Sequential design approaches for bioequivalence studies with crossover designs. Pharmaceutical Statistics. (Potvin et al 2008: Pharm. Stat. 7:245–262)

Additional results for 'Sequential design approaches for bioequivalence studies with crossover designs'. (Montague et al 2011: Pharm. Stat. online)

PSI Journal Club
21-July-2011
Timothy H Montague
Director Clinical Statistics, GlaxoSmithKline

Standard Bioequivalence (BE) Assessment

- Standard two-way crossover design
- Two one-sided *t*-test (type I error of 5%)
  - 90% CI Test/Reference within 80-125%
- Planning sample size based on:
  - Within-subject variance
  - Effect Size (e.g. 95-105.3%)
  - Power (e.g. 80%)
- Prior information on variance and/or effect size being poor or nonexistent leads to inaccurate sample size
  - If variance chosen is too low or the effect size overly optimistic
    - Study underpowered $\Rightarrow$ failing BE

Methods for when prior information on variance poor or nonexistent

- Add-on designs (TPD, WHO)
- Plan another BE study with more subjects (CDER-FDA)

- Pilot $\Rightarrow$ pivotal
  - Pool data (“double dip method”)
  - Never pool data

- Group Sequential Design
- Sample size re-estimation design

- Question: What about a group sequential design with sample size re-estimation?

The Product Quality Research Institute (PQRI) Initiative

- Adaptive sample size sequential method
  - Final sample size required can be re-evaluated after first stage
  - Endpoint evaluated more than once with early stopping when criteria met
- Properties considered
  - Overall type I error ≤ 5.2%
  - # of stages = 2
  - Allow stopping after each stage if criteria met
  - No blinding
  - Provide unique unambiguous result
- Four methods explored
  - Sample size re-estimation (Method A)
  - Group sequential design with sample size re-estimation (Methods B, C & D)

Potvin et al (2008) - Methods

- Sample size re-estimation (Method A)

Evaluate power at stage 1 using α-level of 0.05

If power ≥ 80%, evaluate BE at stage 1(α = 0.05) and stop

Pass or fail

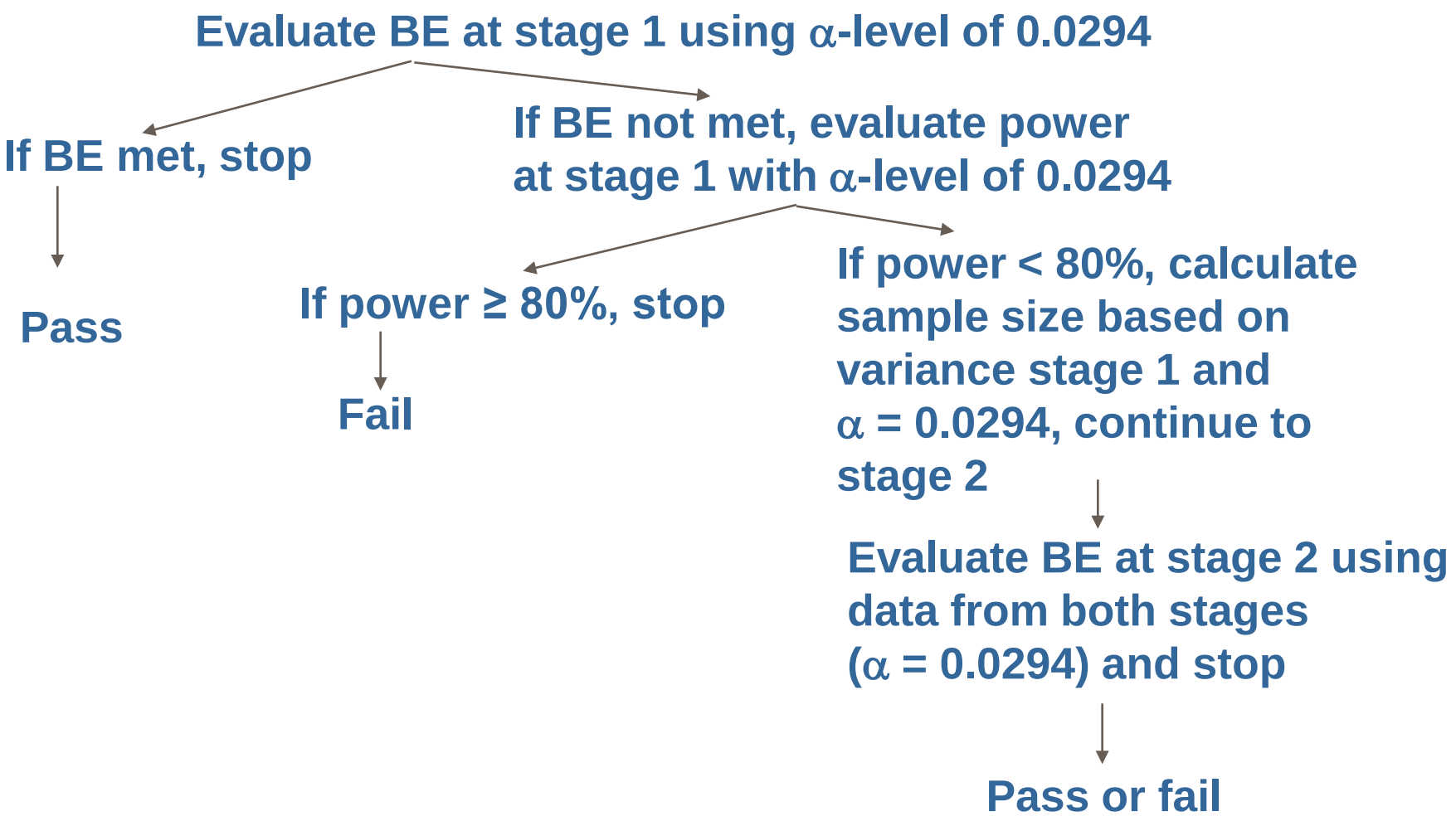
If power < 80%, calculate sample size based on variance stage 1 and α = 0.05, continue to stage 2

