An Interim Analysis Example

Mark A. Weaver, PhD
Family Health International

Office of AIDS Research, NIH
ICSSC, FHI
Goa, India, September 2009

Example

- We are designing an HIV vaccine trial
  - HIV negative, high-risk women will be randomized to either vaccine or placebo
- We desire 90% power to detect a 50% reduction in HIV incidence (HR = 0.50), 2-sided $\alpha = 0.05$
  - Tomorrow, we will see that power depends on number of events (HIV infections)
  - The events provide all of the "information" for test
  - For now, just know that trial requires at least 88 events

Analysis Plan

- Primary efficacy analysis:

  H_0: HR = 1 vs. H_A: HR ≠ 1

- Plan to test this using Cox PH model

  $h(t) = h_0(t) \exp\{ \beta * \text{Trt} \}$

  where Trt = 1 for vaccine, 0 for placebo

  $\beta = \log(\text{HR})$

  ($\beta < 0$ for HR < 1, $\beta > 0$ for HR > 1)

- Test statistic: $Z = \dfrac{\hat{\beta}}{\text{SE}(\hat{\beta})}$

3

Plan for Early Stopping

- Desire to stop the trial early in the event of:

1. clear early evidence of efficacy
2. clear early evidence of harm (wrt HIV)
3. early evidence of harm based on additional safety data (AEs, labs, vital signs, etc.)

- Also might decide to stop for futility
  - What is “futility”?
- Could also stop for poor trial quality, slow enrollment, high loss, etc.

4

"No single statistical test or monitoring procedure ought to be used as a strict rule for decision-making, but rather as one piece of evidence to be integrated with other evidence."

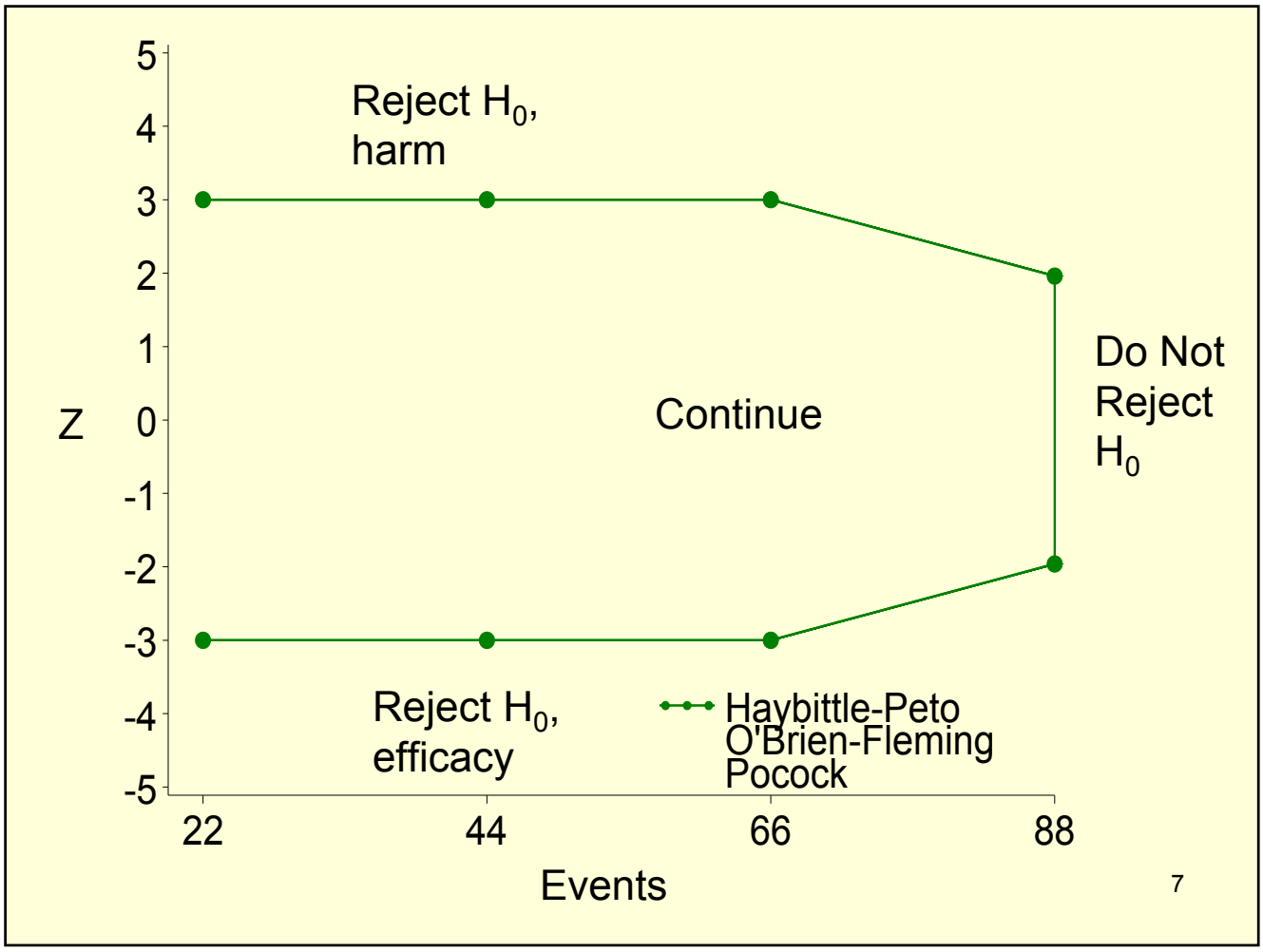
Friedman, Furberg, and DeMets

Fundamentals of Clinical Trials, 3rd ed.

