Statistical considerations in evaluation of CV risk in diabetes programs conducted pre- and post-approval

CSRC Thinktank

April 13, 2011

J Todd Sahlroot, Ph.D.
Deputy Director and Statistics Team Leader
Diabetes Group
FDA/CDER

Disclaimer

The views and opinions expressed in the following PowerPoint slides are mine and not necessarily those of the FDA

Guidance for Industry Diabetes Mellitus — Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)

December 2008
Clinical/Medical

Dec 2008 Diabetes Guidance: CV results decision tree

UPPER BOUND OF 95% CI FOR HAZARD RATIO (HR)	CONCLUSION
>1.8	Inadequate to support FDA approval
>1.3 but <1.8*	Postmarketing trial(s) needed to show definitively <1.3
<1.3*	Postmarketing cardiovascular trial(s) generally not necessary

Note: hazard ratio (HR) considered interchangeable with risk ratio

*with a “reassuring” point estimate Table courtesy of Hylton Joffe

Sample size

event driven

Number of primary CV events needed			
Non-inferiority margin for hazard ratio	power		
	80%	85%	90%
1.8	91	105	122
1.3	456	523	611
Logrank statistic, 1:1 Randomization, true HR=1			

Hypothesis testing structure

- Guidance formulates two sets of hypotheses
  - **H_{01}: HR > 1.8 vs H_{11}: HR ≤ 1.8**
  - **H_{02}: HR > 1.3 vs H_{12}: HR ≤ 1.3**
- Goal: test each null hypothesis statistically and reject both hypotheses, equivalent to demonstrating non-inferiority for both margins. I.e., 95% CI for HR is < 1.8 and ultimately is < 1.3
  - Simultaneously – Pre-market data are sufficiently strong to reject 1.8 and 1.3 null hypotheses (95% CI for HR is < 1.3) OR
  - Sequentially -- Pre-market data are sufficient to reject 1.8 null hypothesis but not 1.3 null hypothesis (i.e., 95% CI for HR is < 1.8 but is > 1.3). After approval, collect additional data to reject 1.3 hypothesis (i.e., 95% CI for HR is < 1.3)
- What is type 1 error rate?
  - We've assigned each null hypothesis its own type 1 error rate (α) equal to 2.5% (one-sided) consistent with the Guidance requiring 2-sided 95% CIs for assessing HR

Two null hypotheses

- Not the usual statistical testing problem – complexities arise from
  - two-dimensional testing space (null space, time space)
  - hypotheses tested at the same or different times
  - each hypothesis can be tested one or more times (repeated testing over time)

Repeated testing over time

- Type 1 error for each hypothesis should be controlled for any multiple analyses
- Most sponsor proposals have involved repeated testing over time
- Even if not explicitly acknowledged, repeated testing implied for 1.3. Sponsors testing for 1.8 may also claim a win wrt 1.3 if 95% CI for HR is < 1.3. Otherwise, sponsor will conduct one or more additional analyses post-market to rule out 1.3. *At least 2 chances to win.*

How can we efficiently address statistical aspects of this hypothesis testing framework?

- Single group-sequential boundary
- Double group-sequential boundaries
- Adaptive designs – *won't address this here*
- Other approaches?

