THE SCIENCE of POSSIBILITY

Adjusting for Multiplicity in Clinical Trials - Composite and Secondary Endpoints

Abdul J. Sankoh and Haihong Li
Vertex Pharmaceuticals

Abdul_sankoh@vrtx.com

SCT Meeting, 2013

Source of Multiplicity (1)

Need to adjust for multiplicity to control the Family-wise Error Rate (FWER)

◆ Multiple endpoints, need at least one to be significant (union-intersection)

◆ Multiple treatments/doses

◆ Interim analysis (multiple looks)

◆ Subgroup analysis

◆ Composite endpoint

Source of Multiplicity (2)

No need to adjust for multiplicity

- Multiple endpoints, need at least one to be significant (intersection-union)
- Multiple analysis sets (e.g., PPS)

Depends…

- Secondary endpoints
- Multiple statistical methods (*e.g., sensitivity analysis, selecting primary analysis based on data confirming assumptions: equal variance, covariance structure, missingness, presence of carryover effect, etc.*)
- Multiple regulatory agencies
- Addressing different objectives in a single study

Primary endpoint

The ideal clinical and statistical situation for design of confirmatory clinical trials is to prospectively specify a single primary endpoint that

- completely characterizes disease under study and
- permits efficient evaluation of treatment effect.

Desirable features of the ideal endpoint:
- Relevance
- Reliability
- Validity
- Sensitivity
- Reproducibility
- Interpretability

Desirable features of the primary endpoint (1)

♦ **Clinical relevance:** must focus directly on study primary objective, mechanism of action of intervention, and impact patients' well-being.

♦ **Reliability:** must be easily diagnosable and capable of being assessed in all subjects consistently, i.e., lack of measurement error

♦ **Validity of comparison:** must be ascertainable and classifiable in an unbiased way so as to allow unbiased between treatment group comparisons

Desirable features of primary endpoint (2)

◆ **Reproducibility:** meaningful and acceptable demonstrated clinical evidence of treatment effect for target population must be reproducible over time

◆ **Optimality of clinical and statistical significance:** must be sensitive to meaningful changes induced by treatment and associated with readily available simple statistical analysis methods of optimal precision

◆ **Completeness and interpretability:** must offer broad and comprehensive ascertainment and unambiguous interpretation of treatment effect

Composite Endpoints (1)

◆ Treatment effect is assessed by composite endpoints in some disease areas

- ◆ the disease manifests itself in a multi-faceted form of the same underlying cause
- ◆ clinically important events are rare (for binary/time to event endpoint case) or not sensitive enough (for continuous endpoint case)

◆ Binary, time to event, or continuous/index/responder

Composite Endpoints (2)

◆ It is often expected in addition to statistical significance for the composite endpoint,

- ◆ Components trend positively
- ◆ Overall treatment effect is not driven by softer components

◆ Multiplicity adjustment needed if also interested in pre-specified components

Adjustment for Multiplicity for Composite Endpoints

◆ Traditional fixed sequence strategy

◆ Flexible fixed sequence strategy (FFS) (Huque and Alosh 2008)

◆ Adaptive alpha allocation approach (Li and Mehrotra 2008, Li, Sankoh, and D'Agostino 2013)

◆ Consistency-adjusted alpha-adaptive strategy (Alosh and Huque, 2010)

◆ Multi-branched testing strategy (Huque, Alosh and Bhore 2011)

Specification of single primary endpoint may not be practical if

◆ Disease under study manifests itself in a multi-faceted form

⇒ Composite endpoints: Allergic rhinitis, Crohn's disease, RA

◆ Clinically important events are rare (for binary) or not sensitive enough (for continuous) variable

⇒ Composite endpoints: CV, Allergic rhinitis

◆ No consensus in clinical community for such a single endpoint

⇒ Multiple primary and secondary endpoints: Ulcerative colitis

◆ Areas with prevailing methods for assessment of efficacy dictating multi-dimensional approach both for primary endpoint selections and evaluations

⇒ Multiple primary and secondary endpoints: Device

◆ Desire for broader and more complete evaluation of treatment benefit.