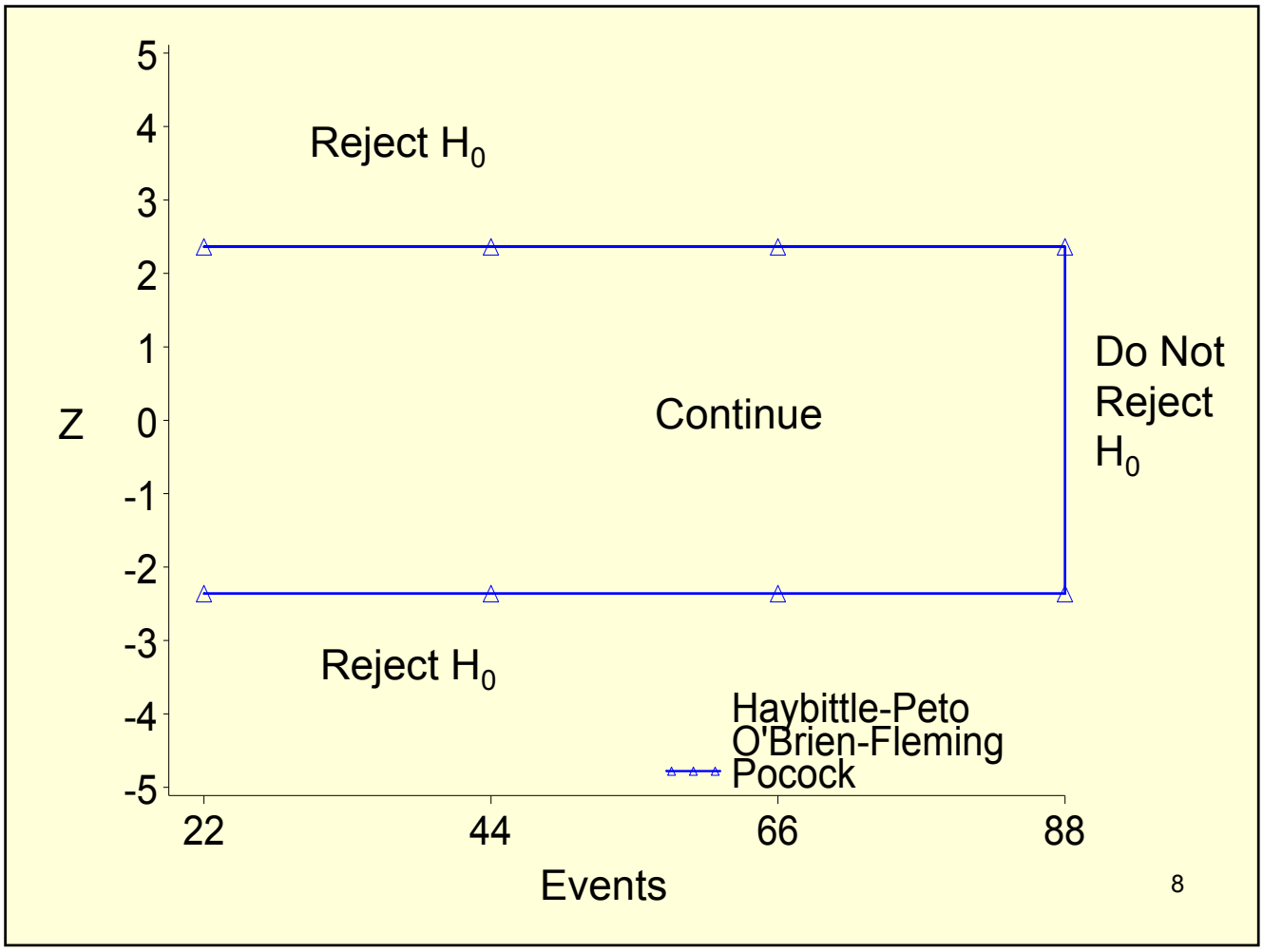
Planning the Interim Analyses

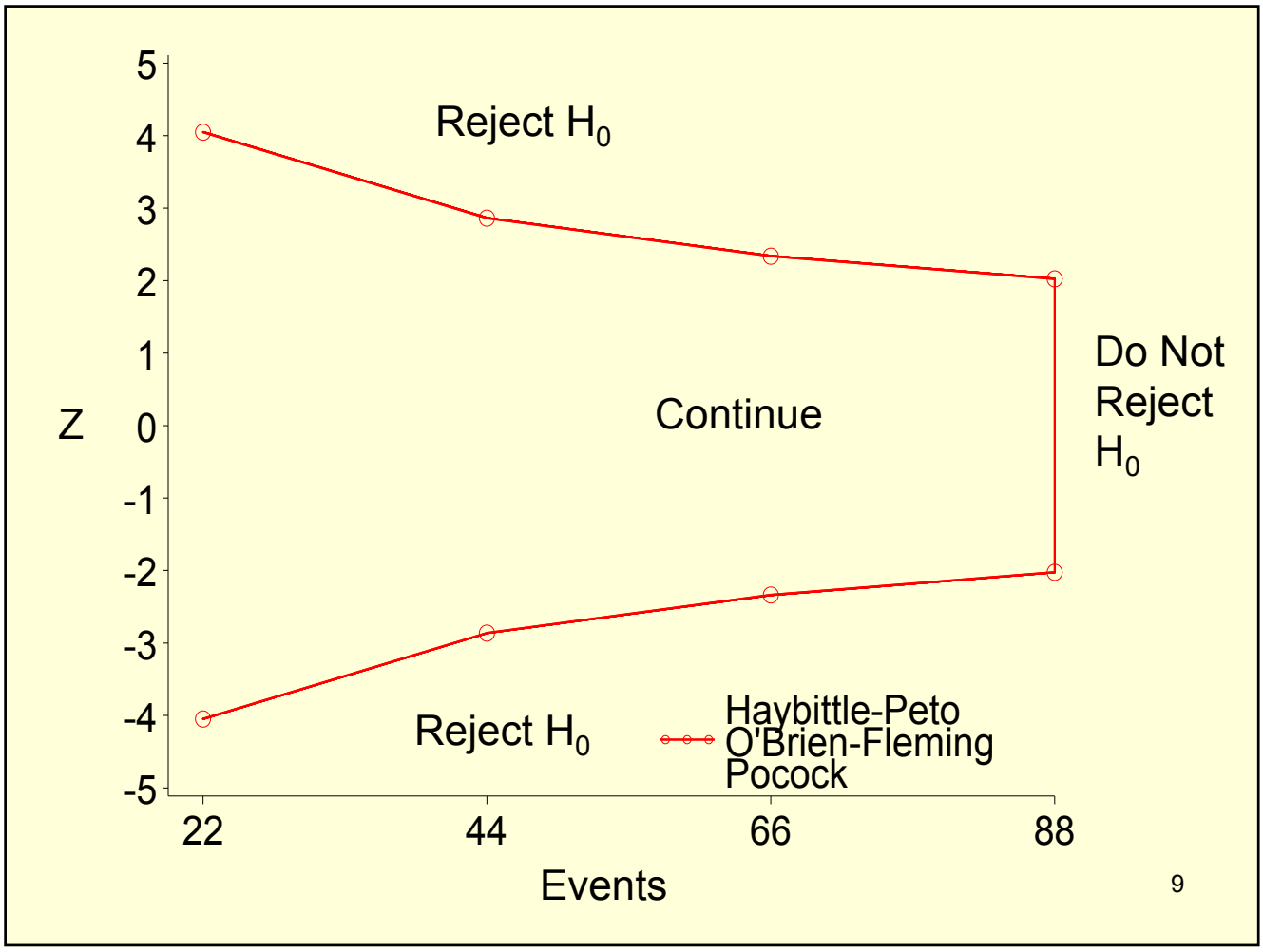
- Plan to test primary hypothesis at 4 points, 3 interim analyses and 1 final analysis
  - After 22, 44, 66, and 88 events
- Goal: maintain the overall probability of a type 1 error (i.e., α) at specified 0.05 level
- Choose a "group sequential" stopping boundary to control α
  1. Haybittle-Peto
  2. Pocock
  3. O'Brien-Fleming

