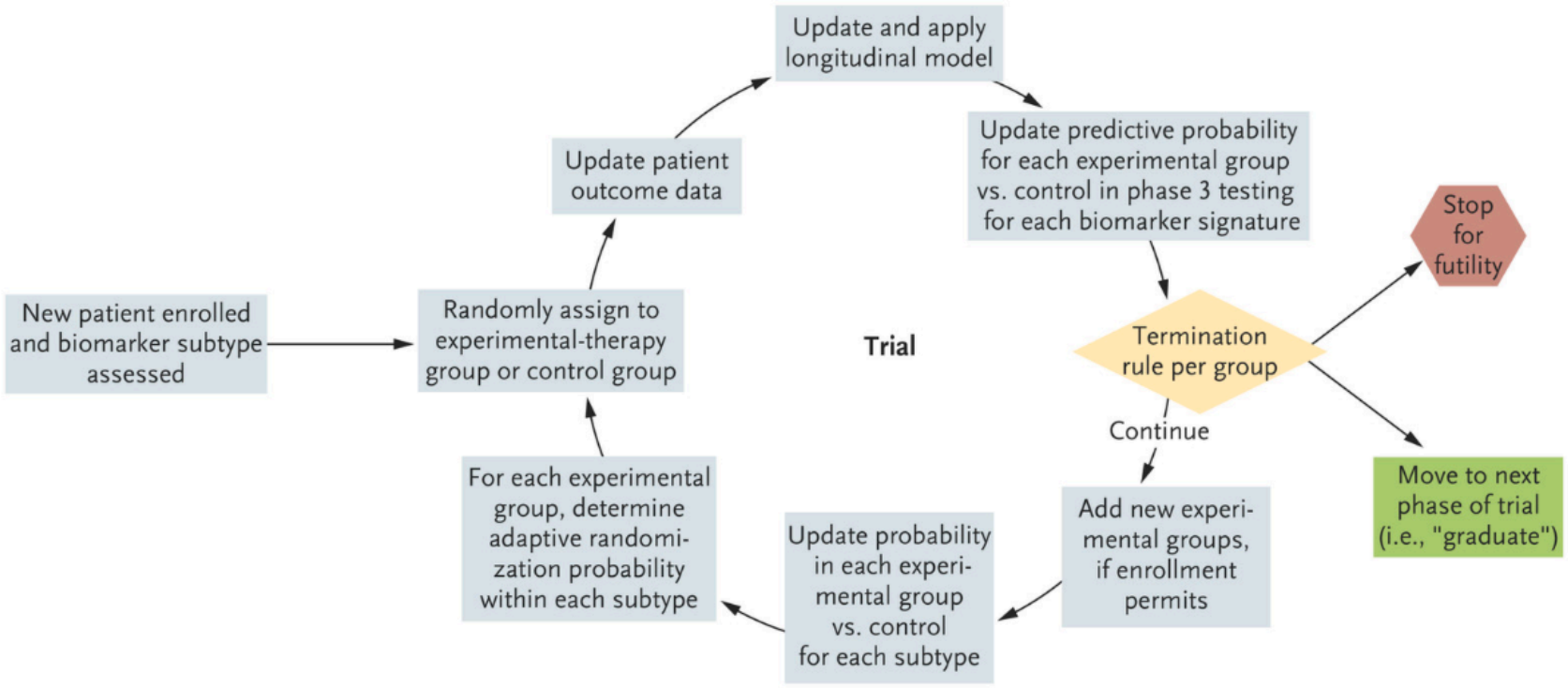


Update assumption of response probabilities (posterior distribution)

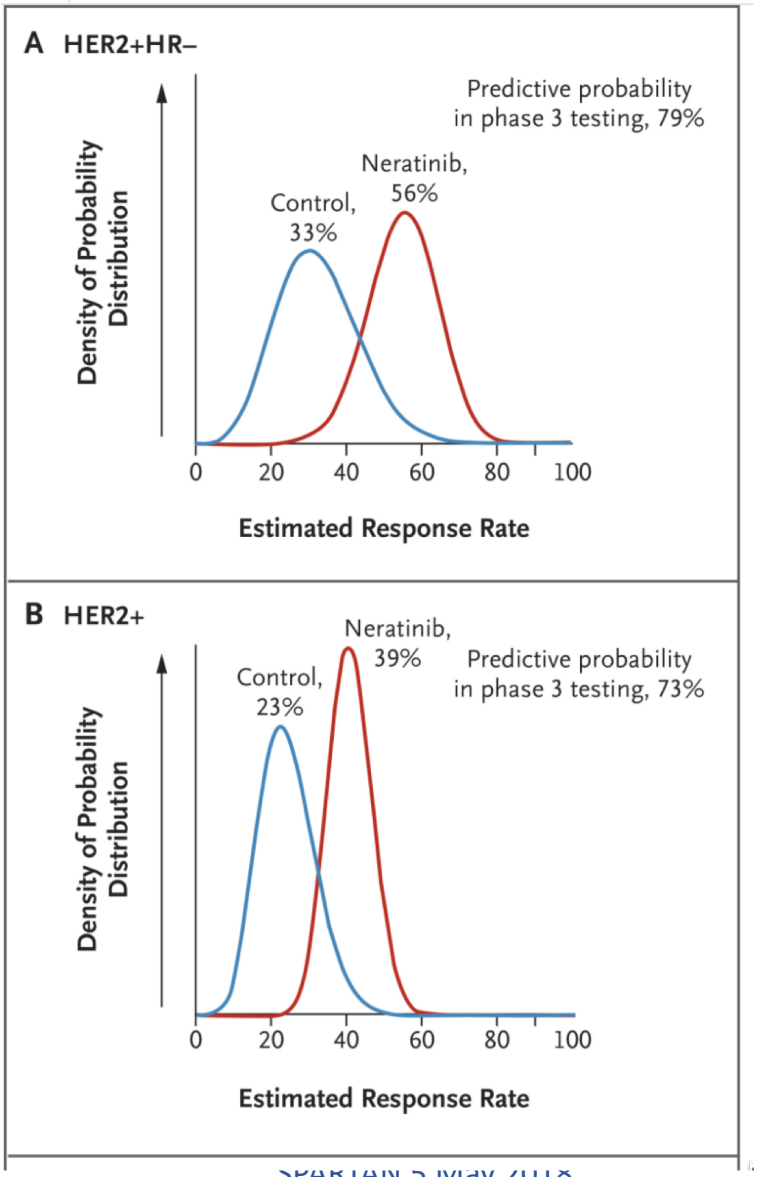
I-SPY 2 (N Engl J Med 2016; 375:11-22)