⇒ Multiple primary and secondary endpoints

In practice …

◆ Several efficacy response variables in randomized clinical trials are identified and classified as

    ◆ Primary & Secondary endpoints.

◆ Primary (composite) endpoint(s) address(es) directly the primary study objective.

    ◆ Focus of study design and primary statistical analysis method, including study power calculation, is on primary endpoint(s).

    ◆ Sought indication and subsequent labeling claim is often limited to trial findings based on primary endpoint(s).

Secondary Endpoints

◆ Secondary (components of composite) endpoints serve a number of important roles

- ◆ Key endpoint, critical on their own (e.g., overall survival): Adjustment required
- ◆ Supportive, provide more comprehensive understanding of drug effect: No need for multiplicity adjustment

◆ Findings based on secondary endpoints do not generally lead to labeling claim if primary objective not met.

Example 1: Carvedilol in Heart Failure (CHF)

◆ Four studies in support of NDA of Carvedilol in patients with moderate and severe heart failure.

   ◆ Primary endpoint: Exercise capability

◆ Due to emerging concerns of excessive mortality risk caused by other CHF drugs:

   ◆ FDA asked sponsor to also study mortality as an "additional" endpoint.

   ◆ A DSMB was formed "to review unblinded data … to ensure that an excess of events in either therapy is not occurring that should mandate a modification or termination of the clinical trials program".

The Carvedilol case (2)

◆ Primary endpoint of exercise capability was not significant in any of the studies:

  ◆ P-value > 0.05.

◆ However, results for mortality rate were great:

  ◆ P-value =0.0001

◆ DSMB recommended stopping trials and offering carvedilol to all placebo recipients.

◆ Key questions:

  ◆ **Should carvedilol be approved?**
  ◆ **Should label include mortality?**

The Carvedilol case (3)

◆ 1995 (1st) AC meeting voted NO

  ◆ 2nd AC meeting voted YES

◆ Some interesting conversations at the meetings:

  ◆ "…the overall consistency through analyses after analyses and analyses that were specified by the sponsor and the analyses that were asked by the FDA… all seemed to go in the same direction"

  ◆ "…In the absence of prospective statements by the sponsor, … , I again have the freedom to choose a very conservative track, and the conservative track is concerned for the risk of a type 1 error in the population at large…."

  ◆ "…if the trial is negative on the primary endpoint, and overall alpha I spend even for a phenomenal finding for mortality, still winds up being unacceptably high, …"

  ◆ "… I agree with everything that X says, …, but I for exactly that reason vote no. …"   (X voted yes)

The Carvedilol case (4)

◆ Carvedilol was approved based on recommendation of majority of 2nd Cardio-Renal AC Meeting.

◆ But FDA wanted clarification-

"… But before we quit, I'd like to find out two things so I understand the sense of the committee.

   ◆   So what does carvedilol do? …,

   ◆   should it be allowed to claim that it saves lives? …"

◆ Carvedilol was approved but without mortality in the label!

   ◆ This case culminated in 3 articles and the birth of PAAS

      ◆ Moye´ (CCT 20,1999), Fisher (CCT 20, 1999), D'Agostino (SIM, 2000)

VERTEX™

Example 2: Tivozanib for Renal Cell Carcinoma (RCC)

◆ A single Phase 3 study in support of NDA of Tivozanib in patients with RCC.

  ◆ Primary endpoint: Progression free survival (PFS)

  ◆ First secondary endpoint: Overall survival (OS)

The Tivozanib case (2)

◆Primary endpoint of PFS was significant:

   ◆HR (vs. Sorafenib) = 0.80

   ◆Median Survival 11.9 vs. 9.1 months

   ◆P-value = 0.04

◆However, results for OS were not great:

   ◆HR (vs. Sorafenib) = 1.25

   ◆Median Survival 28.8 vs. 29.3 months

   ◆P-value =0.11

The Tivozanib case (3)

- **2013 ODAC meeting:**
  - FDA asked in the briefing document:

    "In considering the results from a single randomized trial submitted in support of marketing approval of a new molecular entity, FDA expects that the trial will be adequately designed and well conducted and that the results will be internally consistent. We are asking the ODAC's advice on whether this single trial is sufficient to support approval of tivozanib for the indication of treatment of patients with advanced renal cell cancer or whether an additional trial is necessary before considering marketing approval."