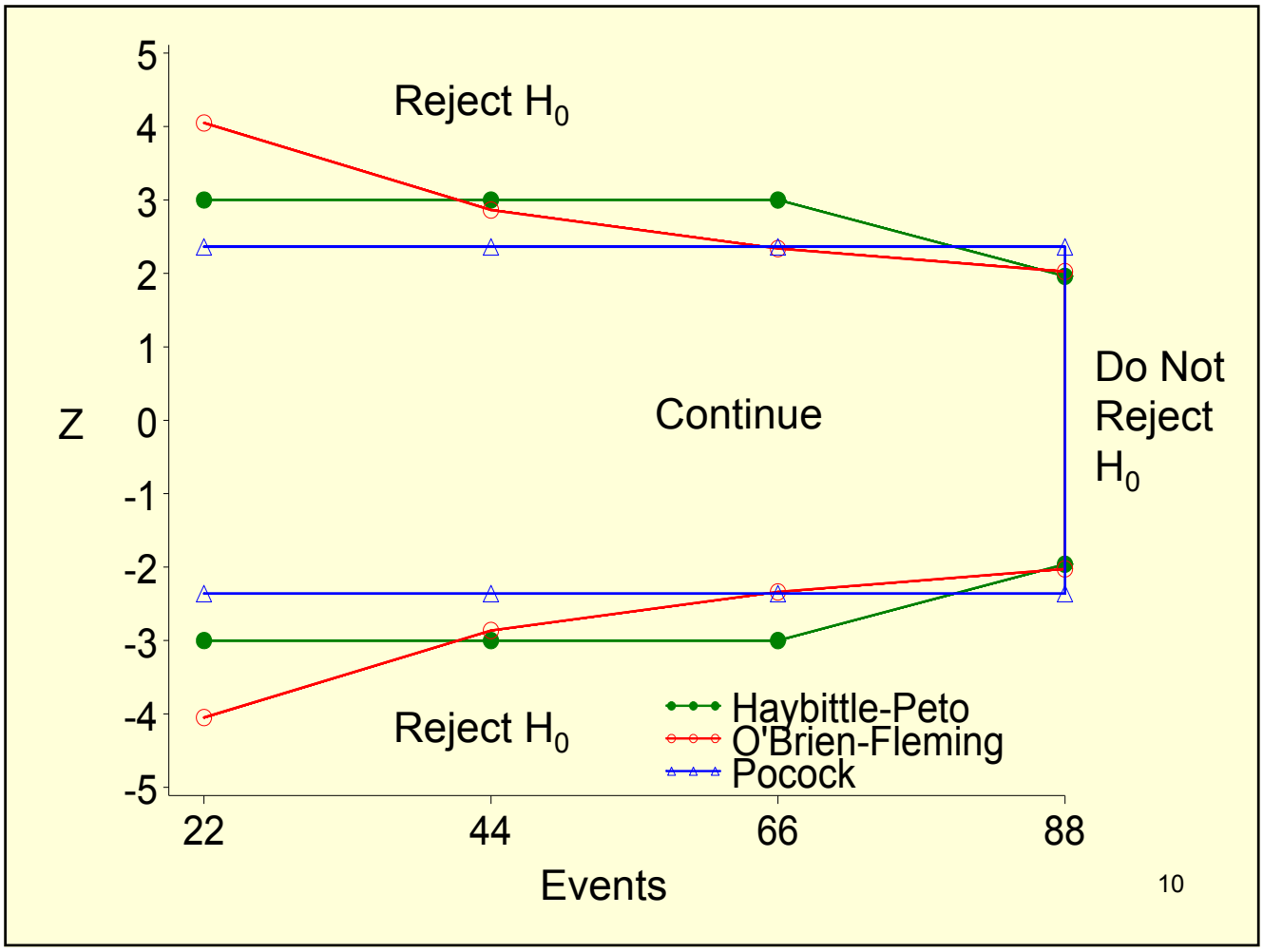
6


Reject H_0, harm
Do Not Reject H_0
Continue
Reject H_0, efficacy
Haybittle-Peto
O'Brien-Fleming
Pocock
Z
5
4
3
2
1
0
-1
-2
-3
-4
-5
22
44
66
88
Events

7


5
4
3
2
1
0
-1
-2
-3
-4
-5
Z
Reject H_0
Continue
Do Not Reject H_0
Reject H_0
Haybittle-Peto
O'Brien-Fleming
Pocock
22
44
66
88
Events
8

