Statistics•Collaborative

Interim Boundaries or Guidelines – A Guide for and from the Perplexed

Janet Wittes, PhD

5th Annual Janice Pogue Lectureship in Biostatistics

June 28, 2022

Column

CLINICAL TRIALS

Clinical Trials
2014, Vol. 11(5) 601–604

DOI: 10.1177/1740774514534347
ctj.sagepub.com
SAGE

Clinician-trialist rounds: 23. When an RCT's Data Center Report drowns vital information in seas of data: Where's Waldo?

Janice Pogue[1,2] and David L Sackett[3]

Boundaries or guidelines

- Many people object to calling the lines we use as
  - Rules
  - Boundaries
- They prefer to call them "guidelines"
- My preference
  - These are boundaries
  - We use them as guidelines

Types of α-preserving boundaries

- Statistically formal - usually use Lan-DeMets spending functions
  - Pocock
  - O'Brien-Fleming
  - G-rho
- Statistically informal
  - Haybittle-Peto
  - Yusuf type

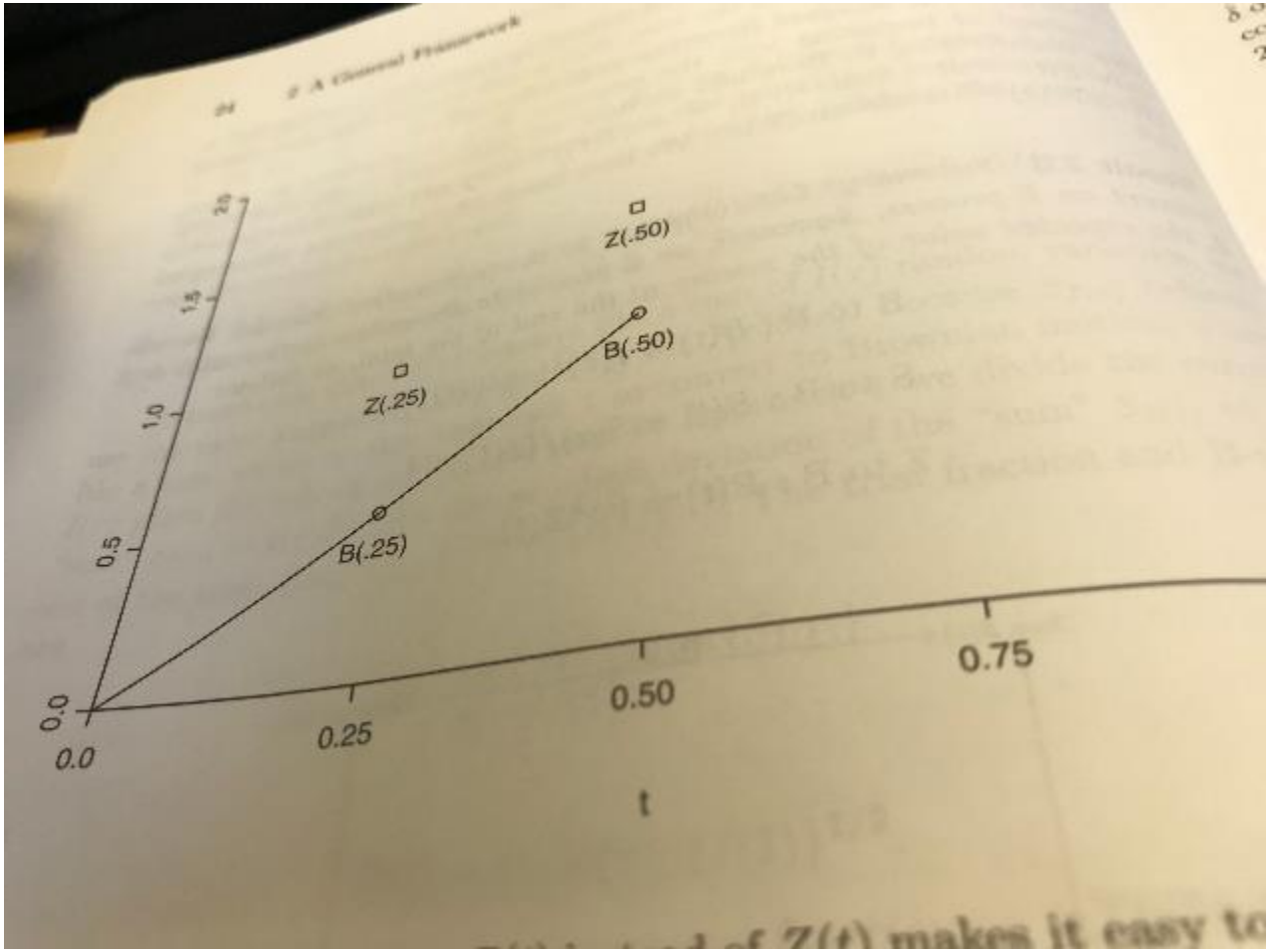
- For our examples, we are going to use 5-look boundaries
  - In practice, we apply Lan-DeMets use functions

