Roles of Multiplicity Adjustment in Regulatory Applications

H.M. James Hung
DBI/OB/OTS/CDER, US FDA

Presented in NISS Webinar, 05/06/2020

Disclaimer

This presentation reflects the views of the author and should not be construed to represent FDA's views or policies.

Outline

- Drug development process
- Clinical trials
- Clinical trials for proving efficacy
- Multiplicity control/adjustment
- Many challenges remain
- Remarks

Drug Development Process

- Discovery and development
- Preclinical research:
- Clinical research
- FDA review
- FDA post-market safety monitoring

https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process

Drug Development Process <cont'd>

- Preclinical research

  Drugs undergo laboratory and animal testing to answer basic questions about safety.

- Clinical research

  Drugs are tested on people to make sure they are safe and effective.

https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process

Clinical Trials

- Efficacy assessments
  - Providing proof of concept
  - Learning/Exploring based on the data of relatively homogeneous humans with disease or condition
  - Proving efficacy/effectiveness of test drug in a larger number of intention-to-treat patients
- Randomized to test vs. control (e.g., no test)

Clinical Trials For Proving Efficacy <Cont'd>

Each trial tests pre-specified clinical/statistical hypotheses for:

- multiple doses / dosing regimens
- multiple endpoints (primary, secondary)
- multiple patient subgroups (marker +/-)
- multiple times (e.g., when drug starts to work? How long drug's effect maintains?)
- A hypothesis may be tested repeatedly

Clinical Trials For Proving Efficacy <Cont'd>

Relationships among the hypotheses are often complex

In most practices for efficacy assessments:

- Search for at least two "positive or successful" trials
- Decide on what trial results to include in labeling

Multiplicity Control/Adjustment

Control the overall (familywise or studywise) type I error probability at a fixed level of α (e.g., two-sided $\alpha = 5\%$)

- Overall type I error probability: probability of mistakenly rejecting a **null** hypothesis (**mostly, no treatment effect**) regardless of other hypotheses
- Within each trial, each null hypothesis is tested at level of α or less (mostly, $p \leq \alpha$)

Multiplicity Control/Adjustment <cont'd>

Multiplicity control is applied to screen out the null hypotheses for the purpose of regulatory decision and labeling

After multiplicity control, p value is often used to assess strength of statistical evidence (against null hypothesis) within individual trials in a uniform fashion

Multiplicity Control/Adjustment <cont'd>

For rejected null hypothesis, should p value be adjusted following the pre-specified multiple testing procedure?

- If adjusted, definition of p value is changed
- If not adjusted, the same magnitude of p value against the same null hypothesis may have different implications in different trials

Multiplicity Control/Adjustment <cont'd>

Multiplicity control is applied within each pivotal trial, not to the drug development program.

In learning trials, selecting hypotheses using nominal p value without some kind of multiplicity control may affect success probability in pivotal trials.

Multiplicity Control/Adjustment <cont'd>

Multiplicity control is applied within each pivotal trial, not to the drug development program.

Ex. For a hypothesis tested in two pivotal trials, p value was small but did not pass the multiple testing scheme within each trial. This hypothesis is very likely to have been falsely ignored.

Many Challenges Remain

- Clinical contexts behind hypotheses are often different, e.g., dose, endpoint, subgroup, ….
- Widely used is a single-string hierarchical testing; i.e., continue testing the string of null hypotheses until a null hypothesis cannot be rejected. Often this is not sensible, per clinical contexts.

Many Challenges Remain

- Relationships between relevant hypotheses need to be considered. How?
- What if the primary endpoint requires two positive trials but a secondary endpoint needs only one positive trial?

Many Challenges Remain

- In RCT, test a reasonably likely surrogate (RLS) for accelerated approval, followed by testing clinical benefit endpoint. In case RLS fails or does not appear to predict clinical benefit during the trial, could still continue the trial to directly test the clinical benefit endpoint. What is the overall type I error?

Many Challenges Remain <cont'd>

- What if …
  - only one positive trial but $p < 0.0001$
  - only two out of 10 trials are positive
  - trial results are simply described but not formally tested

Many Challenges Remain <cont'd>

- In complex application scenarios such as use of Master Protocol under which several medical products can be tested,
  - what if these products have the same mechanisms of action for efficacy
  - what if these products have different mechanisms of action for efficacy

Remarks

- Multiplicity control has served reasonably well in screening null hypotheses within the desirable limit for regulatory applications
- Many questions remain even within each trial for each drug. Q: Should dose selected for marketing be based mainly on hypothesis testing?

Remarks

- Once a drug is shown effective on an endpoint, may need further description of the drug's performance on that endpoint, e.g., for composite endpoint, need to assess treatment effect on the components of the composite. Such description can render a competitive edge. Q: Shouldn't they be subject to some kind of multiplicity control?

Thank you