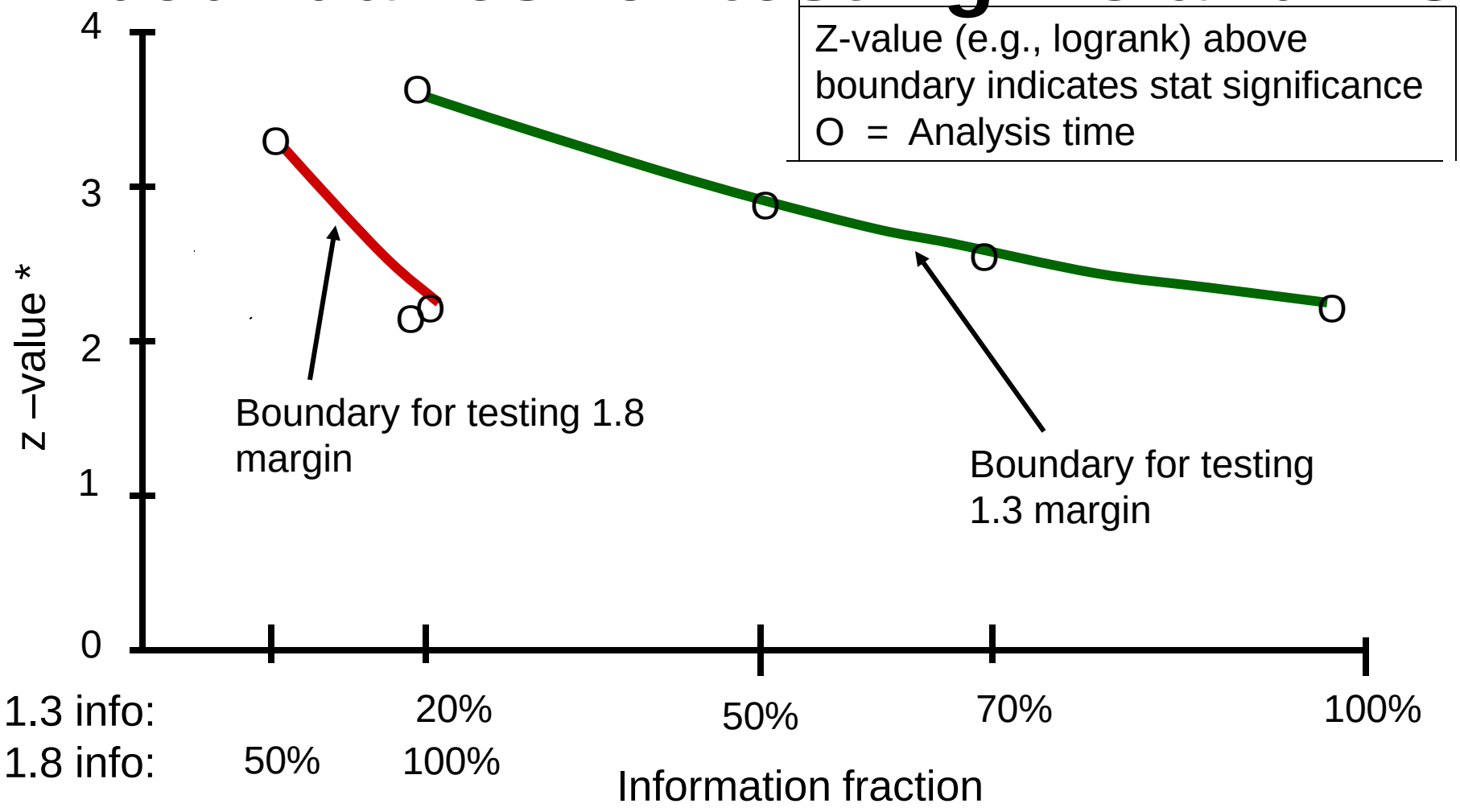
Single group-sequential boundary

- e.g., O'Brien-Fleming
- At interim or final analysis, test 1.8 at alpha = α_0 corresponding to the GS procedure. If "win" then test 1.3 at same α_0
- Approach preserves type 1 error across repeated analyses over time and across both null hypotheses
- Conservative
  - needlessly controls for multiplicity across the two null hypotheses
  - More difficult to reject 1.8 hypothesis early because interim alpha(s) for testing 1.8 should be small to have sufficiently large final alpha (close to 2.5%) to test for 1.3.

Separate group-sequential boundaries

- We've been suggesting to sponsors two distinct group-sequential boundaries, one for each null hypothesis
- Allows simultaneous testing of 1.8 and 1.3 at different alphas or testing for one hypothesis and not the other

Separate group-sequential boundaries for testing 1.8 and 1.3



*standardized with respect to relevant margin

How can we efficiently address statistical aspects of this hypothesis testing framework?

- Single group sequential boundary
- Double group-sequential boundaries
- Adaptive designs – *won't address this here*
- **Other approaches?**

Other statistical themes

- Study heterogeneity
  - Different controls, populations, durations, etc
- Analysis populations
  - Intent-to-treat vs on study drug
  - National Research Council recent report recommends following all patients for the primary endpoint out to the end of the study even if the patient is off drug (ITT)
  - Both approaches are needed