I

Z-values and B-values

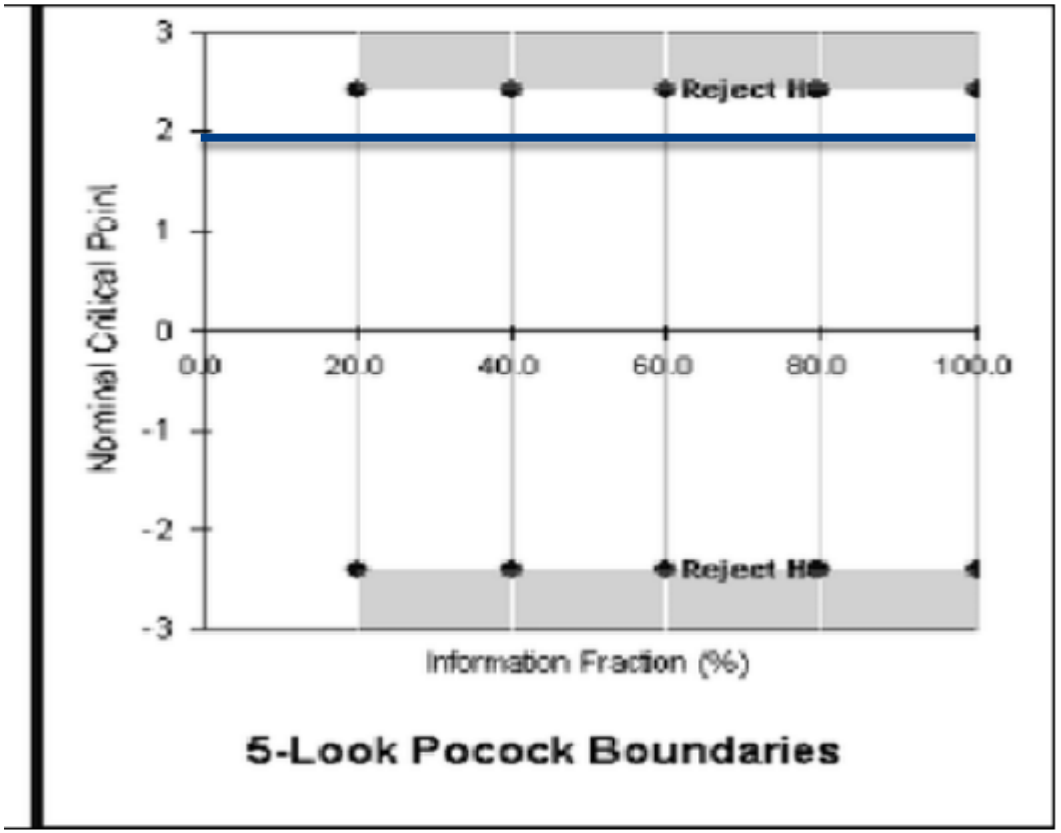
- Let *t* be the proportion of information in the trial
  - At beginning of trial $t=0$
  - At end of trial $t=1$
- At time *t*, the Z-value is Z(t)
  - At end of trial, Z=Z(1)
  - Z does not increase linearly over time
- BUT, if we multiply Z(t) by $\sqrt{t}$, we get $B(t)= \sqrt{t}\, Z(t)$
  - B(1)=Z(1)
  - And, B(t) increases linearly over time