  - Committee voted NO (13:1)

- **FDA at the ODAC:**
  - "…Obviously OS is a much more important endpoint than PFS…"
  - "Extremely disappointed" with the proposed labeling of tivozanib, which does not mention overall survival data
  - "Progression-free survival is primarily a radiological endpoint" and thus has limitations, … "Overall survival assures that both safety and efficacy is critical for the risk-benefit analysis and is an important endpoint for patients."

VERTEX™

Multiplicity adjustment for secondary endpoints

◆ Current consensus is secondary endpoints can be tested only after statistical significance on the primary endpoint(s)

◆ Analysis must be pre-specified

◆ Strong control of the FWER is a minimal prerequisite for confirmatory

◆ Adjustment methods:
  ◆ Sequential
  ◆ Gatekeeping (Dmitrienko, Offen and Westfall 2003, and many subsequent papers)

Proposed multiplicity adjustments for Secondary endpoints when primary is not significant:

◆ Some "key secondary" endpoints can provide clinical evidence on their own and are like primary endpoints

◆ Fallback procedures (Wiens 2003, Wiens and Dmitrienko 2005)

◆ More generally, "feedback" (Zhao, Dmitrienko and Tamura 2010), including adaptive alpha allocation

   ◆ Rewards consistency between primary and secondary endpoints

   ◆ Significance level for the secondary endpoints depends on the strength of evidence from the primary endpoints

Example: alpha level from an extended method of adaptive alpha allocation, with a single primary endpoint
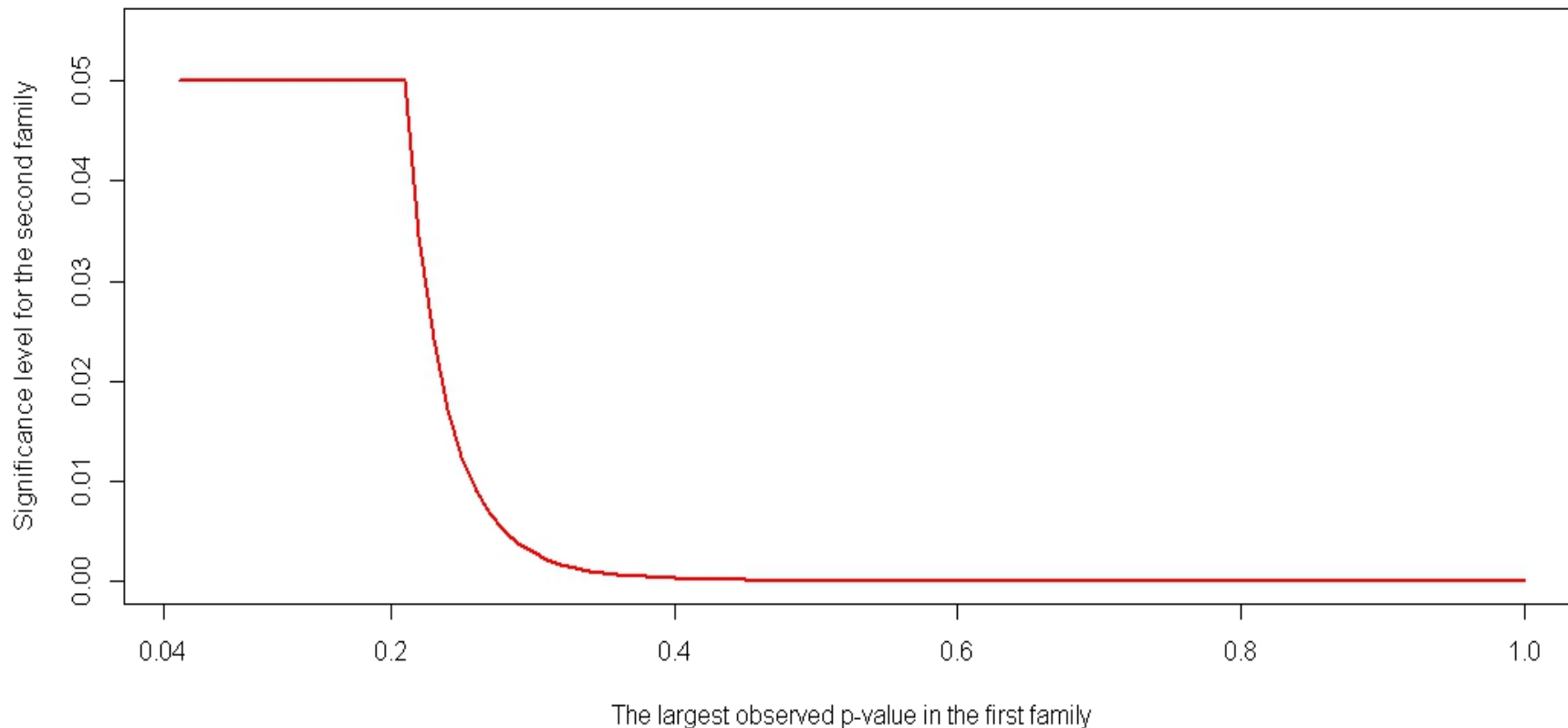
Conclusion??

