Thoughts on the presented statistical analysis

Daniel Gillen, PhD
Consultant to the Committee
Department of Statistics, UC Irvine

Typical criteria for approval

- Submission of two independent well-controlled clinical trials as substantial evidence for effectiveness
- Goal of statistics is to quantify uncertainty in samples in order to make inference or generalize to the larger population

Typical criteria for approval

- A primary reason for requiring consistent results on two independent trials is to broaden the generalizability of observed results
  - Clinical centers
  - Training
  - Patient pools / cohort effects

Current reference standard for statistical evidence

- P-value - Probability of observing results as or more extreme than those actually observed if the null hypothesis were true
  - In the current setting null hypothesis is equal rates of preterm births in each treatment arm
- Reference standard for a single trial is a one-sided P-value of 0.025 or less

Statistical evidence from a single confirmatory trial

- In order to provide sufficient statistical evidence from a single confirmatory trial it has been suggested that one require a P-value of $0.025^2=0.000625$

(the threshold corresponding to 2 independent level .025 tests)

Results reported by the study sponsor (ITT)

- 37 week endpoint
  - Obs proportions: 0.371 vs. 0.549
  - Obs difference: -0.178
  - 95% CI: -0.28, -0.07
  - Corresponding P-value: 0.0003

Results reported by the FDA

- FDA notes the use of an interim monitoring plan
  - 2-sided level .05 O'Brien-Fleming rule
  - 2 interim analyses one final analysis
- Adjusted results
  - Obs difference: -0.178
  - 95% CI: -0.28, -0.07

Results adjusted for interim analyses

- Assumptions:
  - 2-sided level .05 O'Brien-Fleming boundary
  - Three equally spaced analyses (actually took place at 15.2% and 70.2% of maximal sample size)
  - Final analysis sample sizes: 310 vs. 153
  - Baseline event rate of 0.549

Results adjusted for interim analyses

- Adjusted results (based upon sample mean ordering)
  - Obs difference: -0.178
  - Bias adjusted diff: -0.165
  - Adjusted P-value: 0.0035

Results adjusted for interim analyses

- Adjusted results for other endpoints

Endpoint	Obs Diff	Adj Diff	Adj P-value
35 week	-0.091	-0.086	0.068
32 week	-0.070	-0.066	0.156
28 week	-0.005	-0.005	0.919

Final note

- P-values only represent one criteria of evidence
- Also need to consider clinical significance of observed point estimates
  - Observed rate of pre-term births in placebo arm
  - Mean time to birth
- Generalizability of findings
- Safety profile
- Urgency of clinical need