The linearirty of B(t)



Pocock boundary

- Blue line – 1.96 not α-preserving
- The dots - Pocock
  - z= 2.413
  - p=0.0158



Message – please stop as early as you can and still preserve Type I error rate.

O'Brien-Fleming boundary

Pocock

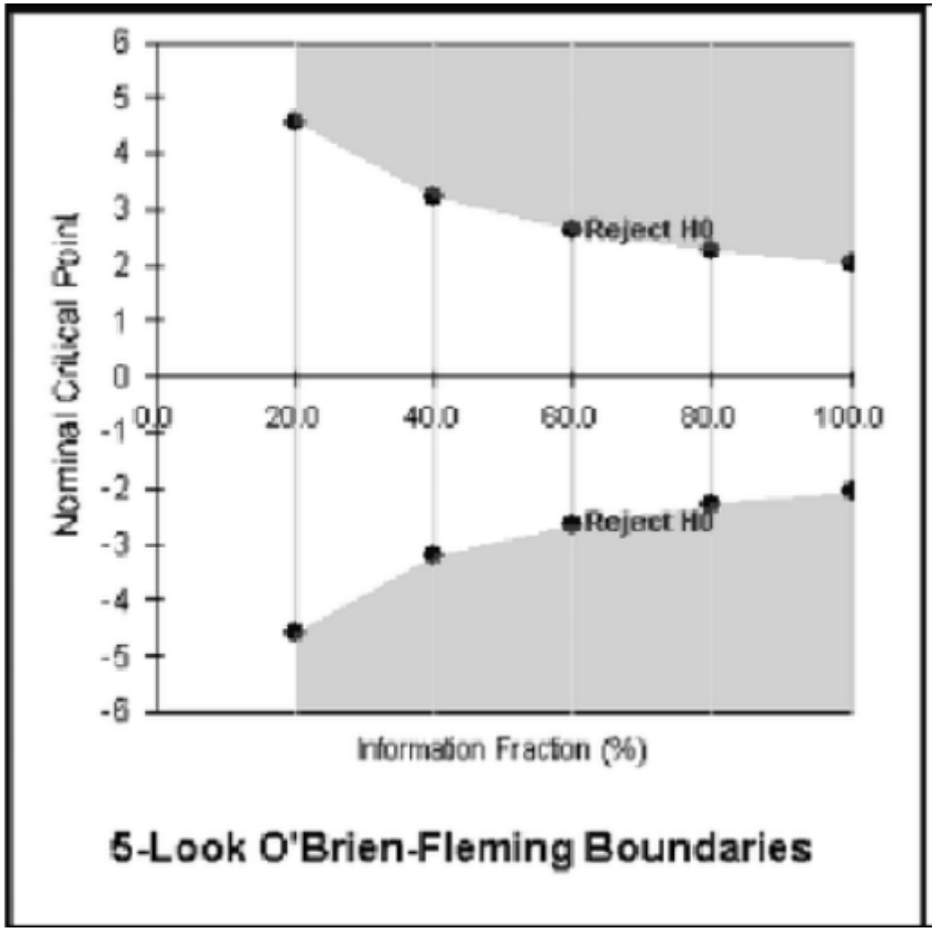
z= 2.413

p=0.0158

Message: please stop early

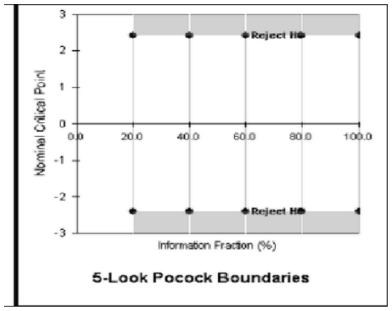
Look	z	p
1	4.555	0.000005
2	3.221	0.013
3	2.630	0.0085
4	2.277	0.0228
5	2.0317	0.0417

Message: don't stop early!