I-SPY 2 (N Engl J Med 2016; 375:11-22)



I-SPY 2 (N Engl J Med 2016; 375:11-22)



Bayesian versus Frequentist Approaches in Clinical Trials.

Table 1. Bayesian versus Frequentist Approaches in Clinical Trials.

Variable	Bayes	Frequentist
Differences		
Main goal of inference	Predict outcomes of future trials and absolute risk for future patients.	Estimate population average effects.
Assumptions	Requires explicit specification of prior distributions of unknown population parameters. Incorporates a priori knowledge and clinical judgment formally. May be sensitive to specification of prior distributions.	Does not require explicit specification of prior distributions of unknown population parameters. Incorporates a priori knowledge and clinical judgment informally.
Interim monitoring	Only the data actually obtained are relevant for final conclusions (e.g., a credible interval or predictive probability). Whether or not a clinician examines accumulating evidence with the possibility of stopping the trial does not affect inference.	Both the data actually obtained and the probabilities of data not obtained are relevant for final conclusions (e.g., a P value). Whether or not a clinician examines accumulating evidence with the possibility of stopping the trial does affect inference.
Ease of use	Often computationally complex; careful modeling often requires simulation-based calculations.	Often computationally simple, though careful modeling may require simulation-based calculations.
Similarities		
Adaptation	Can incorporate adaptive designs, multistage trials, early stopping, and adaptive randomization.	
Role of statistical judgment	Options for data-driven analyses are available. Skill and substance-area knowledge of the data analyst are important in drawing correct conclusions.	
Compatibility	It is feasible to combine a Bayesian design with a frequentist analysis or a frequentist design with a Bayesian analysis.	
Prior knowledge	Both approaches rely on prior knowledge and clinical judgment (though they incorporate them in different fashions).	

Harrington D, Parmigiani G. N Engl J Med 2016;375:7-9.

The NEW ENGLAND JOURNAL of MEDICINE

Bayesian vs. Hypothetical Standard Frequentist Design.

Table 2. Bayesian vs. Hypothetical Standard Frequentist Design.

