Considerations in Enhancing Flexibility of Clinical Trial Design in Drug Development*

H.M. James Hung
Division of Biometrics I, Office of Biostatistics, OTS, CDER, FDA

Presented in the General Session "Flexibility in Clinical Trials: How Do We Deal With It?", FDA/ASA Workshop, Washington, DC, September 28, 2006

*The view expressed here is not necessarily of the U.S. Food and Drug Administration

Flexibility

Risk

Outline

- Some design considerations
- Some analysis considerations
- Logistics considerations
- Some review case studies
- Remarks

Planning Phase III (confirmatory) trial requires careful consideration on:
- primary endpoint, key secondary endpoints
- effect size
- dose(s)
- nuisance parameters
- dropout, missing value
- non-compliance
- statistical assumptions
……..

Recently there is fast growing interest in enhancing flexibility in clinical trial design, particularly for pivotal trial, e.g.

- interim modification of design
- combining phases II and III
- composing different stages of a pivotal trial where each stage can be planned using previous-stage data

Flexibility versus costs/risk is hardly well studied

Some Design Considerations

- Carefully explore useful information from Phase II trials, pilot trial, historical experiences, ***with some control of excessive false positives***

  Ex. Select an effect size for planning Phase III trial

  1) use CI est., rather than point est.

  2) threshold below which drop planning

  3) average power performance

Some Design Considerations

- Consider impact of necessary multiplicity adjustment

  Ex. Hierarchical testing of multiple doses

  Power of each subsequently tested dose is also a function of the true effect size of the previously tested dose(s).

  Similar problem with hierarchical testing of endpoints (1º vs. 2º)

  Difference: In the former, reallocating N is feasible and may improve power

Some Design Considerations

- Consider impact of ancillary or nuisance parameter

  Ex. Study drug effect in 2x2 factorial trial

  nuisance parameter: trt by trt interaction

  If the interim data suggests that trt by trt interaction may be of concern, can reallocating sample size to study the resulting major hypothesis help?

Some Design Considerations

- After careful design considerations as described above, some limited design modification (flexibility?) for Phase III trial probably makes sense

However, design modification based on internal data path is always of concern.
Type I error is not the only issue.

Some Analysis Considerations

Ex-1. GSD trial allowing trial extension

1º endpoint: composite of clinical outcomes (death, MI, stroke, etc.)

Sample size, total # of events may be increased, based on a new projected effect size on 1º endpt

How to test secondary endpoint to be included in product label?

- relationship in location parameters of and correlation between the endpoints
- already a tough problem with GSD
- additional complication with adaptation?

Just to see that the issue is already complicated with GSD without trial extension.

Ex-1. Compare two treatment groups

Effect size: Δ_1 (1º endpt: E1), Δ_2 (2º: E2)

Hypotheses (tested at two possible times t):

H_{0i}: $\Delta_i = 0$ vs. H_{1i}: $\Delta_i > 0$, i = 1, 2

H_0: $\Delta_1 = \Delta_2 = 0$ (global null)

Test statistic, critical value, rejection region:

T_{it}, C_{it}, $[T_{it} > C_{it}]$, i=1, 2 ; t=1, 2

Ex-1

Test 2° endpoint only after 1° endpoint wins.

2° endpoint can only be tested at most once.

1:1 sample size allocation ratio

$$\begin{pmatrix} T_{1t} \\ T_{2t} \end{pmatrix} \sim N\left(\begin{pmatrix} \sqrt{\frac{n_t}{2}}\Delta_1 \\ \sqrt{\frac{n_t}{2}}\Delta_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad t = 1, 2$$

$\rho \geq 0$

Ex-1

Global Type I error probability

global Type I error probability

≤ Type I error probability for 1º = α

Type I error probability for 2º

α_2 =: Pr(falsely rejecting H_{02} | H_{02})

= Pr($T_{11} > C_{11}$, $T_{21} > C_{21}$ | H_{02}) +

Pr($T_{11} \le C_{11}$, $T_{12} > C_{12}$, $T_{22} > C_{22}$ | H_{02})

Ex-1

(C_{11}, C_{12}) based on specified α-spending

Two intuitive strategies for testing E2:

Strategy 1

$C_{21} = C_{22} = Z_\alpha$, ($1-\alpha$)-th percentile of N(0,1)