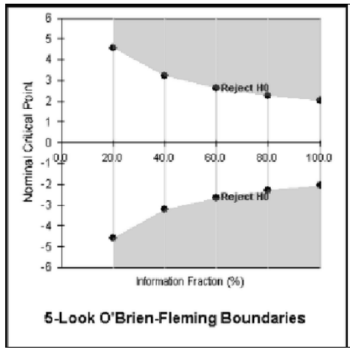
Haybittle-Peto boundary

Pocock



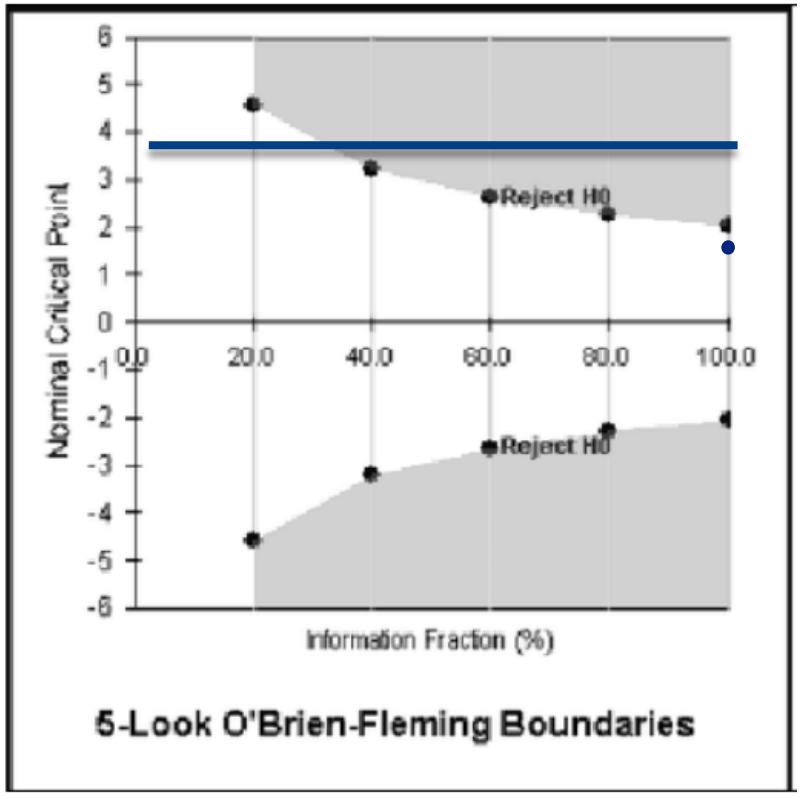
Message – stop as early as you can

O'brien- Fleming



Message: don't stop too early

Haybittle-Peto: blue line: z=3.3, then 1.96 p=0.001 until the last; at end p=0.05



Please try to hang on to the end.

And the Yusuf ("please don't stop") boundary

- Two z-values of 4 in a row
- Two z-values of 4 followed by a z of at least 3
- Three z-values of 4 in a row

- Message – you better have a REALLY GOOD reason for recommending early stopping!!!

And some other messages

- If you stop early, the FDA will not approve this drug
  - So no boundaries for efficacy at all
  - Pushback from journal reviewer
    - Every trial should allow early stopping for efficacy

- FDA – you have to have a futility boundary
  - Response from α-police: Yeah! We can recapture α
  - My retort – it better not be binding!
    - There may be a good reason for continuing
    - Don't even think of recapturing α

Aducanumab

On March 21, 2019, Biogen and Eisai announced they would terminate all currently ongoing aducanumab trials, following an interim analysis that predicted EMERGE and ENGAGE would miss their primary endpoints (see Mar 2019 news). On April 24, 2019, Biogen announced it would not initiate an anticipated Phase 3 secondary prevention program with aducanumab (Biogen Q1 Update), and removed it from its pipeline (May 2019 conference news).