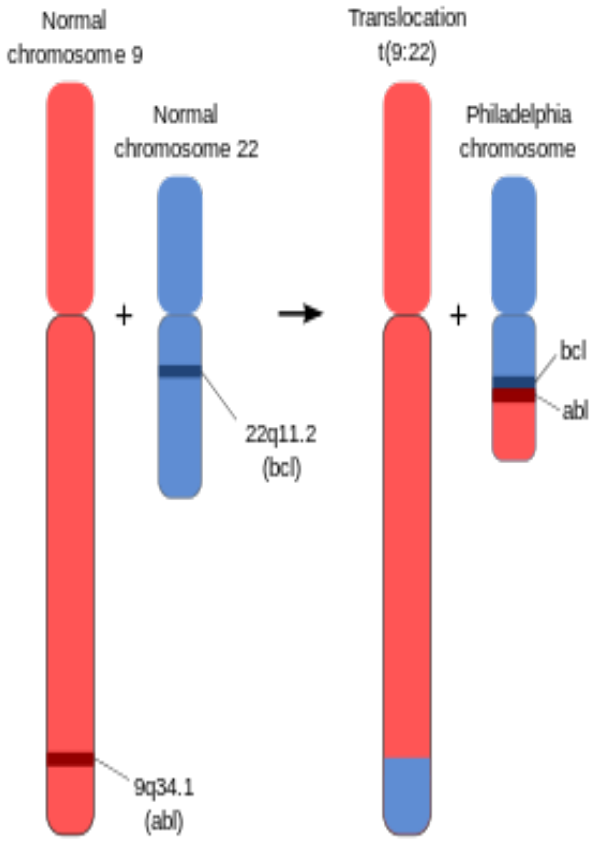
Variable	I-SPY 2 Design	Standard Frequentist Design for I-SPY 2
Main goal of inference	Posterior distributions of rates of pathological complete response for the investigational drug (neratinib or veliparib) and the control. Predicted probability of success in a subsequent phase 3 trial.	Odds ratio or relative risk of response, investigational drug vs. control, with confidence interval and P value
Assumptions	Specification of prior distributions of response rates for investigational drug and control; specification of model for adapting randomization fraction as information becomes available, including model for imputation of pathological complete response based on imaging in previous patients	Specification of anticipated rates of pathological complete response in the control group and of clinically relevant target differences; specification of prior stratification for randomization of subtypes of breast cancer; distributions of unknown population parameters
Randomization	Adaptive randomization increases likelihood of participant receiving treatment assignment that may be of benefit. Estimates of pathological complete response rates must be model-based because of lack of balance of patients' baseline characteristics across treatments.	Constant randomization probabilities do not preferentially target patients who may benefit from a treatment; heterogeneity of patient groups receiving a treatment may dilute estimates of treatment effects. Constant randomization probabilities ensure approximate balance of baseline characteristics across treatments and allow direct comparisons.
Interim monitoring	A treatment is declared potentially successful if predicted probability of success in phase 3 trial is at least 85%. Predicted probability of success is evaluated frequently during the trial. Experimental treatment is dropped for futility if predictive probability of success in a phase 3 trial is <10% in all 10 signatures.	Summary test statistics calculated a small number of times (typically 3 or 4) with P values checked against interim monitoring boundaries for futility and efficacy. Treatment effect estimates can be used for future trials, but groups are not selected on the basis of predicted success rates of future phase 3 trial.
Ease of use	Software for calculation of posterior distribution of pathological complete response or predictive probability of success not generally available. Accruing information must be updated frequently and accurately for adaptive randomization.	Summary and test statistics based on ratios or differences of proportions of pathological complete response. Open-source or other software for design and analysis widely available. Few software packages available for adjusting estimates for treatment effects after early stopping. Accurate data updates required for interim monitoring.

Harrington D, Parmigiani G. N Engl J Med 2016;375:7-9.

