

Designing Cost Efficient Phase II Trials with the Goal of Getting to Phase III: Alpha? Power? Irrelevant

Karen L. Kesler and Ronald W. Helms
Rho, Inc.

Contact: kkesler@RhoWorld.com

Introduction

- Approval Process—Phase I, II, & III
- Phase II goals:
 - Profile safety.
 - Choose the most appropriate outcome.
 - Choose the most efficacious dose or regimen.
 - Eliminate ineffective compounds quickly.
- In general, get enough information to get to Phase III quickly.

Statistics Supports the Clinical Goals

- These goals all have implications for the statistical methods chosen and especially the properties of the trial design and analysis.
- How do we maintain sound statistical practices that guarantee correct interpretation of the trial in a more efficient trial framework?
- Or, how do we use statistical methods to “get to Phase III or kill the compound more quickly”?

Statistics Supports the Clinical Goals

What statistical methods can help us achieve our clinical goals?

- Clinical Goals:
 - More flexible design due to higher levels of uncertainty.
 - Ability to detect clinically relevant scenarios for the doses or regimens.
 - Ability to stop enrolling to specific treatment arms or the whole study quickly for safety reasons.
 - Ability to gather a wider range of information (in doses and populations) without unduly increasing the sample size.

Statistics Supports the Clinical Goals

What statistical methods can help us achieve our clinical goals?

- Statistical Methods:
 - Group sequential designs instead of fixed designs.
 - Measure our potential for mistakes using trial-wise error instead of hypothesis-wise error.
 - Design properties generated from simulations of clinically relevant scenarios.
- Introduce an adaptive design and then discuss the statistical methodology underlying the design and how it creates a more efficient trial framework.

Example Adaptive Design

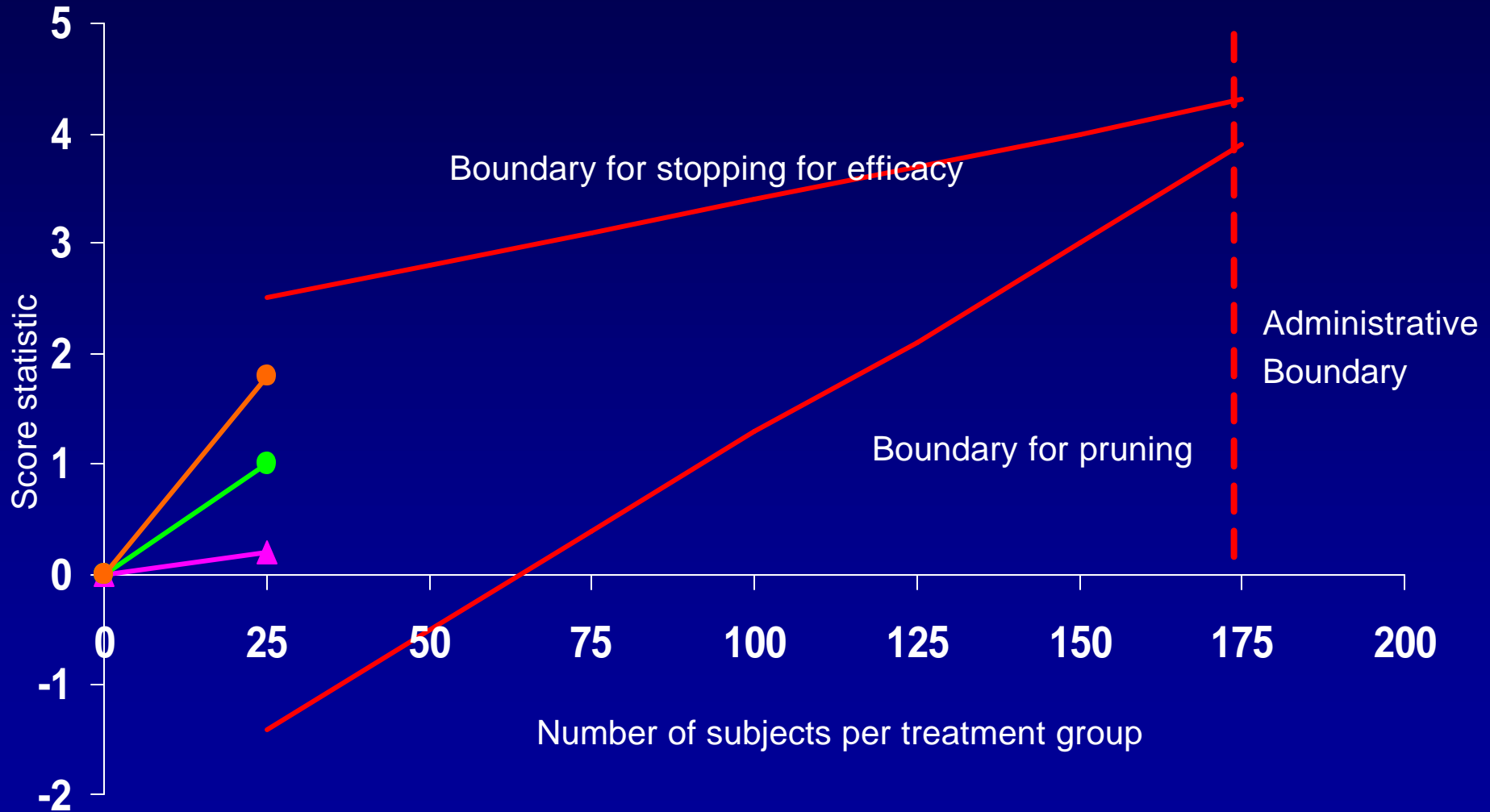
- Design a Phase II study with enough arms to cover all plausible regimens, plus placebo.
 - K possible doses $\Rightarrow K+1$ arms
- Conduct a sequence of frequent interim analyses (using group sequential methods).
- At each interim analysis, potentially prune treatment arms, except control.
- Ultimately, at some interim analysis, the treatment arms are reduced to one active treatment regimen and the placebo.