On October 22, 2019, Biogen announced that the interim futility analysis was wrong, and that subsequent analysis of a larger data set instead showed EMERGE had met its primary endpoint. People on the highest dose, 10 mg/kg, had a significant reduction in decline on the primary endpoint, the CDR-SB. This group also declined less on secondary endpoints MMSE, ADAS-Cog, and ADCS-ADL-MCI. The low-dose group had some slowing of progression, but the differences were not statistically significant from placebo.

The ENGAGE trial did not meet the primary endpoint; however, an exploratory analysis suggested that a subgroup of people who had received 10 or more 10 mg/kg doses declined more slowly, similar to comparable EMERGE participants.

What happened

- There were two studies
- Very rigid futility rules
- DMC followed them to the letter
- Most of the press reports say that both had to show futility
- What they failed to say, "futility based on pooled data"
- One study was showing benefit, the other harm
  - But the two together satisfied the futility criteria

Biogen's AdCom for aducanumab: 6-Nov-2020

- Half-way through the study, the DMC was to declare futility if the conditional power for each of the two studies was <20% **based on pooled data from the two studies**

- High dose vs. placebo
  - Study 301: Conditional power= 0%
  - Study 302: Conditional power =12%

What if you don't choose and define roles carefully? E.g., Biogen's AdCom for aducanumab: 6-Nov-2020

- Half-way through the study, the DMC was to declare futility if the conditional power for each of the two studies was <20% based on pooled data from the two studies

- High dose vs. placebo
  - Study 301: CP= 0%
  - Study 302: CP=12%

"The FDA acknowledges that the Applicant followed the prespecified plan by announcing the termination of the aducanumab Phase 3 studies in response to the futility analysis."

The FDA's comment about the June 14, 2019 Type C meeting

"It would have been more appropriate if futility had not been declared for those studies."

The unpooled data…

Study	% diff hi dose vs. placebo	Conditional power based on study-specific results
301	**18% (Harm)**	0%
302	**15% (Benefit)**	59%

Cases

- Crossed boundary; recommended not stopping
  - CURE –crossed a Haybittle-Peto-type (Yusuf)
  - REWIND – crossed an OF-boundary (Gerstein)
- No boundary; recommended stopping
  - ANCHOR and MARINA – didn't stop
  - N-MOmentum – did stop
- Boundary not crossed; recommended stopping

Crossed boundary – didn't stop

Two cases where the DMC followed their guidelines (and I believe they shouldn't have!)

- In both cases, I'm shall only discuss what is in the public domain
  - Pimavanserin (Acadia Pharmaceuticals): Alzheimer's psychosis
  - Aduhelm (Biogen): Alzheimer's disease

Dementia Related Psychosis (DRP)

Study 045: Primary Endpoint and Statistical Analysis Plan

- Primary endpoint: time from randomization to relapse of psychosis in double-blind period
- Prespecified interim efficacy analysis (after 40 relapses) with stopping criteria
  - One-sided p-value less than O'Brien-Fleming stopping boundary of alpha = 0.0033
- All analyses prespecified for full analysis set in all DRP patients

Acadia https://www.fda.gov/media/159318/download

What the DMC saw

	Events, n/N (%)				
	Pimavanserin	Placebo		HR (95% CI)	Two-sided p-value
DRP	12/95 (12.6%)	28/99 (28.3%)		**0.35** (0.17, 0.73)	**0.005**

"Boundary" was p=0.0033

Since 0.005<0.0033, DMC recommended stopping

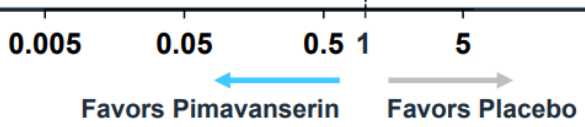
Trial was stopped.