Strategy 2

Use the same rejection boundary of E1

$C_{2t} = C_{1t}$, t = 1, 2

Ex-1

1) Type I error probability of E2, α_2 , depends on ρ and relationship between Δ_1 and Δ_2

2) When $\rho = 0$, type I error probability of E2 for Strategy 1 is

$\alpha_2 = \alpha \times$ power function of E1

$\le \alpha$

Likewise, $\alpha_2 \le \alpha$ for Strategy 2

Ex-1

3) When $\rho = 1$, type I error probability of E2

$$\alpha_2 = \Pr(\ Z_{11} > \max(C_{11} - \sqrt{0.5 n_1}\Delta_1, C_{21})\ ) + \Pr(\ Z_{11} \le C_{11} - \sqrt{0.5 n_1}\Delta_1,\ Z_{12} > \max(C_{12} - \sqrt{0.5 n_2}\Delta_1, C_{22}))$$

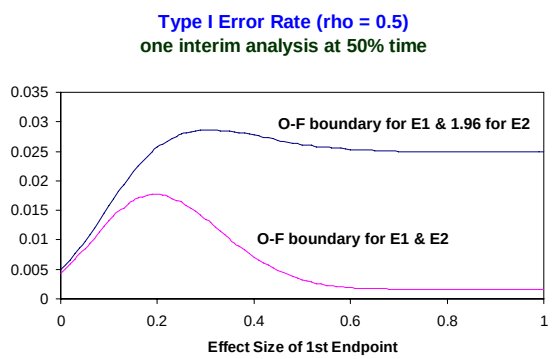
where

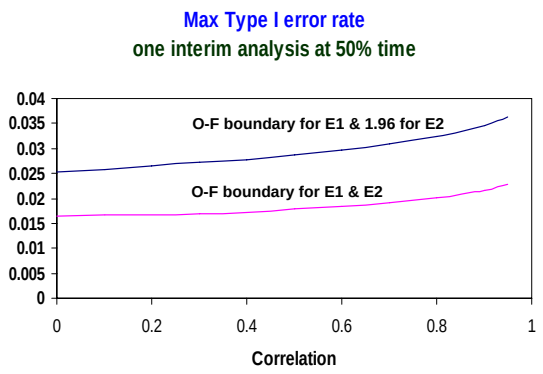
Z_{1t} is the standard Brownian Motion process

Ex-1: Type I Error Performance

Planned sample size of 200 per group
Target $\alpha = 0.025$
Strategy 1:
Use O-F α-spending for 1º endpt
Use $\alpha = 0.025$ for 2º endpt
Strategy 2:
Use O-F α-spending for both endpts

Based on 1 million replications; max type I errors are evaluated over $\Delta_1 = 0\ (0.01)\ 1$ at $\rho = 0\ (0.1)\ 1$





One interim analysis (t=.5, O-F)

Strategy 1 (use of $Z_{.025} = 1.96$ for all analyses) does not always control Type I error rate of E2. Maximum error rate over $\Delta_1 \in [0, 1]$ and $\rho \in [0, 1]$ is 0.0409

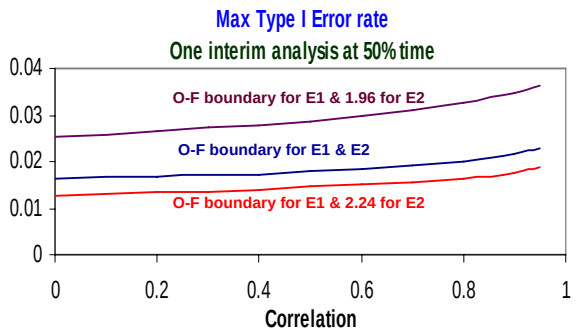
Strategy 2 (use the same alpha boundary of E1) always control Type I error rate of E2. Maximum error rate is 0.0249

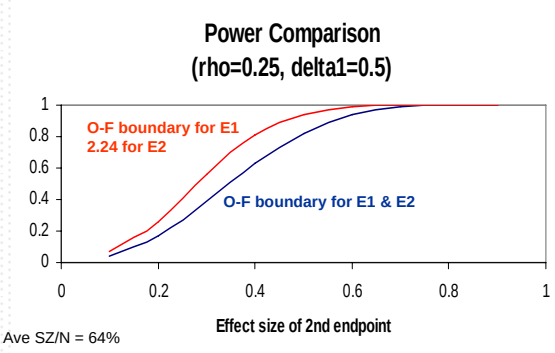
One interim analysis (t=.5, O-F)