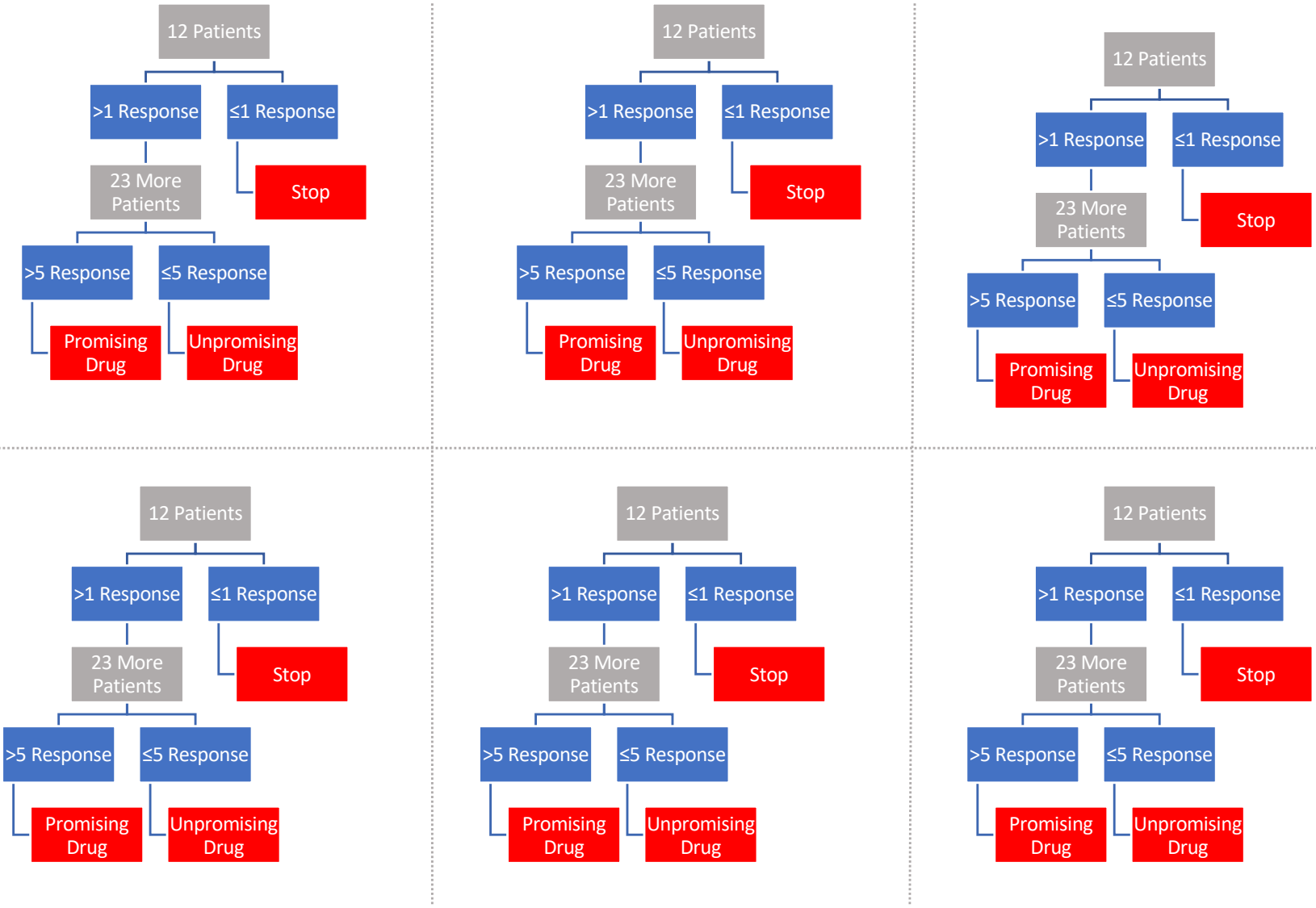
The NEW ENGLAND JOURNAL of MEDICINE

Basket trials

- Basket trials test therapies in diseases that are
  - Phenotypically different
  - Similar in molecular or genetic signatures
- Most often used in Phase 2 trials in cancer
- Example: BCR-ABL Translocation
  - Parts of two chromosomes (9 and 22) switch places
  - Results in a “fusion gene”: juxtaposition of ABL1 gene (9q34) to the BCR gene (22q11)

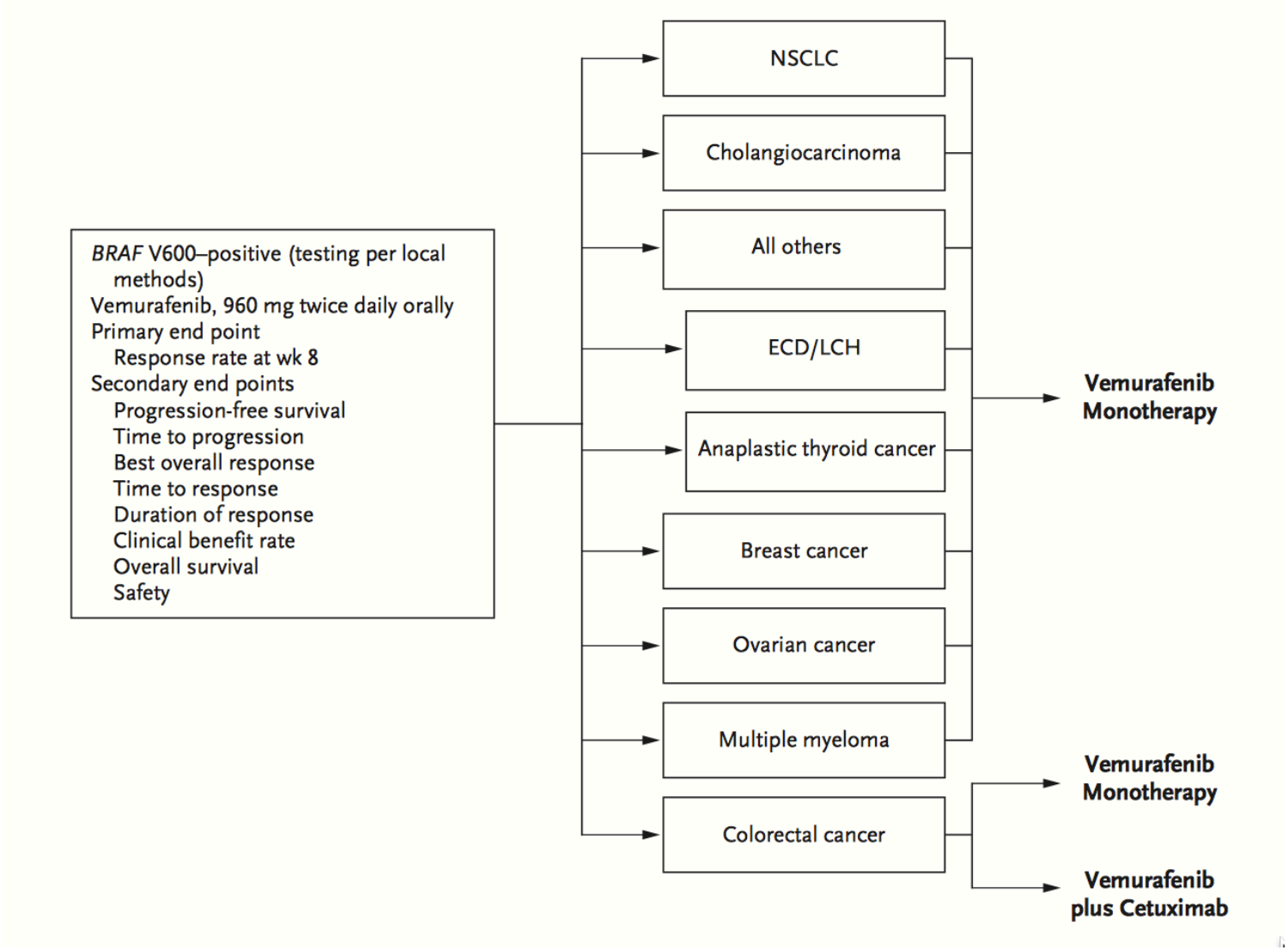


Using the Traditional Design in Basket Trial



Source: M. Gonen, MSKCC

Hyman, et al. (NEJM 373:8, Aug 20, 2015)