Example Adaptive Design: Setup

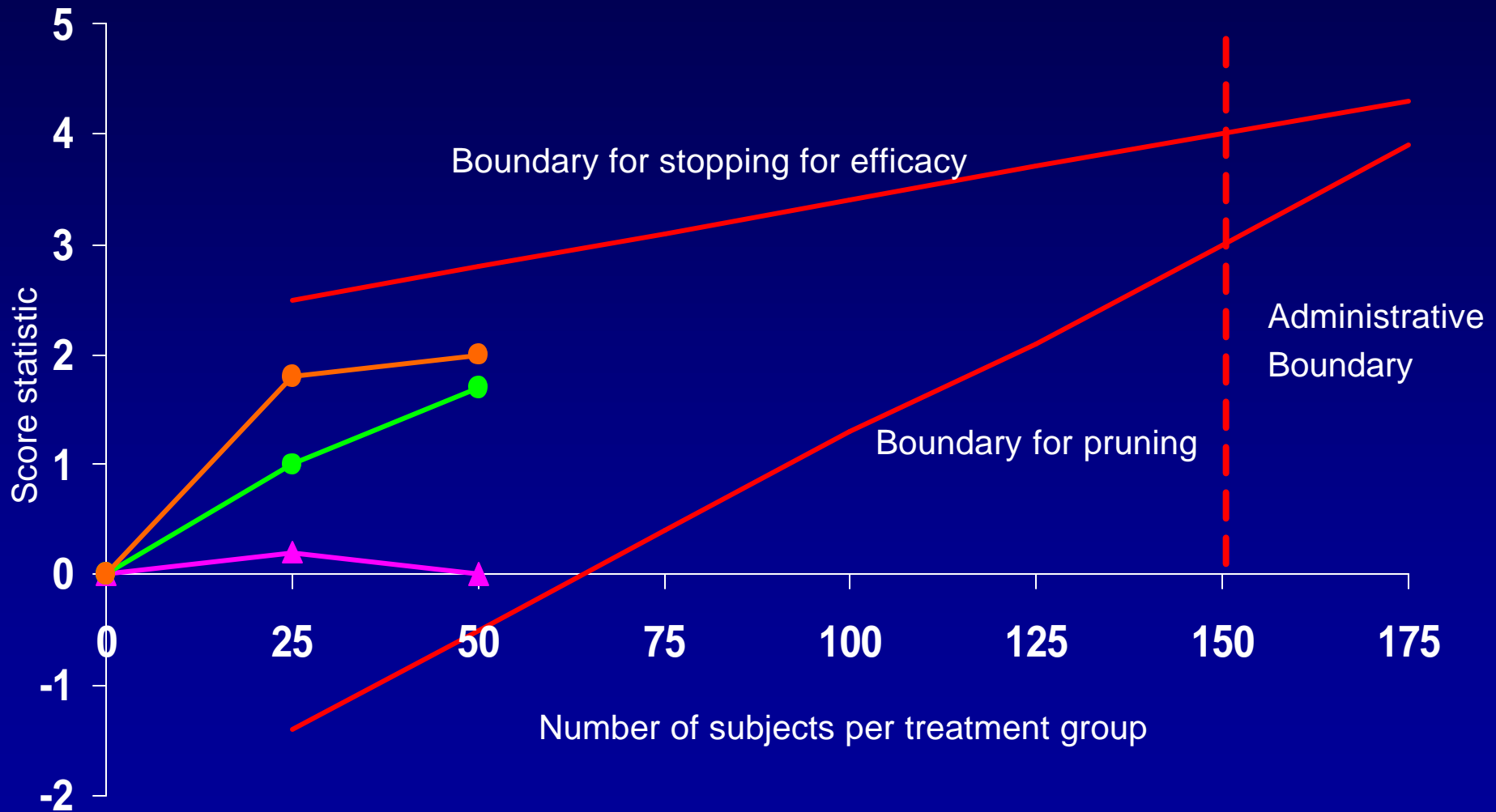
Hypothetical Trial Setup

- Therapeutic area: Skin infection
- Design features: double blind, randomized, placebo controlled, multicenter, baseline, final assessments at 4 weeks.
- Outcome: Patients with infection cured. (Log odds vs. Control)
- Safety profile (from Phase I data): mild but potentially unpleasant side effects (e.g., rash).
- Four dose levels to be considered: Placebo, Low, Medium, and High