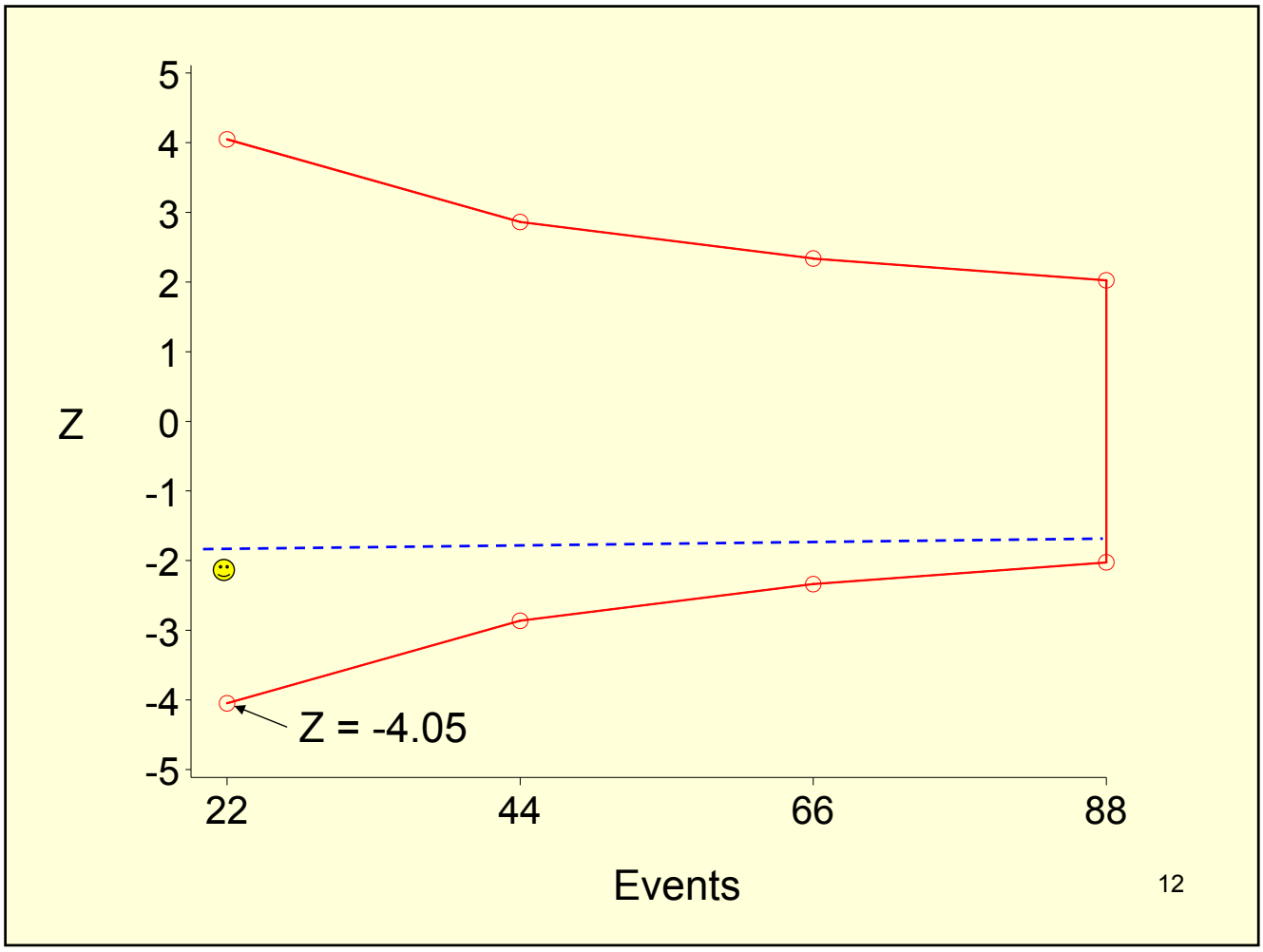

Reject H_0
Continue
Do Not Reject H_0
Reject H_0
Haybittle-Peto
O'Brien-Fleming
Pocock
Z
5
4
3
2
1
0
-1
-2
-3
-4
-5
22
44
66
88
Events
9


Reject H_0
Continue
Do Not Reject H_0
Reject H_0
Haybittle-Peto
O'Brien-Fleming
Pocock
Z
5
4
3
2
1
0
-1
-2
-3
-4
-5
22
44
66
88
Events
10

Monitoring the Trial

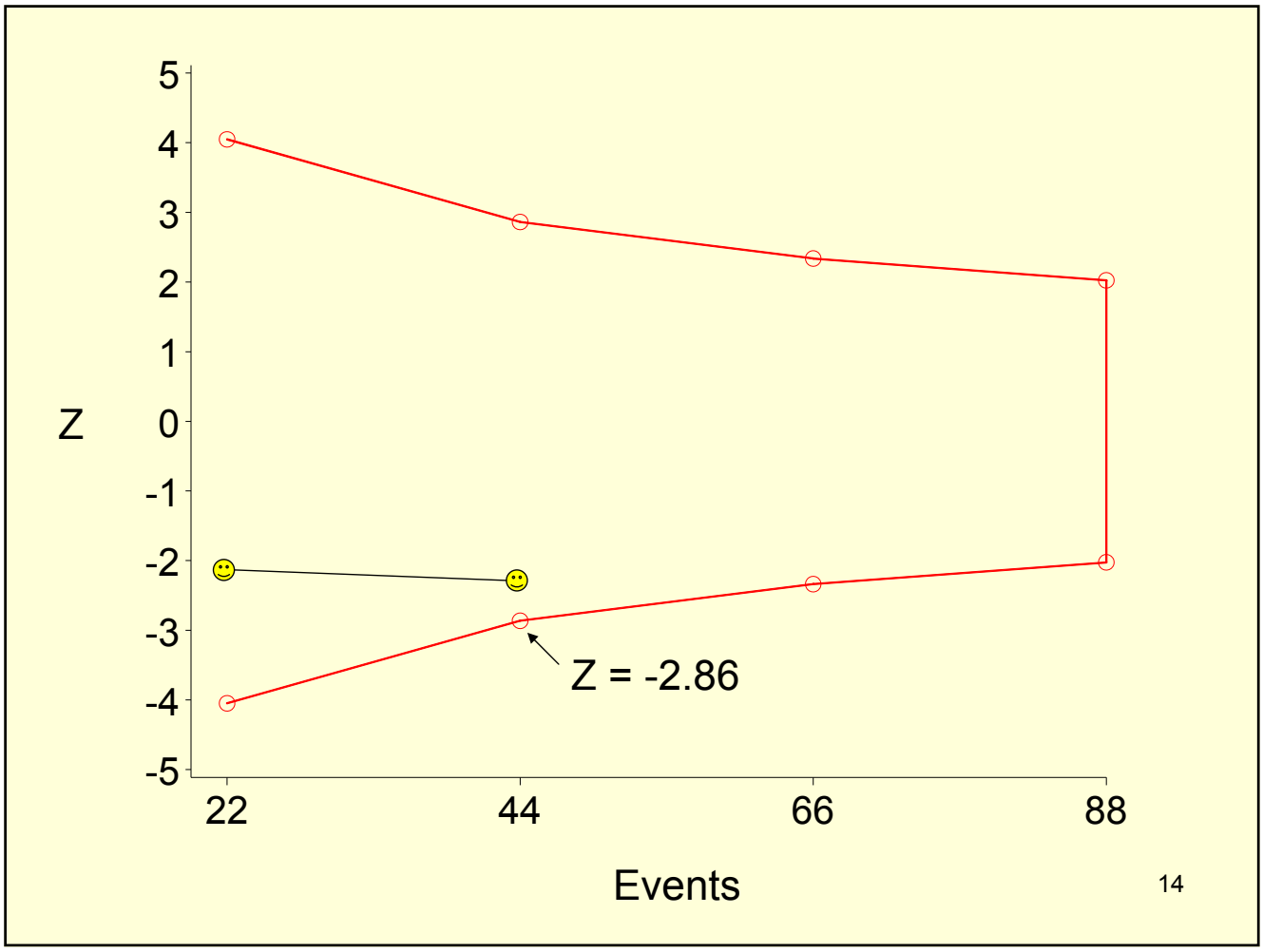
- Suppose the trial started enrolling and following patients on January 1, 2008.
- On June 15, 2008, event number 22 detected
- Data were locked, interim analysis run, DSMB meeting held following week
- 6 events in vaccine arm, 16 in placebo arm
  - Estimated HR = 0.38
  - $\beta = -0.97$, $SE(\beta) = 0.48$
  - $Z = -0.97 / 0.48 = -2.02 \rightarrow$ unadjusted p = 0.0434
  - Looking good, but can we stop at this point?