◆ Failure to win on at least one primary endpoint will usually lead to a negative trial

◆ Winning on a key secondary endpoint (e.g., mortality) may "salvage" a negative trial if:

  ◆ Probability of incorrectly rejecting even a single primary or key secondary hypothesis is controlled strongly at a pre-allocated alpha level.

  ◆ Adjustment method for controlling type I error is pre-specified.

◆ Number of reasonable methods exist in the literature for strong control of type I error for testing "family" of hypotheses containing multiple primary and secondary endpoints.

Reference (1)

◆ Huque M, Alosh M. A flexible fixed-sequence testing method for hierarchically ordered correlated multiple endpoints in clinical trials. *Journal of Statistical Planning and Inference* 2008; **138**:321–335.

◆ Li J, Mehrotra DV. An efficient method for accommodating potentially underpowered primary endpoints. *Statistics in Medicine* 2008; **27**:5377–5391.

◆ Li H, Sankoh AJ, D'Agostino RB. Extension of adaptive alpha allocation methods for strong control of the family-wise error rate. *Statistics in Medicine* 2013; **32**:181–195.

◆ Alosh M, Huque M. A consistency-adjusted alpha-adaptive strategy for sequential testing. *Statistics in Medicine* 2010; **29**:1559–1571.

Reference (2)

- Huque MF, Alosh M, Bhore R. Addressing multiplicity issues of a composite endpoint and its components in clncal trials. *Journal of Biopharmaceutical Statistics* 2011; **21**:610–634.

- Moye LA. End-point interpretation in clinical trials: the case for discipline. *Controlled Clinical Trials* 1999; **20**:40–49.

- Fisher LD. Carvedilol and the Food and Drug Administration-approval process: a brief response to Professor Moye's article. *Controlled Clinical Trials* 1999; **20**:50–51.

- D'Agostino RB. Controlling alpha in clinical trials: the case for secondary endpoints. *Statistics in Medicine* 2000; **19**:762–766.

Reference (3)

◆ Dmitrienko A, Offen WW, Westfall PH. Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Statistics in Medicine* 2003; **22**:2387–2400.

◆ Moye LA. End-point interpretation in clinical trials: the case for discipline. *Controlled Clinical Trials* 1999; **20**:40–49.

◆ Wiens BL. A fixed sequence Bonferroni procedure for testing multiple endpoints. *Pharmaceutical Statistics* 2003; **2**:211–215.

◆ Wiens BL, Dmitrienko A. The fallback procedure for evaluating a single family of hypotheses. *Journal of Biopharmaceutical Statistics* 2005; **15**:929–942.

◆ Zhao YD, Dmitrienko A, Tamura R. Design and analysis considerations in clinical trials with a sensitive subpopulation. *Statistics in Biopharmaceutical Research* 2010; **2**:72–83.