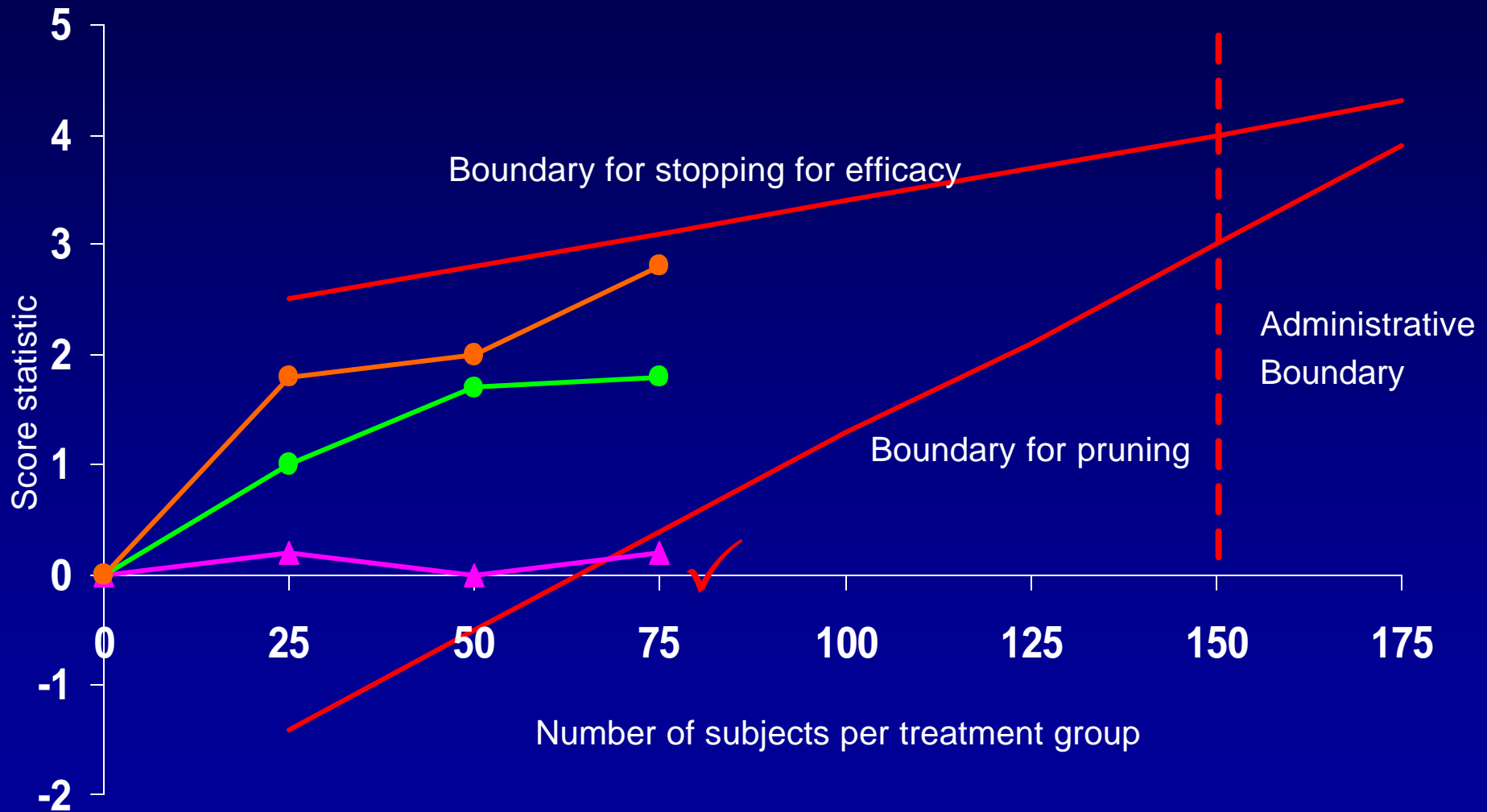
Example Adaptive Design: First Interim



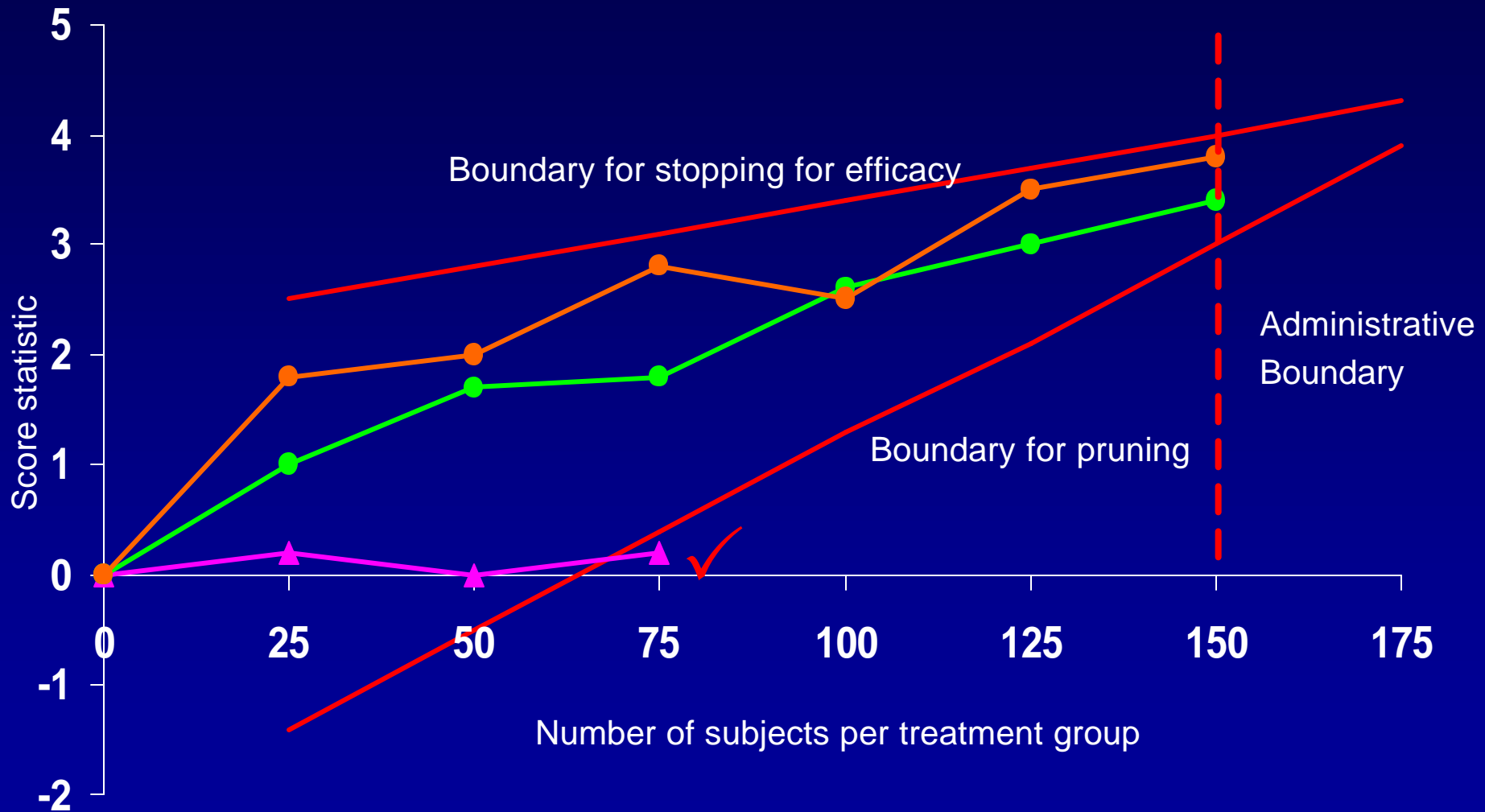
Example Adaptive Design: Second Interim



Example Adaptive Design: Third Interim



Example Adaptive Design: Administrative Stop



Example Adaptive Design: Review of Methods