11


5
4
3
2
1
0
-1
-2
-3
-4
-5
Z
Z = -4.05
22
44
66
88
Events
12

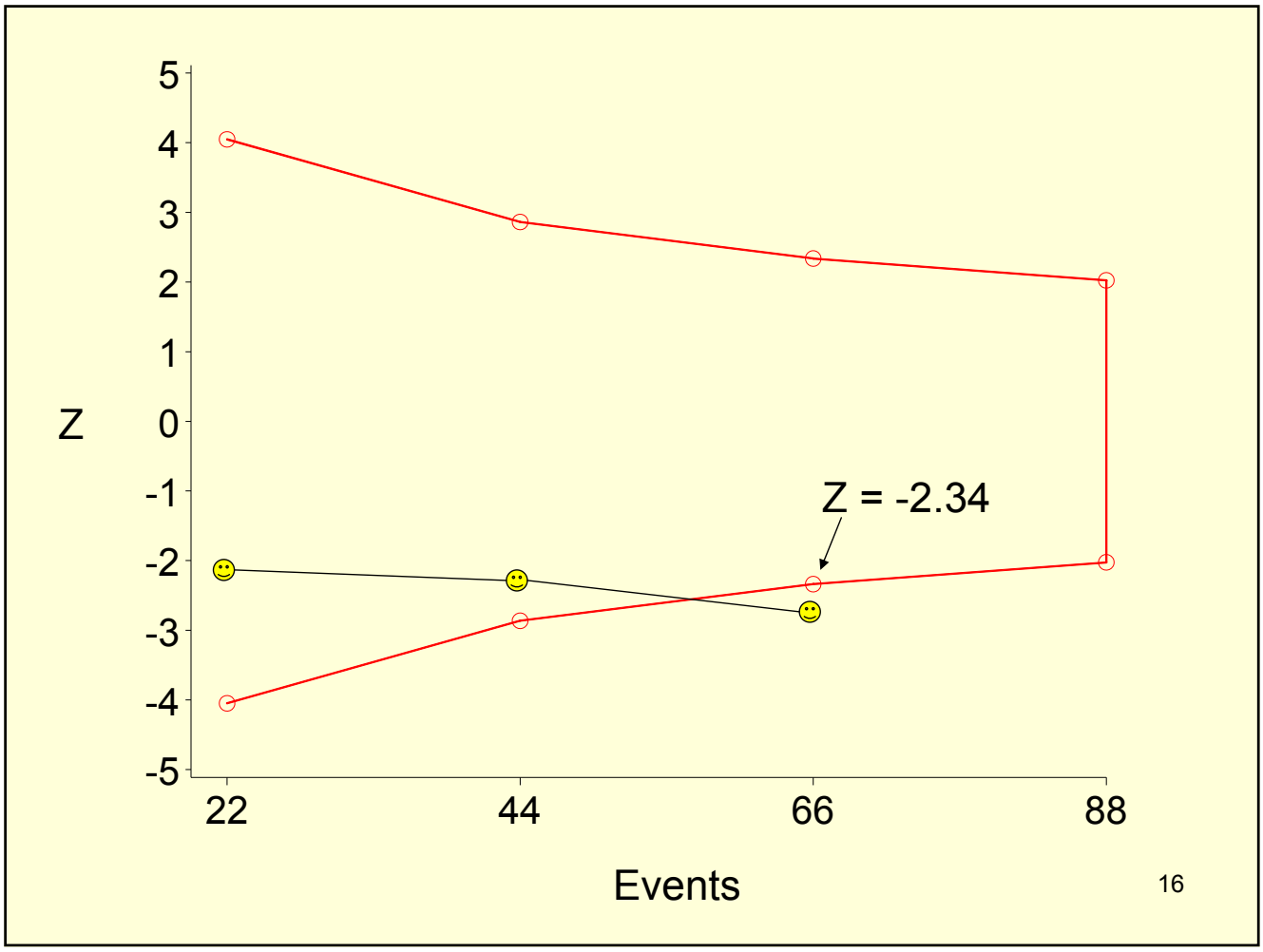
Monitoring the Trial

- On January 13, 2009, event number 44 detected
- Data were locked, interim analysis run, DSMB meeting held following week
- 14 events in vaccine arm, 30 in placebo arm
  - Estimated HR = 0.47
  - $\beta = -0.76$, $SE(\beta) = 0.32$
  - $Z = -0.76 / 0.32 = -2.38 \rightarrow$ unadjusted p = 0.0174
  - Wow, that's really "significant" … but can we stop at this point?


5
4
3
2
1
0
-1
-2
-3
-4
-5
Z
22
44
66
88
Events
Z = -2.86
14

Monitoring the Trial

- On May 31, 2009, event number 66 detected
- Data were locked, interim analysis run, DSMB meeting held following week
- 21 events in vaccine arm, 45 in placebo arm
  - Estimated HR = 0.47 (same as before)
  - $\beta = -0.76$, $SE(\beta) = 0.26$
  - $Z = -0.76 / 0.26 = -2.92 \rightarrow$ unadjusted p = 0.0036
  - Wow! … can we stop now?


5
4
3
2
1
0
-1
-2
-3
-4
-5
Z
Z = -2.34
22
44
66
88
Events
16

SUCCESS!!