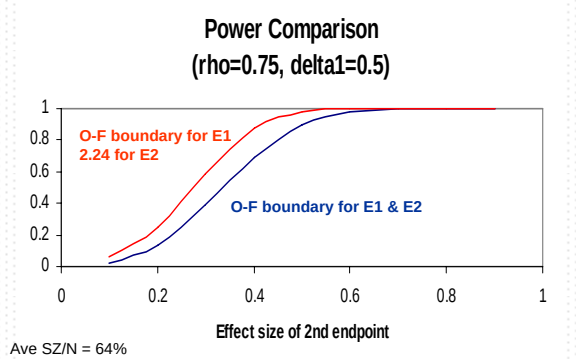
An interesting strategy

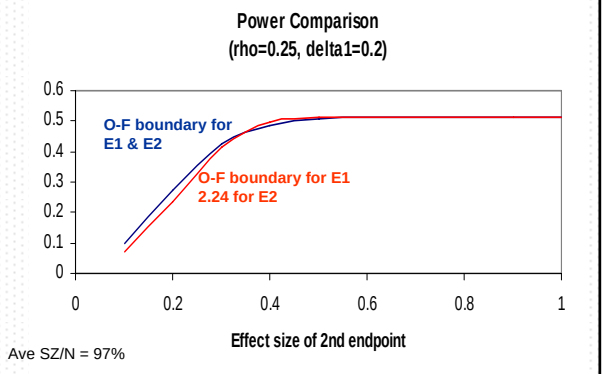
Using $Z_{0.025/2}$ = 2.24 for all analyses always control Type I error rate of E2. Maximum error rate is 0.0209, which is a little smaller than the maximum type I error rate of Strategy 2.

This strategy can be more powerful than Strategy 2 (using same alpha spending function for both endpoints).

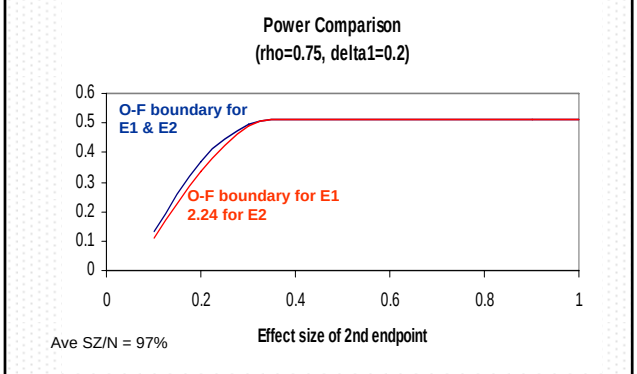








Ave SZ/N = 97%



Ave SZ/N = 97%

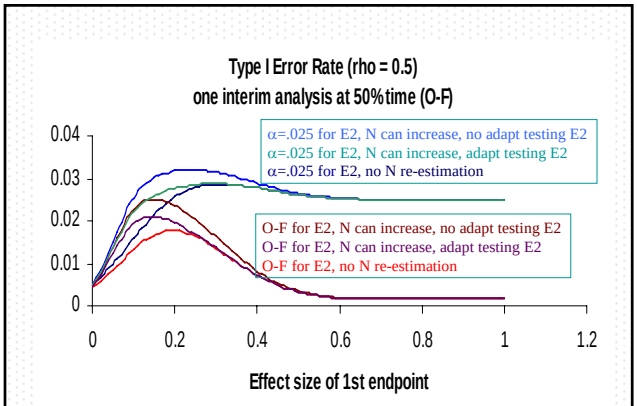
Adaptation may add inflation to the error assoc. w/ testing secondary endpoint.