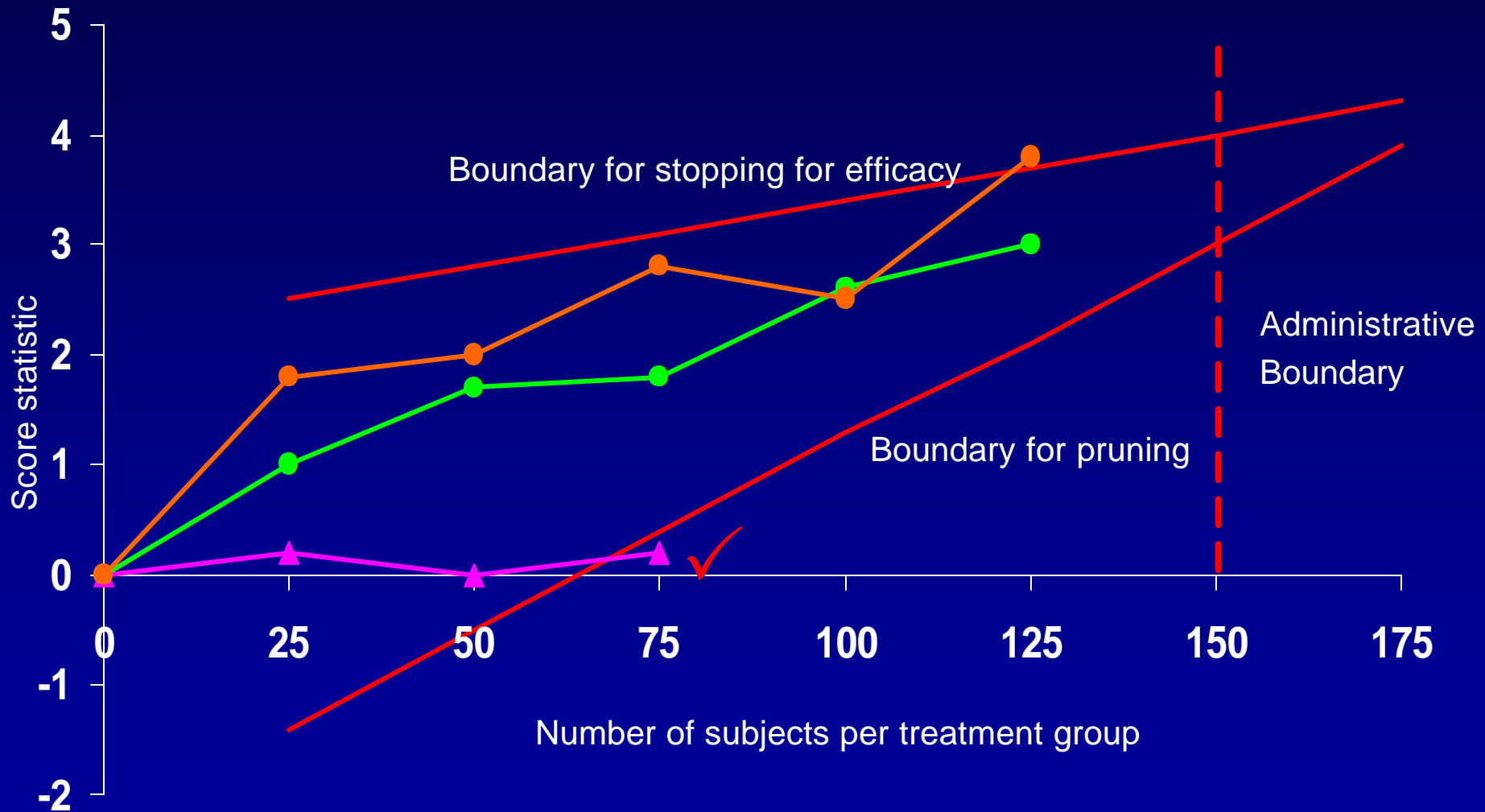
- Now that you've seen the concept, recall the methods we're using:
 - Group sequential methods
 - Trial-wise error instead of hypothesis-wise error
 - Simulation-based measurement of properties

Let's look at some specifics. . .

Group Sequential Trials: What is it?

- A group sequential design makes use of interim analyses to assess the data as the trial progresses and make a decisions to stop early for efficacy or futility.
- Test statistics are compared to predetermined futility and efficacy boundaries.
- If we had crossed an efficacy boundary instead. . .