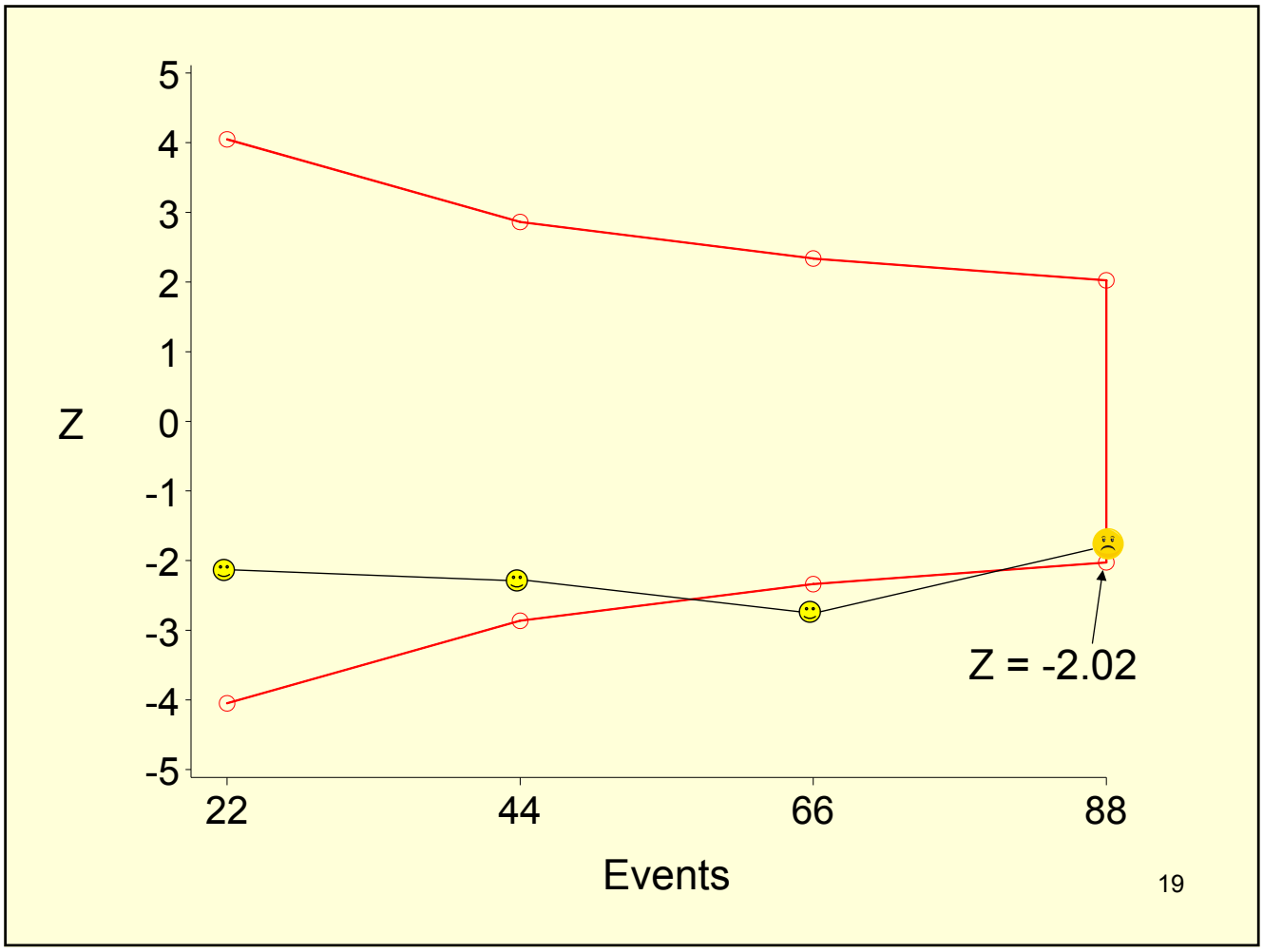
But, what else should a DSMB consider prior to recommending that the trial stop?

17

Monitoring the Trial

- DSMB noted increased incidence of troubling AE (severe apparent allergic reactions) in vaccine arm
  - Due to a bad batch, possibly?
- DSMB recommends to not stop the trial
  - Note, this is all that the project team would get to know!
- Trial proceeded to completion (88 events)
- 35 events in vaccine arm, 53 in placebo arm
  - Estimated HR =0.66
  - $\beta = -0.42$, $SE(\beta) = 0.21$
  - $Z = -0.42 / 0.21 = -2.00 \rightarrow$ unadjusted p = 0.0456

18


5
4
3
2
1
0
-1
-2
-3
-4
-5
Z
22
44
66
88
Events
Z = -2.02
19

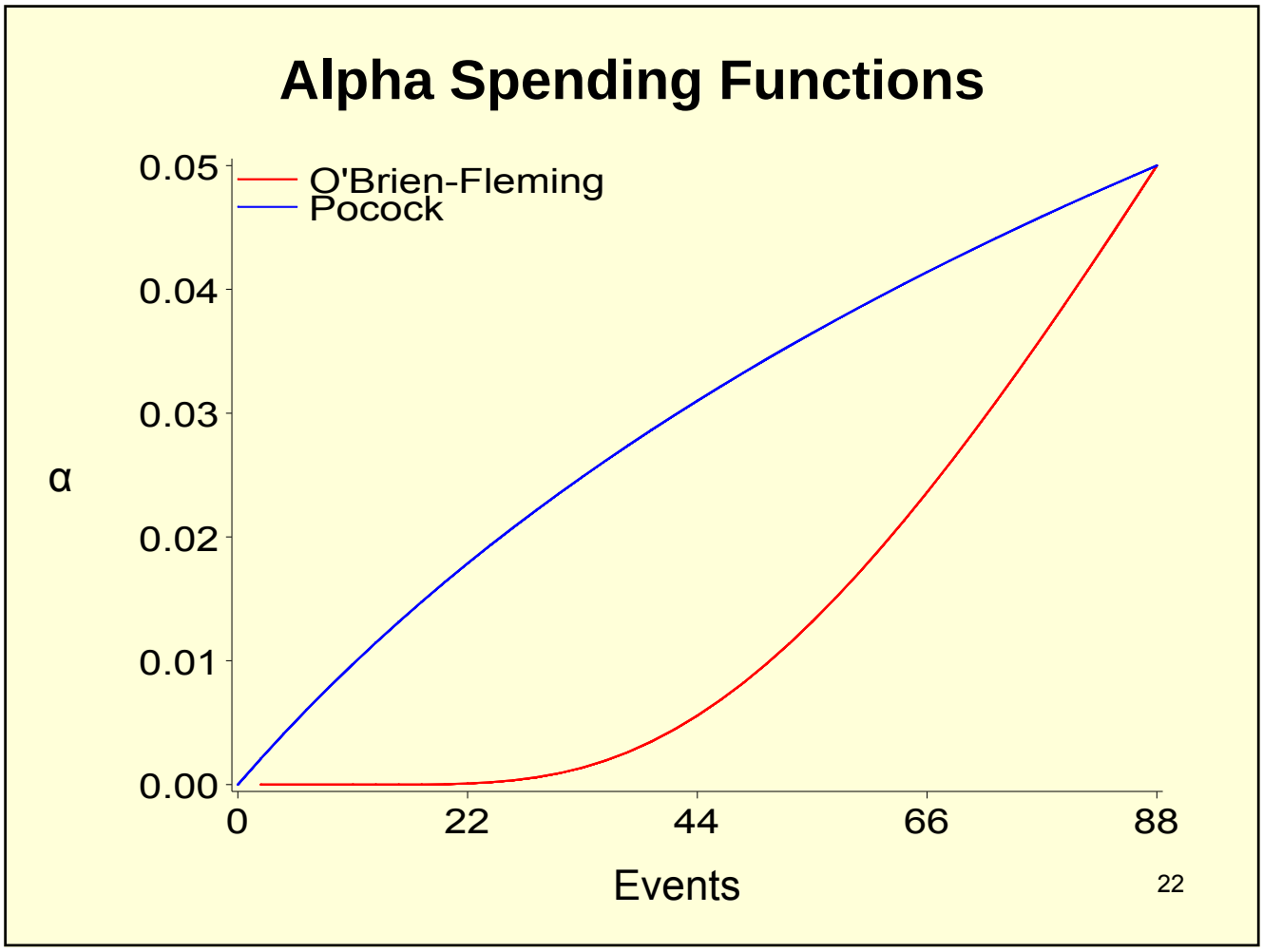
What if All Doesn't Go According to Plan?