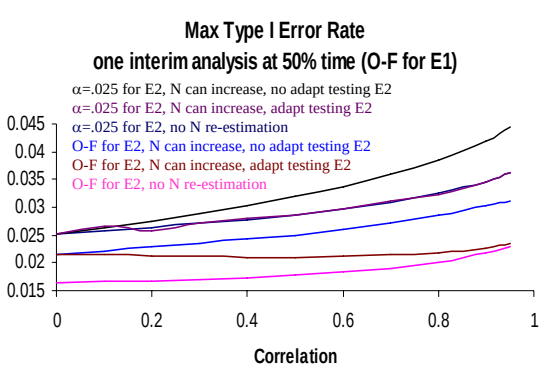
N = f(E1) = f(g(E2))
g may be an implicit function

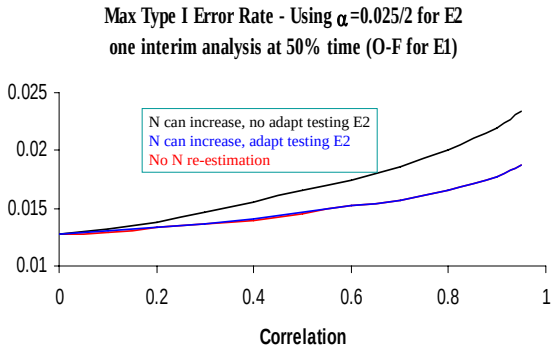
The classical Z-statistic may be biased in favor of rejecting the null hypothesis for E2 more often than it should.

Q: how much additional bias?

For illustration purpose only, assume that sample size allowed to be increased to 4 times (exaggeration) of the original sample size based on an interim observed effect size of 1º endpoint.







Some Analysis Considerations

Ex-2. Test H_0: $\Delta_1 \le 0$ or $\Delta_2 \le 0$ … or $\Delta_K \le 0$

ancillary parameter: $\lambda_{ij} = n^{1/2}(\Delta_i - \Delta_j)$

Rejection region: $\cap[Z_i > C_i]$

Type I error rate

$= \max_{(-\infty < \forall\lambda ij < \infty)} \Pr(\cap [Z_i > C_i] \mid H_0)$

How realistic can each λ_{ij} range from $-\infty$ to ∞? If not, how can we know the plausible range of λ_{ij}? Note: $|\lambda|$ involves n, so can easily be large.

Reallocating sample size per predicted values of nuisance parameters might help.

Some Analysis Considerations

Ex-3. In longitudinal clinical trials, suppose that

$E(Y_{it}) = \alpha + \beta t + \gamma Z + \theta Z*t$ t = visit time

R_{it} = 1 if measurement of patient i missing at t

0 if not missing at t

Z = independent variables

How can MAR be distinguished from MNAR? At least, implicitly, $P(R_{it} = 1 \mid Z)$ also depends on $Y_{i,t+1}$. Thus, it may also be MNAR.

Need to learn whether it is MAR or MNAR from 'external' trials

Some Review Case Studies

CS-1. Pooling trials for testing 2° endpoints

Two almost identical parallel trials are planned to study the effects of a test drug in complimentary patient populations. If the primary endpoint achieves statistical significance, then there are two positive studies to support primary endpoint.

It is planned to pool two studies for testing some key 2° endpoints.

What is experimentwise type I error rate?

Some Review Case Studies

CS-2. Phase II-III combination design

In Phase II, at least one dose will be selected based on a marker or surrogate endpoint to study in Phase III. The marker is thought to be predictive of the clinical response or the potential treatment effect on the clinical endpoint of interest. The analysis will be based on the combined data of the two phases.

Some Review Case Studiesy

CS-2. Phase II-III combination design

Major Issues

1) Alpha implications, depending on selection criteria and win criteria
2) How informative of the clinical endpoint data in Phase II?
3) Any change of other design specifications after examining Phase II marker data and clinical endpoint data?

Some Review Case Studies

CS-3. Changing focus on patient populations

A trial is planned to allow the flexibility of changing focus on patient populations, overall population vs. subpopulation, based on interim trial data and then possibly increasing sample size.

- Alpha issues are complicated
- If deciding to focus on subpopulation, analysis of overall population at the study end might not be interpretable
- Logistics issues

Logistics Considerations

Standard Operation Procedure

- which party to implement adaptation?
- who sees what?
- what knowledge of internal trial to protect from investigator/patient, sponsor management, etc.?
- how to minimize possible influence of adaptation on investigator/patient behavior?

Logistics Considerations

Compliance with SOP

- ability to comply ?
- how to ensure compliance with SOP?
- check and monitor quality of compliance
- paper trails regarding compliance data
- carrots and sticks

Confidentiality

- confidentiality to internal constituents
- confidentiality to external communities
- ability to mask ?