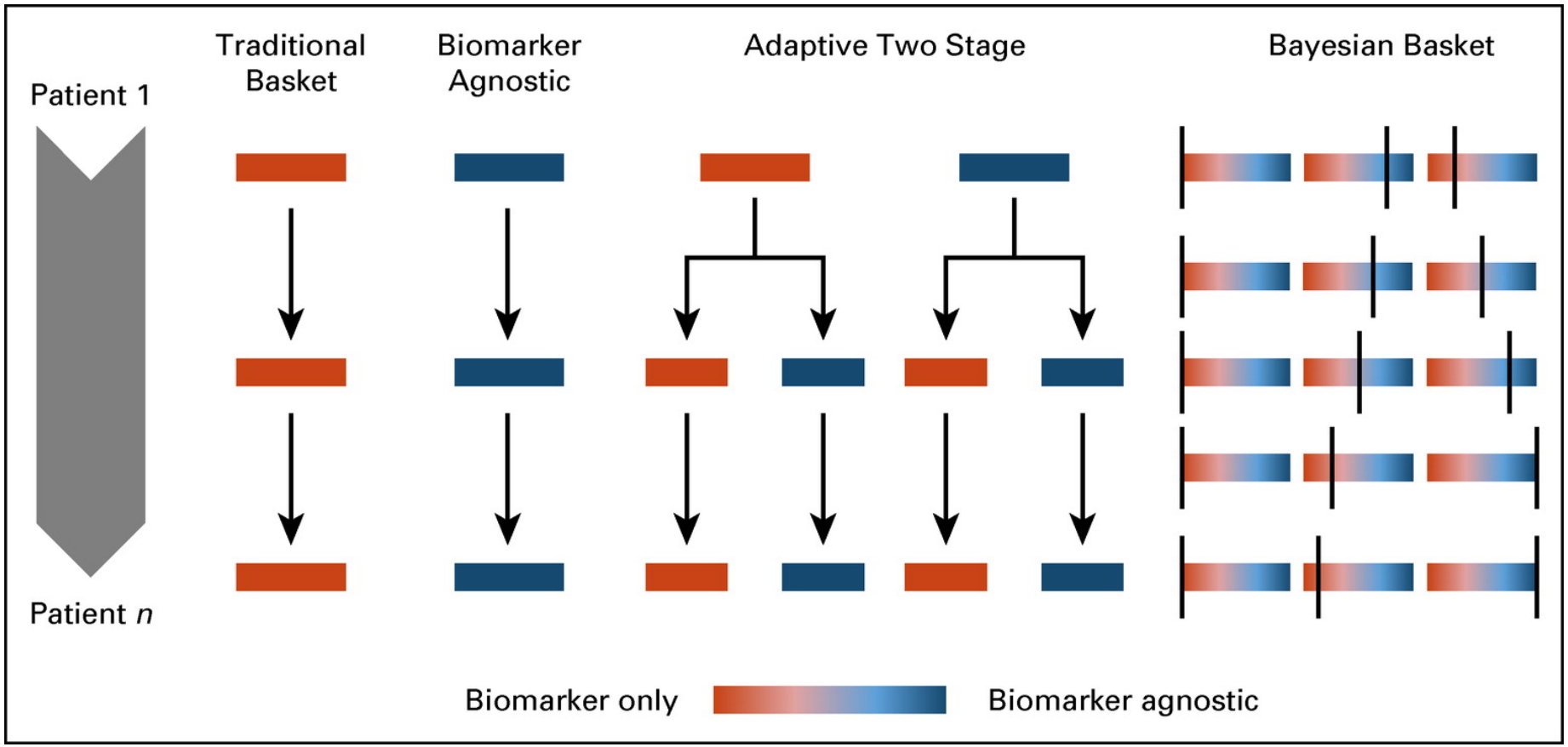
Hyman, et al. (NEJM 373:8, Aug 20, 2015)

- *"BRAF* V600 appears to be a targetable oncogene in some, but not all, nonmelanoma cancers.
- Preliminary vemurafenib activity was observed in non–small-cell lung cancer and in Erdheim–Chester disease and Langerhans'-cell histiocytosis.
- The histologic context is an important determinant of response in *BRAF* V600–mutated cancers."

Aggregation Design (M. Gonen, et al., MSKCC working paper)

- Stage 1
  - Allocate a (modest) number of patients in each basket
  - Evaluate if response rates are heterogenous across baskets to determine if baskets should be treated as independent or aggregated
  - Apply a futility rule to individual baskets or the aggregated basket, depending on assessment of heterogeneity
- Stage 2
  - Continue allocating patients to the aggregated basket and use a one-sample test for efficacy
    - OR
  - Continue allocating patients only to selected baskets and test for efficacy using many one-sample tests, adjusting for multiple testing