- Boundaries technically designed for fixed number of equally spaced analyses
  - What if 77 events (rather than 66) had occurred prior to the 3rd look at the data?
  - Or, what if DSMB had requested an additional look between 66 and 88 events to explore AE issue?
  - O'Brien-Fleming and Pocock boundaries are fairly robust to unequally spaced analyses, but neither accommodates additional, unplanned analyses
    - What about Haybittle-Peto procedure?

20

Flexible Monitoring – "Alpha Spending"

- Lan and DeMets developed a very flexible boundary procedure that accommodates unequal timing, additional looks, even extending the trial
  - Called "alpha spending" functions
  - Researchers can choose to "spend" their alpha (i.e., conduct interim analyses) any way they want
  - Ensures that total alpha "spent" is no more than 0.05 (or whatever was specified)
- There are other spending function options


Alpha Spending Functions
O'Brien-Fleming
Pocock
0.05
0.04
0.03
0.02
0.01
0.00
α
0
22
44
66
88
Events
22

Flexible Monitoring – "Alpha Spending"

- In protocol, we typically specify: "The Lan-DeMets spending function with O'Brien-Fleming type boundaries will be employed to preserve the overall one-sided type I error rate for effectiveness at the 0.025 level, regardless of the timing of the analysis."

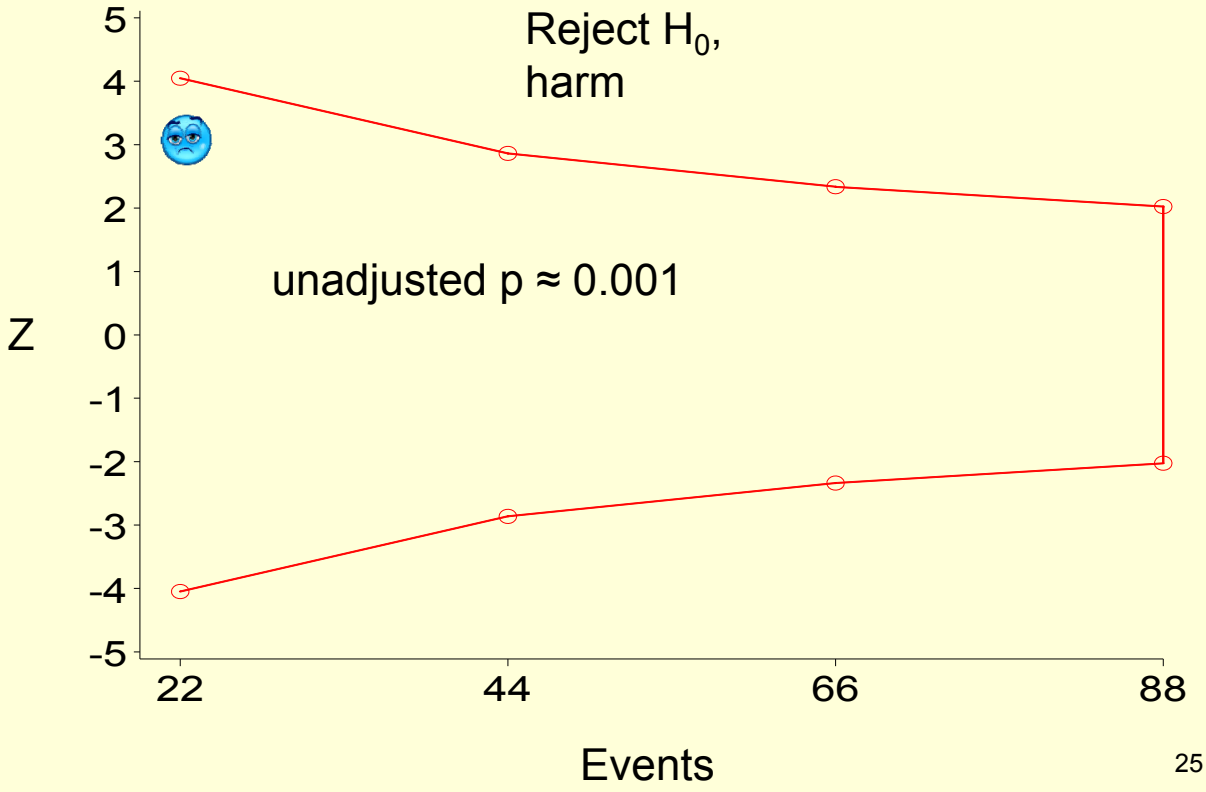
- Software available for calculating (East, "ldbounds" package in R – this is a "free" download at http://www.r-project.org/, others?)

Alpha Spending - Example

- Suppose our 3rd interim analysis happened after 77 rather than 66 events (i.e, after 88% of total rather than 75%)
  - L-D spending function would have used Z = -2.14 rather than -2.34 at 77 events
  - At 88 events, would use z = -2.02
- Suppose DSMB requested additional interim analysis at 77 events given safety issues?
  - L-D spending function at 77 events would have used Z = -2.21 (having used -2.34 at 66 events)
  - At 88 events, would use z = -2.06 (slight penalty)

24

What if our first look had gone other way?