Group Sequential Methods: Crossing the Efficacy Boundary



Group Sequential Trials: What do we gain?

- Flexibility to stop the trial.
 - If we cross an efficacy boundary, we can stop early.
 - Can see if the compound is not effective and stop the trial.
- Smaller expected sample size with more information.
 - Group sequential trials have a smaller sample size on average than their corresponding fixed design counterparts, so we can have more information on a range of doses.
- Planning time.
 - Interim analyses allow a focused look at the results as the trial progresses.

Measuring Error

- We have more flexibility, but we can go farther by co-opting the concepts of statistical error to better reflect our clinical goals.
- Recall, at Phase II, we want to:
 - Pick the best dose if the compound is effective.
 - Kill it quickly if it's not.
- So, what constitutes an “error”?
 - Given an ineffective compound, we conclude it is effective.
 - Given an effective compound, we conclude that it is ineffective.

Measuring Error: How do we do it?

- How can we measure these errors in our adaptive design?
- Picking an ineffective dose/regimen
 - The rate of an ineffective dose crossing the efficacy boundary.
 - The rate of an ineffective dose having highest test statistic at the administrative boundary.
- Missing an effective dose/regimen
 - The rate of an effective dose crossing the futility boundary.
 - The rate of an effective dose being not chosen at administrative boundary.

Measuring Error: Simulations

- How do we observe these rates when we only get to see one trial's outcome?
- Use simulations to predict the behavior of the design under different situations.
- Allows us to measure the properties of the design, much like performing power calculations for a typical fixed sample design.

Measuring Error: Simulation Specifics

- An overview of the process:
 - We then assume outcomes for all of the treatment arms, creating a “scenario.”
 - Using the most plausible scenario, we create our group sequential boundaries for the trial.
 - We randomly create patient data under the scenario assumptions and measure the outcome of our trial.
 - We repeat an appropriate number of times and see how often we came to the “right” conclusion—according to the scenario assumptions.
 - Change the scenario and start again.

Measuring Error: Scenarios

Hypothetical Trial Setup: Treatment Effect

- Three doses: Low, Medium, and High
- Placebo is always 0%

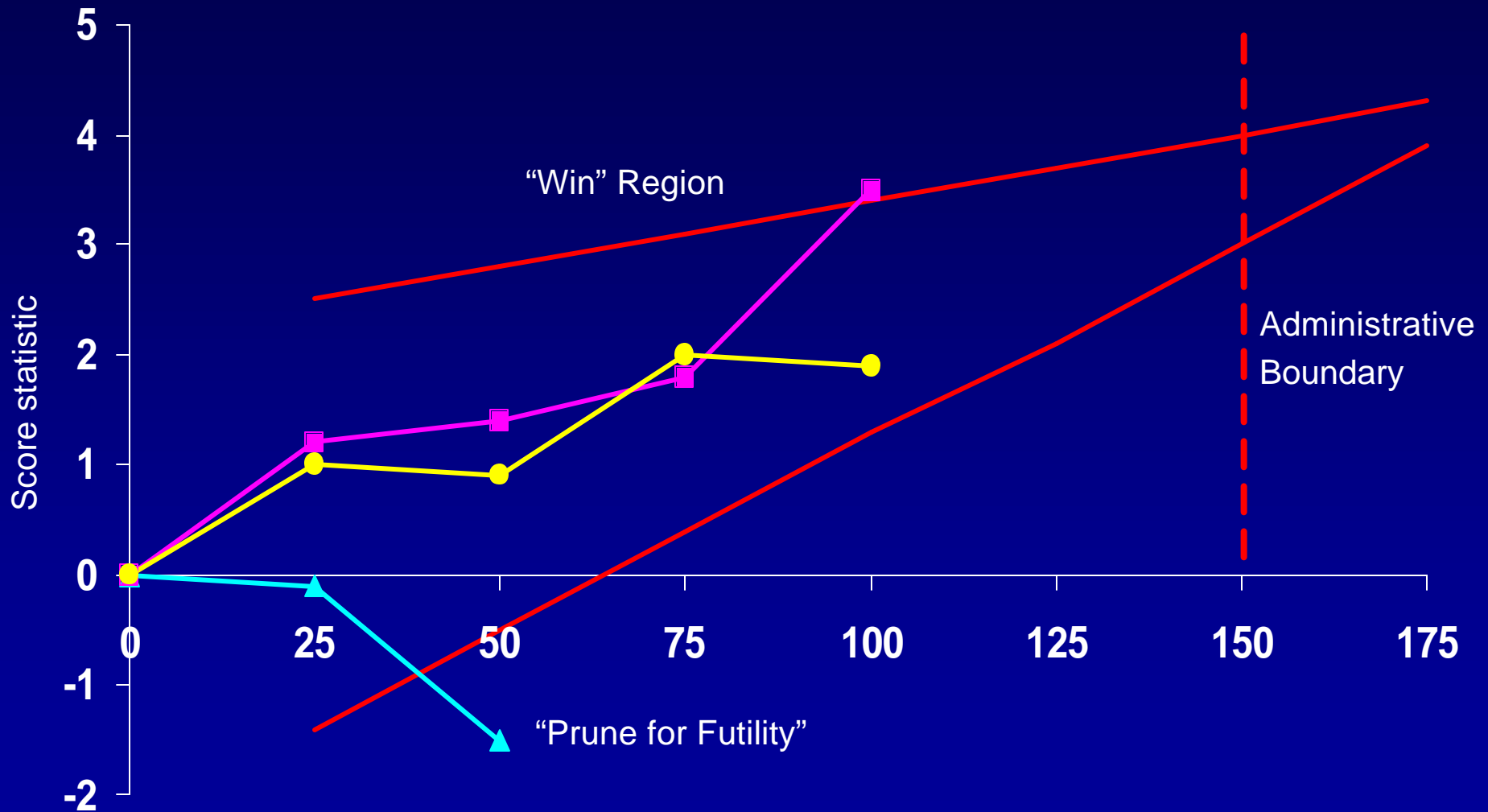
<i>Scenario</i>	Low	Medium	High
A: Linear	0%	30%	60%
B: Linear 2	25%	42.5%	60%
C: Plateau	25%	60%	60%
D: Early Plateau	60%	60%	60%
E: No effect	0%	0%	0%