Remarks

- For Phases I and II trials, adaptive design is probably a good choice in the sense of better exploring drug effects without excessive false positives and with an option of dropping inferior doses to help design Phase III trial

Remarks

- For Phase III trial, **clinical trial simulations** are needed at the design stage to help with the previously laid out
  - design considerations
  - analysis considerations
- Logistics considerations is a big issue
  - trial integrity vs. flexibility
- Reviewability by regulatory agency

Remarks

- Efficiency of adaptive design paradigm vs. current non-adaptive design paradigm for entire clinical development program is unclear and needs research and practical experience
  - trade-off between flexibility, probability of success and logistical burden

Remarks

- For Phase III (or pivotal) trial planning, the document submission for review should contain:
  - pre-Phase III exploratory information
  - explanation of how the values of design parameters of phase III trial are projected (scenario planning)
  - clinical trial simulation results and computer programs

Selected References

Bauer, Köhne (1994, Biometrics)
Bauer, Röhmel (1995, Stat. In Med.)
Lan, Trost (1997, ASA Proceedings)
Proschan, Hunsberger (1995, Biometrics)
Fisher (1998, Stat. In Med.)
Shen, Fisher (1999, Biometrics)
Cui, Hung, Wang (1999, Biometrics)
Bauer, Kieser (1999, Statistics in Medicine)
Posch, Bauer (1999, Biometrical J.)
Kieser, Bauer, Lehmacher (1999, Biometrical J.)
Lehmacher, Wassmer (1999, Biometrics)
Müller & Schäfer (2001, Biometrics)
Hellmich (2001, Biometrics)
Liu & Chi (2001, Biometrics)

Wang, Hung, Tsong, Cui (2001, Statistics in Medicine)
Bauer, Brannath, Posch (2002, Method Inform Med)
Brannath, Posch, Bauer (2002, JASA)
Cui (2002, Encyclopedia of Biopharmaceutical Statistics)
Li, Shih, Xie, Lu (2002, Biostatistics)
Lawrence (2002, J. of Biopharmaceutical Statistics)
Lawrence (2002, J. of Pharmaceutical Statistics)
Lawrence & Hung (2002, Biometrical J.)
Posch, Bauer & Brannath (2003, Statistics in Medicine)
Jennison & Turnbull (2003, Statistics in Medicine)
Tsiatis & Mehta (2003, Biometrika)
Liu, Proschan, Pledger (2003, JASA)
Proschan, Liu, Hunsberger (2003, Statistics in Medicine)
Chen, DeMets, Lan (2003, Controlled Clinical Trials)
Brannath, Bauer, Maurer, Posch (2003, Biometrics)
Chen, DeMets, Lan (2004, Statistics in Medicine)

Wang, Hung, O'Neill (2004, ASA Proceedings)
Hung, Wang, O'Neill (2004, ASA talk)
Liu (2002, 2003, 2004, ASA talks, other talks)
Lan (2000-2005 ASA, ICSA, FDA talks)
Mehta (2004, talk on FDA-MIT workshop)
Hung, Wang, O'Neill (2004, BASS talk)
Lan, Soo, Siu, Wang (2005, J. of Biopharmaceutical Statistics)
Hung, Cui, Wang, Lawrence (2005, J. of Biopharmaceutical Statistics)
Brannath, König, Posch, Bauer (2005, SOCT talk)
Hung, Wang, O'Neill, Lawrence (2005, SOCT talk)
Anderson, Zhang (2005, SOCT talk)
Hung, Wang, Lawrence, Tsong, O'Neill (2005, WNAR talk)
Shih, Lai, Zhu(2005, WNAR talk)
Liu, Cui (2005, WNAR talk)

Shen et al (2005, WNAR talk)
Wang, Hung (2005, ASA talk)
O'Neill (2005, MCP keynote speech)
Posch et al (2005, FDA/Industry talk)
Hung, Wang, O'Neill, Lawrence (2005, FDA/Industry talk)
Schinder & Kram (2005, FDA/Industry talk)
Dragalin (2005, FDA/Industry talk)
Chen, Hung (2006, ICSA talk)
Shen (2006, ICSA talk)
Hung (2006, IBC talk)
Hung (2006, ASA talk)

Acknowledgments

Sue-Jane Wang (OB/OTS/CDER/FDA)
Robert O'Neill (OB/OTS/CDER/FDA)

This presentation is partly based on the research supported by FDA/CDER RSR funds, #06-14, 05-02, 05-14, 04-06, 02-06, 99/00-008, 96-010A