Study had 3 subgroups of dementia (really 5)

CO-5

Study 045: Exploratory Efficacy by Dementia Subgroup in Double-Blind Period

	Events, n/N (%)		HR (95% CI)	Two-sided p-value
	Pimavanserin	Placebo		
DRP	12/95 (12.6%)	28/99 (28.3%)	**0.35** (0.17, 0.73)	**0.005**
ADP	8/61 (13.1%)	14/62 (22.6%)	**0.62** (0.26, 1.49)	**0.283**
PDD	1/15 (6.7%)	10/20 (50.0%)	**0.05** (0.02, 0.18)	**< 0.001**
Other (DLB, FTD, VaD)	3/19 (15.8%)	4/17 (23.5%)	**0.52** (0.08, 3.38)	**0.490**



ADP: Alzheimer's disease dementia DLB: Lewy body VaD: Vascular dementia

https://www.fda.gov/media/159317/download

Relevant Regulatory History: Complete Response

- Complete Response (CR) April 2021 concluded application did not provide substantial evidence of effectiveness for dementia-related psychosis
- Although Study 045 not powered for subgroup efficacy demonstration, subgroup observations included:
  - Results for Parkinson's disease dementia (PDD) subgroup were highly nominally statistically significant, appearing to drive overall results despite smaller size (n=35)
  - Results for Alzheimer's disease (AD) subgroup not nominally statistically significant despite largest subgroup (n=123)
  - Too few subjects with dementia with Lewy bodies (n=10) or frontotemporal dementia (n=3) to adequately represent those subgroup responses
  - No difference on time-to-relapse for vascular dementia (n=25)

….And the stock..

AdCom Friday June 17, 2022

Monday was Juneteenth –new US Federal holiday

Shares of **Acadia Pharmaceuticals** were crashing 35.4% as of 11 a.m. ET on Tuesday. The steep decline came after a Food and Drug Administration (FDA) advisory committee voted 9-3 against recommending approval of pimavanserin in treating Alzheimer's disease psychosis.

CURE

(Clopidogrel in Unstable Angina to Prevent Recurrent Events)

- Effects of Clopidogrel in Addition to Aspirin in Patients with Acute Coronary Syndromes without ST-Segment Elevation
- …..
- The data and safety monitoring board monitored the incidence of the primary outcome to determine the benefit of clopidogrel, using a modified Haybittle–Peto boundary of 4 SD in the first half of the study and 3 SD in the second half of the study.

CURE

(Clopidogrel in Unstable Angina to Prevent Recurrent Events)

- Effects of Clopidogrel in Addition to Aspirin in Patients with Acute Coronary Syndromes without ST-Segment Elevation

The Clopidogrel in Unstable Angina to Prevent Recurrent Events

- The data and safety monitoring board monitored the incidence of the primary outcome to determine the benefit of clopidogrel, using a modified Haybittle–Peto boundary of 4 SD in the first half of the study and 3 SD in the second half of the study. The boundary had to be exceeded at two or more consecutive time points, at least three months apart, for the board to consider terminating the study early.

CURE

(Clopidogrel in Unstable Angina to Prevent Recurrent Events)

- Effects of Clopidogrel in Addition to Aspirin in Patients with Acute Coronary Syndromes without ST-Segment Elevation
  The Clopidogrel in Unstable Angina to Prevent Recurrent Events

- The data and safety monitoring board monitored the incidence of the primary outcome to determine the benefit of clopidogrel, using a modified Haybittle–Peto boundary of 4 SD in the first half of the study and 3 SD in the second half of the study. The boundary had to be exceeded at two or more consecutive time points, at least three months apart, for the board to consider terminating the study early. There were two formal interim assessments performed at the times when approximately one third and two thirds of the expected events had occurred. Despite the fact that the preset boundary indicating efficacy had been crossed by the time of the second interim analysis, the board recommended that the trial continue until its planned end, in order to define more clearly whether the risks of major bleeding episodes could offset the benefits of therapy.