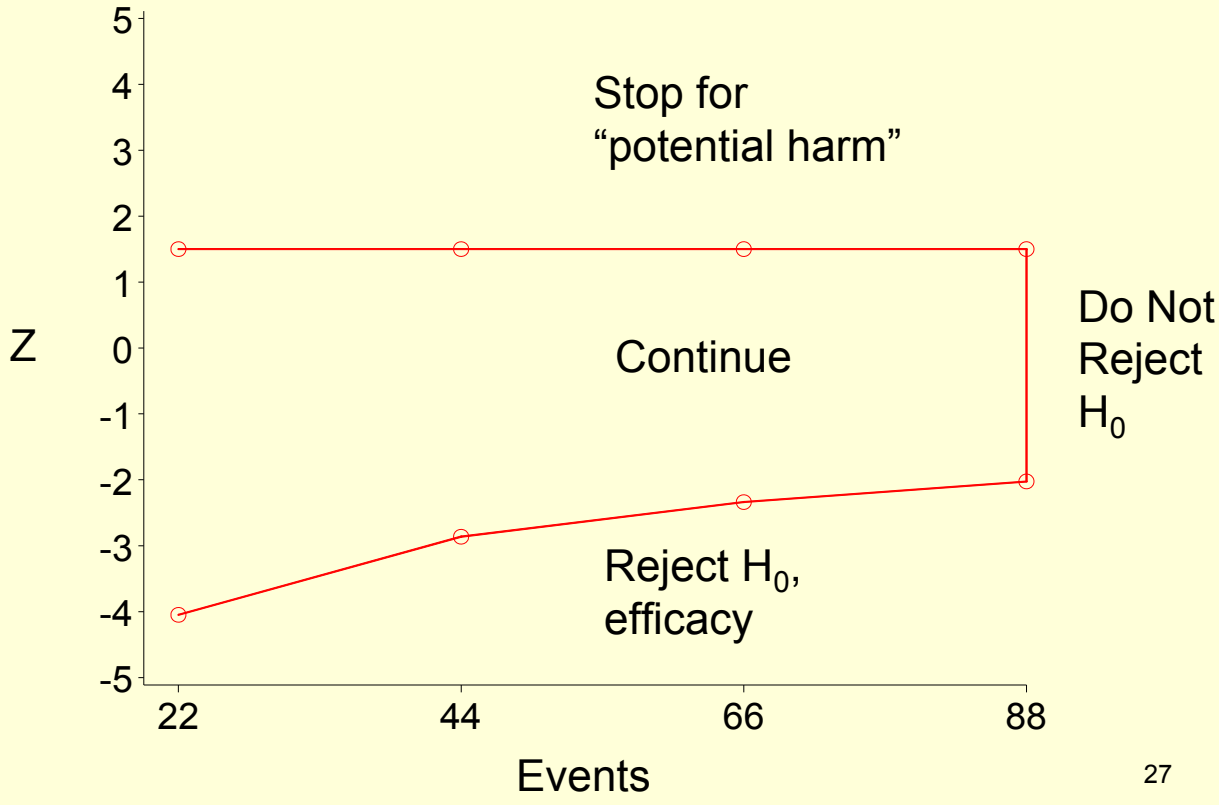


25

Asymmetric Boundaries

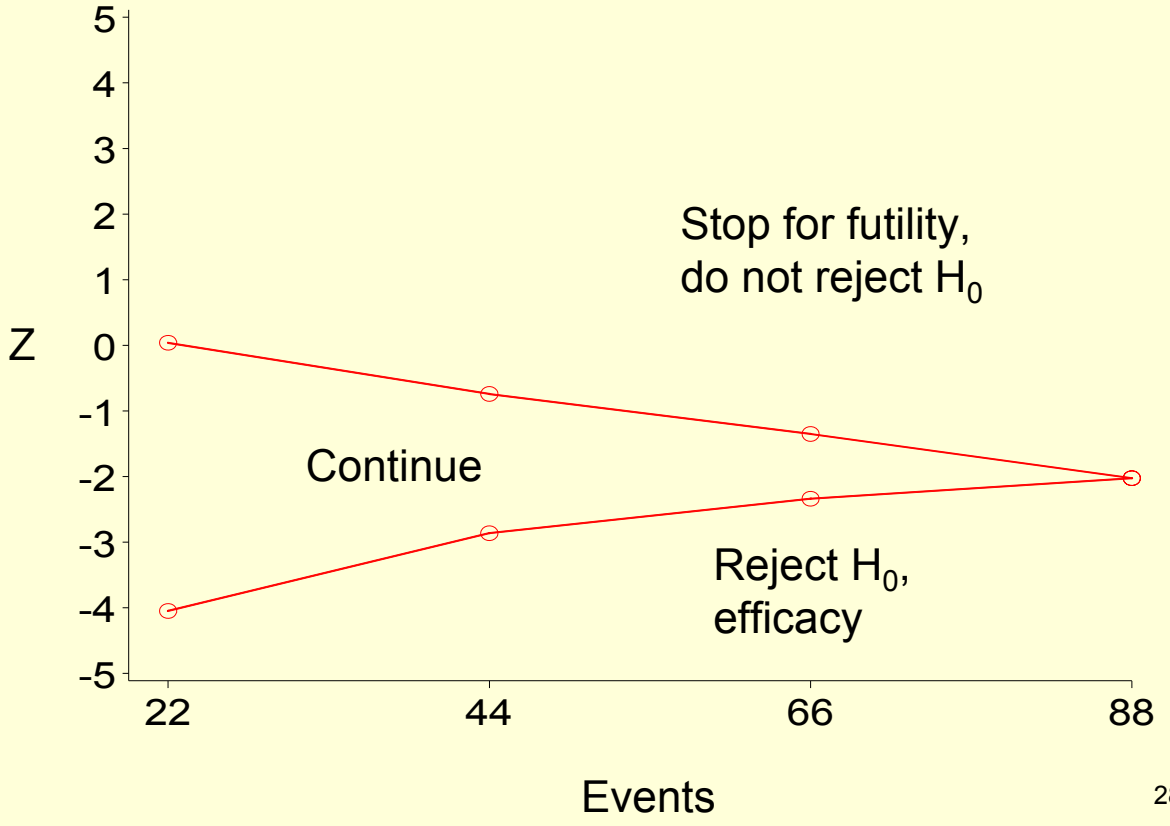
- It can be unethical to continue a study to get definitive evidence that product is harmful
- Typically need to be more “liberal” wrt harm
- “Trending” toward harm might be enough for most DSMB members, anyway
- Also might want to stop for futility
  - E.g., if it’s highly improbable that study will provide a significantly favorable result given current data
  - Fiscally logical to kill an ineffective product ASAP and move onto the next one

“Potential Harm” Boundary



27

Futility Boundary



28

Questions?