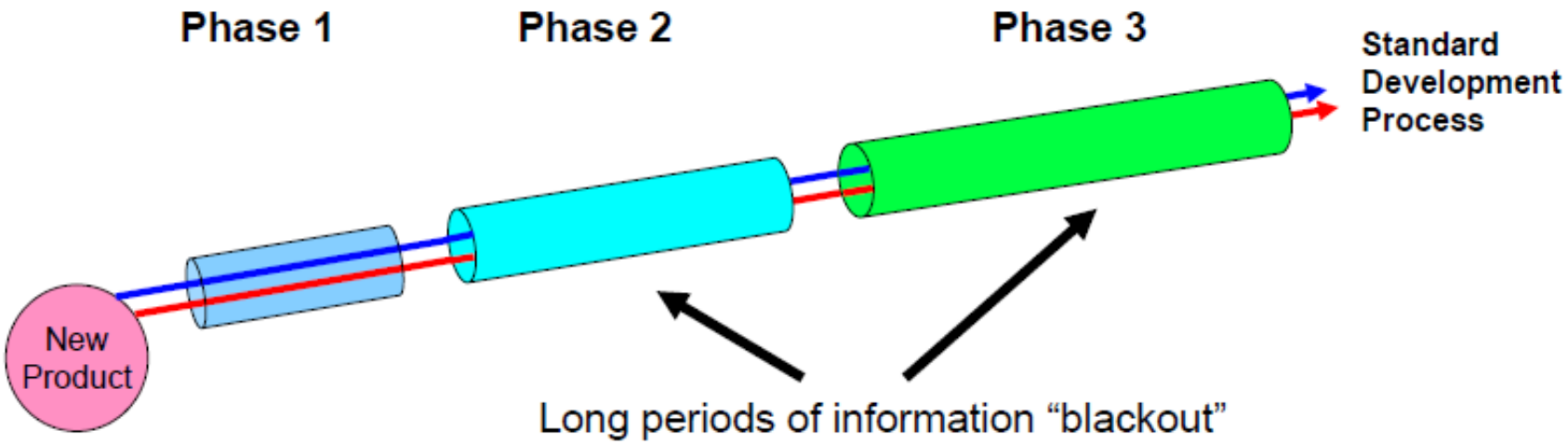
Bayesian Basket Designs, Trippa & Alexander, JCO 2017.



Adaptive, flexible designs

- **Traditional Design:** fix the sample size in advance and only perform one efficacy analysis after all subjects have been enrolled and evaluated.
- **Flexible Design:** monitor the accruing efficacy data at administratively convenient intervals and make important decisions concerning the future course of the study along the way.
- Flexible designs more appealing to pharma than academic trials, but can be useful in both settings.

Traditional Clinical Development Process



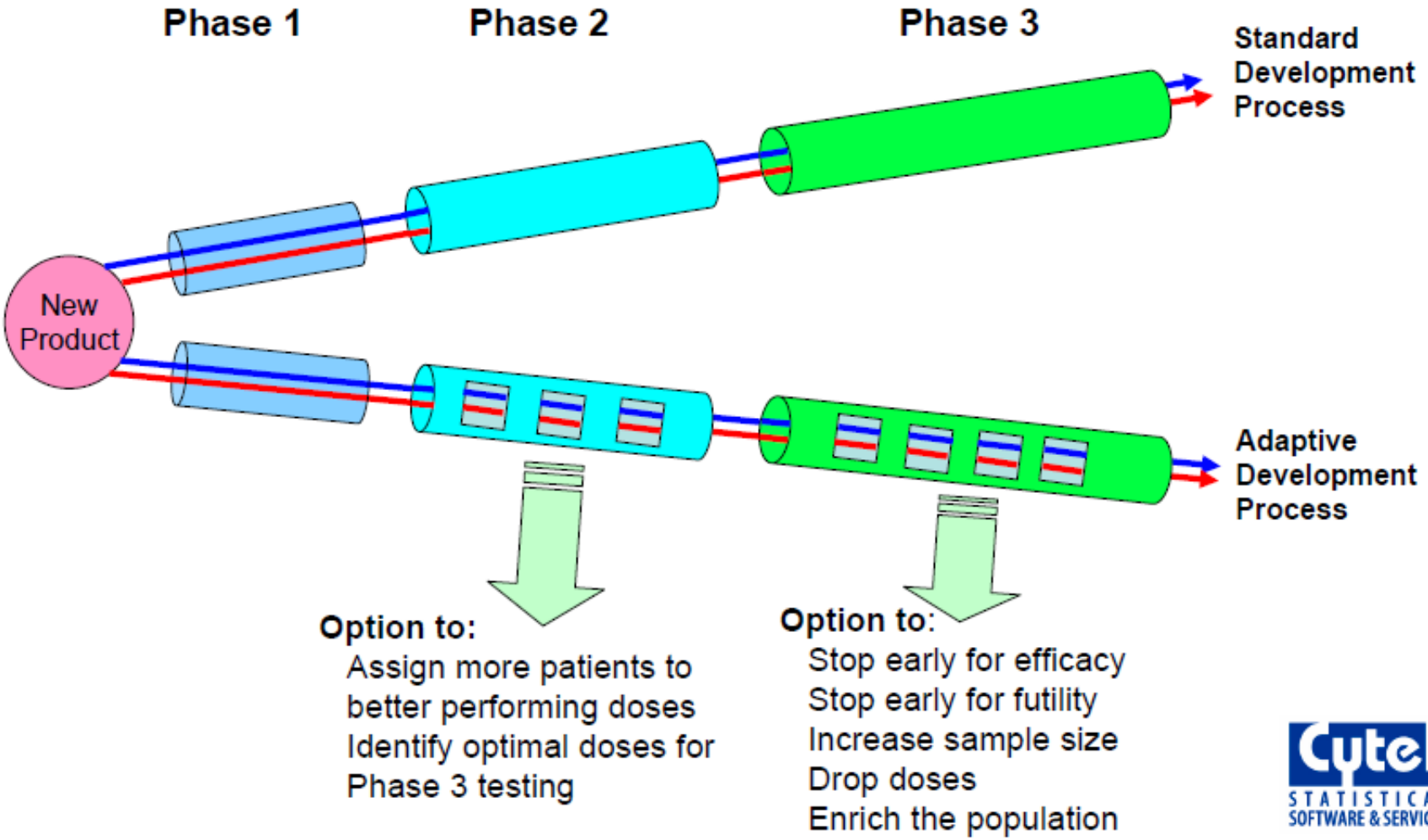
Long periods of information "blackout"
No opportunity to:

- adjust dosing
- make minor adjustments to trial design
- stop for futility

What if we had a "window" into the process to check to see if we are on track?

Cytel STATISTICAL SOFTWARE & SERVICES

Adaptive Clinical Development Process

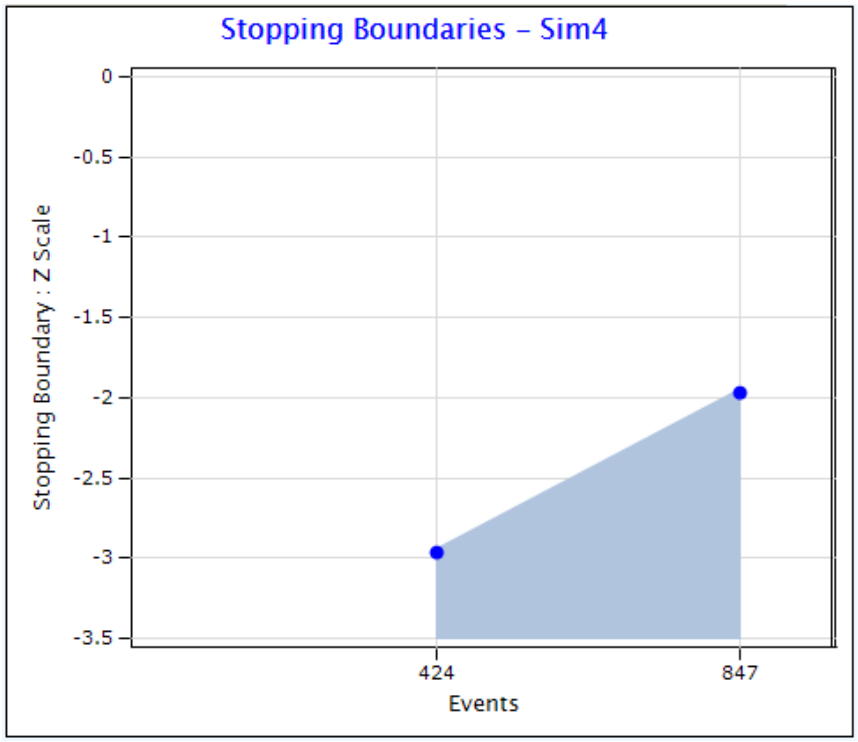


Event Driven Trial to Detect HR=0.75

- Enroll 7000 patients over 4 years; and follow for 1 more year until 530 CV events obtained
- Given 3.25% events/year on placebo, trial has:
  - 91% power to detect HR=0.75 (25% risk reduction)
  - 73% power to detect HR=0.8 (20% risk reduction)
- Possible to recover lost power if HR=0.8?