Measuring Error: Specifics

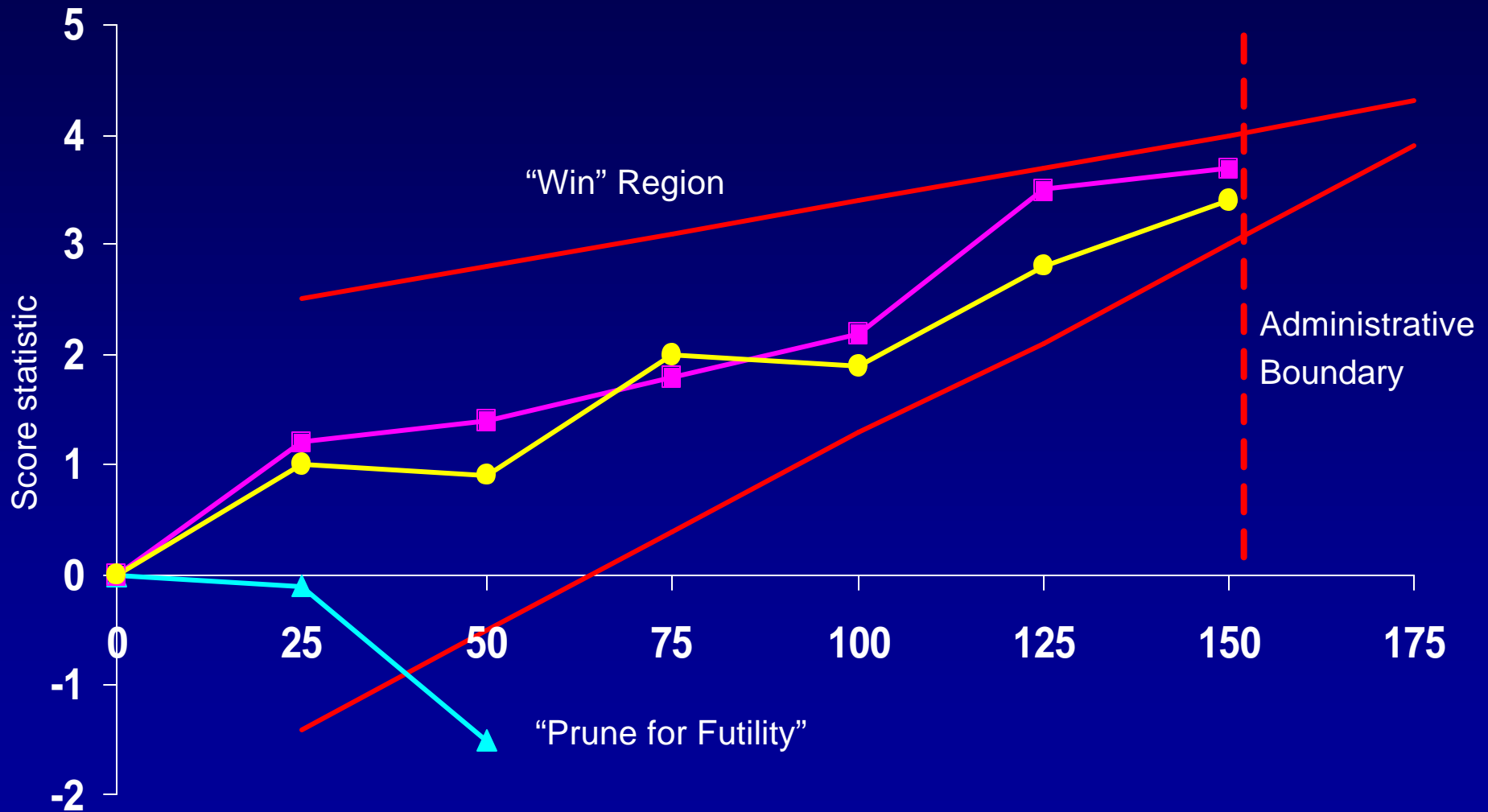
What are we measuring?