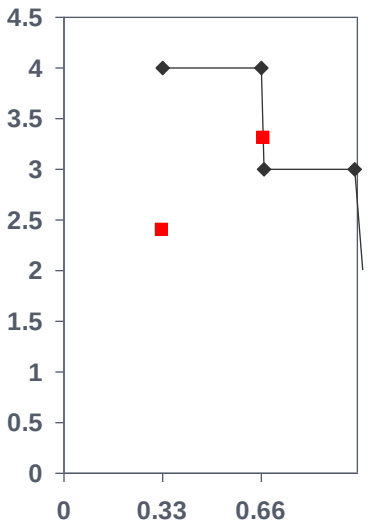
The Manuscript Writing Committee (Salim Yusuf, D.Phil., F.R.C.P.C., Feng Zhao, M.Sc., Shamir R. Mehta, M.D., F.R.C.P.C., Susan Chrolavicius, B.Sc., Gianni Tognoni, M.D., and Keith K. Fox, M.D., F.R.C.P.) assumes responsibility for the overall content of the manuscript.

Data Safety and Monitoring Board: G. Wyse (chair), J. Cairns, R. Hart, J. Hirsh, M. Gent, T. Ryan, J. Wittes

N Engl J Med 2001; 345:494-502DOI: 10.1056/NEJMoa010746

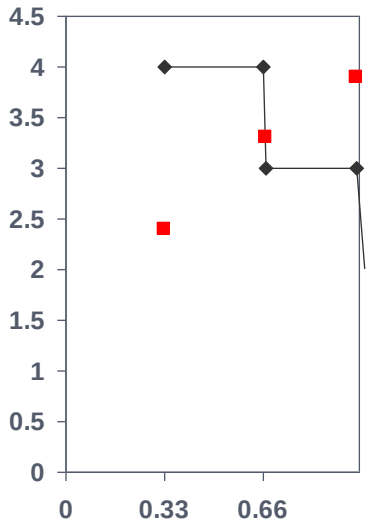
What we saw in CURE

- Z-values for CV death, MI, or stroke
  - Look z
  - ~1/3 2.4
  - ~2/3 3.3
- Why not stop?
  - We saw bleeding
    - Lots of excess minor bleeding
    - Also excess in intracerebral bleeds (7:1)
  - Only three more months to go
  - We did not tell the PI



Final outcome in CURE

- Final z-value was ~4
- Relative risk=0.80
- 95% CI: (0.72, 0.90)
- Intracranical bleeds – 7:5