Option 1: Large Group Sequential Trial

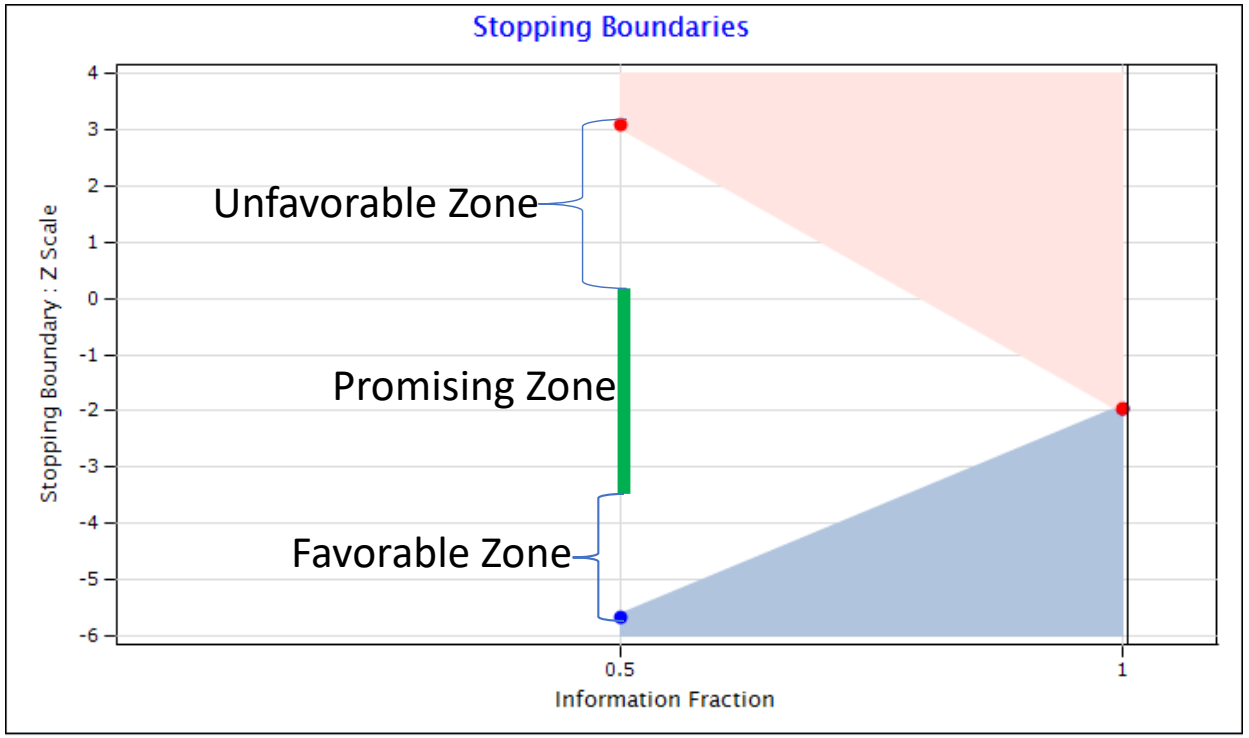
- 11,000 patients and 847 events over 5 years provides 91% power to detect HR=0.8
- Group sequential boundary for early stopping



- Early stopping probability:
  - 27% if HR=0.8
  - 50% if HR=0.75
- **Problem:** Only 1.5 years of average follow-up if early stop

Option 2: Promising Zone Design

- **Start small**: 7000 patients and 530 events
- Increase by 50% if interim results are promising
- Promising zone is defined by **conditional power**



Unfavorable:
CP<50%

Promising:
50%≤CP<90%

Favorable:
CP>90%

Design Optimistically: Assume HR=0.75

Operating Characteristics if in Truth Risk Reduction is 20% (HR=0.8)

Zone	P(Zone)	Power		Sample Size		Study Duration	
		NonAdapt	Adaptive	NonAdapt	Adaptive	NonAdapt	Adaptive
Unfav	24%	38%	38%	7000	7000	4.9	4.9
Prom	36%	**74%**	**90%**	**7000**	**10500**	**4.9**	**5.7**
Fav	40%	95%	95%	7000	7000	4.9	4.9

Operating Characteristics if in Truth Risk Reduction is 25% (HR=0.75)

Zone	P(Zone)	Power		Sample Size		Study Duration	
		NonAdapt	Adaptive	NonAdapt	Adaptive	NonAdapt	Adaptive
Unfav	12%	57%	57%	7000	7000	5	5
Prom	28%	**87%**	**98%**	**7000**	**10500**	**5**	**5.8**
Fav	60%	99%	99%	7000	7000	5	5

Statistical Issues with Adaptive Designs

- For a regulatory filing, must maintain strong control of type-1 error
- Two sources of error inflation in these examples
  - due to unblinded sample size re-estimation
  - due to enrichment by subgroup selection
- Although the trial is expanded only if the IA shows promise of superiority:
  - Actual interim decision should only be conveyed on need to know basis (to drug supply and IVRS teams)
  - Investigators may be told only that this adaptive design has a maximum sample size of xxx patients and possibility of re-powering at IA
- Use a double blind design if possible to avoid operational bias

Operational and Bias Issues

- All design details are included in DMC charter
- DMC buys into design at the kick-off meeting, but reserves right to exercise clinical judgment
- Although the trial is expanded only if the IA shows promise of superiority:
  - Actual interim decision should only be conveyed on need to know basis (to drug supply and IVRS teams)
  - Investigators may be told only that this adaptive design has a maximum sample size of xxx patients and possibility of re-powering at IA
- Use a double blind design if possible to avoid operational bias

Data sharing for complex designs

- As a trial designer, I would like to know what happened in previous trials.
- Information richer than overall summary statistics would be most helpful
  - Patient mix, including correlation among baseline characteristics
  - Time-dependent behavior of the trial
- Help may be on the way

The NEW ENGLAND JOURNAL of MEDICINE

EDITORIALS

Data Sharing Statements for Clinical Trials — A Requirement of the International Committee of Medical Journal Editors

The International Committee of Medical Journal Editors (ICMJE) believes there is an ethical obligation to responsibly share data generated by interventional clinical trials because trial participants have put themselves at risk. In January 2016 we published a proposal aimed at helping to create an environment in which the sharing of explained at www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html. If the data sharing plan changes after registration this should be reflected in the statement submitted and published with the manuscript, and updated in the registry record.

June 8, 2017

ICMJE Policy (NEJM June 8, 2017)

- “1. As of July 1, 2018, manuscripts submitted to ICMJE journals that report the results of clinical trials must contain a data sharing statement as described below.
- 2. Clinical trials that begin enrolling participants on or after January 1, 2019, must include a data sharing plan in the trial’s registration. The ICMJE’s policy regarding trial registration is explained at www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html. “