- “Wins” for each dose
 - Cross the efficacy boundary
 - Have the highest statistic at the administrative boundary
- “Crossing the Boundary” for each dose
 - Futility
 - Efficacy
- Picking an effective dose over the whole trial.

Measuring Error: “Wins” and “Prunes”



Measuring Error: “Wins” and “Prunes”



Measuring Error: Simulation Results

- Take the “Linear” scenario:
 - Goals: Cut low dose for futility (34%) and select the high dose (89%).

			Cross a Boundary	
Dose	Effect Size	% Wins	% Efficacy	% Futility
Low	0%	1%	<1%	34%
Med	30%	10%	3%	9%
High	60%	89%	17%	1%

Measuring Error: Reducing Error

- Minimize “Wins” for ineffective (Low) dose.
- Maximize “Cross Futility Boundary” for ineffective (Low) dose.
- Maximize “Cross Efficacy Boundary” and “Wins” for most effective (High) dose.
- Make sure administrative boundary is sufficient to gather enough information.
- Confirm measurements across all the scenarios.

Measuring Error: Background

How is this different than the traditional hypothesis-wise error?

- Type I Error (alpha)
 - “Rejecting the null hypothesis when it is true.”
 - Deciding the dose is effective when it is not. (Backing a “Loser”.)
- Type II Error (beta or 1-Power)
 - “Accepting the null hypothesis when it is false.”
 - Deciding an efficacious compound has no effect. (Missed opportunity--benefits to patients, profits, etc.)

Measuring Error: What do we gain?

- Better suited to assessing the exploratory paradigm.
 - Considers more than one hypothesis.
- A more realistic picture of the types of errors we could make.
- Simulations provide a more intuitive measure of the potential errors under more clinically relevant assumptions.
- Helps avoid the danger of concluding a compound is efficacious, without inflating the sample size to unfeasible proportions.

Adaptive Designs: What do we gain?

- Flexibility
 - Stop trial early if compound not effective.
 - Stop arms early if not effective.
- More “bang for your buck”
 - More information on dose-response curve
 - Fewer subjects than a fixed design with same number of treatment arms
- Time
 - Reduces time to Phase III by eliminating “down time” between trials.
 - By monitoring the results as they progress, planning can be done in parallel to the trial.

Extensions

- Our example only had four treatment arms, but if there's a large range for the dose, more arms can be included easily—think big, prune often.
- Even more time can be saved in the overall drug development scheme if the Phase III trial can be started directly after the Phase II trial ends.
 - Continue with the last dose chosen (and placebo).
 - Sites continue with enrollment and momentum is maintained.

Cautions

- This is *not* appropriate to use in a confirmatory trial.
 - Designed for an exploratory approach.
- Only appropriate for certain situations, you need relatively few patients enrolled between taking a snapshot of the database for an interim analysis and the meeting to make decisions based on the interim analysis. Implies:
 - Fast data capture, processing.
 - Centralized Randomization.
 - At the start of Phase II, one can specify a set of treatment regimens that will very likely include the regimen(s) to be tested in Phase III.

Conclusions

- We can use current statistical methods to improve efficiency and relevancy of Phase II exploratory trials *without undermining sound statistical principles*.
- Specifically, we can:
 - Gather more information on more doses without huge increases in sample size.
 - Estimate sample size based on clinically relevant scenarios with a view towards exploring the data.
 - Consider the goals of the trial, not just a single hypothesis when evaluating design and power.

A version of this presentation is available
(PDF) at: www.RhoWorld.com