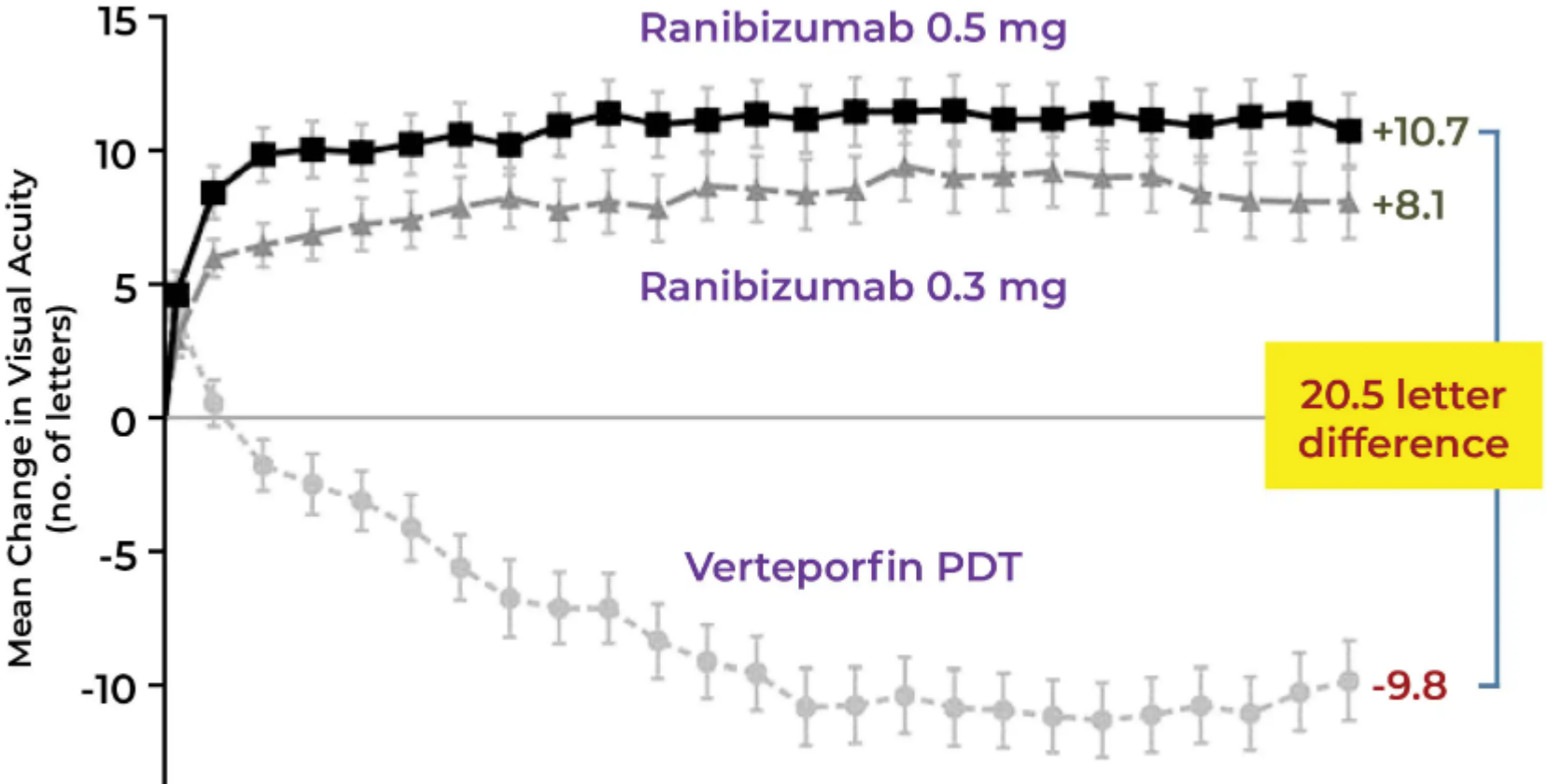
REWIND

- **Gertzel et al.(2019). Dulaglutide and cardiovascular outcomes in type 2 diabetes (REWIND): a double-blind, randomised placebo-controlled trial. Lancet 394: 121-130.**

No boundary – recommended stopping

ANCHOR and MARINA

- Age-related macular degeneration



N-MOmentum

- Double-blind. Placebo-controlled, randomized 3:1
- Neuromyelitis optica spectrum disorder (NMOSD)
- Primary endpoint – time to attack
  - Attack leads to permanent worsening

Didn't cross boundary; did stop

Sometimes the data are so overwhelming…

- That even when there is a boundary the DMC recommends stopping before the first planned look
- Very risky to do but sometimes the data overwhelms the "rule"
  - E.g., early nivolumab trial
  - DMC will create an extreme boundary
    - Will argue: data are so strong that the evidence is clear

Some other issues from Shrikant…

- Boundary has been crossed, but barely
  - Several reported outcomes have not been adjudicated
  - Once adjudicated, the actual final Z value could be below the boundary
  - My comment: ambiguous cases slower to adjudicate
- How should we weigh
  - Safety
  - Important secondary outcomes

Conclusion-futility

- Don't have "binding" futility rules
- Don't stop too early if treatment may have delayed effect

Conclusions-efficacy

- We need
  - Boundaries (guidelines)
  - An understanding of what the investigators want
  - Ability to prepare for stopping
  - Tools to resist stopping
- In preparation for a meeting at which stopping is likely
  - Think of how each DMC member will respond to the data
  - Be prepared to answer those questions
  - Prepare scenarios for the DMC

Conclusions- overall

- For all DMCs – reporting statistician must
  - Understand what
    - The investigators want
    - The regulators need
  - Understand study and data