Evaluate BE at stage 2 using data from both stages (α = 0.05) and stop

Pass or fail

Potvin et al (2008) - Methods

- Adaptive sample size sequential design based on Pocock (Method B)



Potvin et al (2008) - Methods

- Adaptive sample size sequential design based on Pocock (Method C)

Evaluate power at stage 1 using α-level of 0.05

If power ≥ 80%, evaluate BE at stage 1(α = 0.05) and stop

Pass or fail

If power < 80%, evaluate BE at stage 1 (α = 0.0294)

If BE met, stop

Pass

If BE not met, calculate sample size based on stage 1 and α = 0.0294, continue to stage 2

Evaluate BE at stage 2 using data from both stages (α = 0.0294) and stop

Pass or fail

Method D: α = 0.028

Potvin et al (2008) - Methods

- Simulation Methodology
  - Two-way crossover studies, 2 stages
  - Individual ln(T) – ln(R) were simulated
  - Distribution: Normal ($\ln(\theta)$, $2\sigma^2$)
  - θ is the true ratio of T/R geometric means
  - σ^2 is the true intra-subject variance of the drug
  - CV is defined as

$$\text{intra}-\text{subject CV}(\%) \quad = \quad 100\sqrt{e^{\sigma^2}-1}$$

  - Even n_1 and n_2
  - Two one-sided *t*-test
  - Variance estimate at stage 1 based on the standard GLM ANOVA model
  - Variance estimate at stage 2 based on model with *Sequence, Stage, Period(Stage), Treatment, Subject(Sequence x Stage),*

Potvin et al (2008) - Methods

- Simulation Methodology

  - Compaq Visual Fortran (6.1.0)
  - Different randomly selected seed for each scenario
  - Scenario: combination of θ, CV, n_1 and method
  - 1 million studies per scenario
  - θ = 0.95 (power) and 1.25 (Type I error)
  - CV = 0.10 ; 0.20 ; 0.30; … 1.00 (Method A: only 0.20 and 0.30 evaluated)
  - Power = 80%
  - Type 1 error (0.05, 0.0294, 0.028)
  - n_1 = 12, 24, 36, 48, 60

Potvin et al (2008) - Results

- Sample Size Re-estimation Method A (Type I error)
  - Type I error inflated in certain scenario
    - When n_1=8, type I error up to 0.07
    - When n_1=12, inflation of 16% (to 0.058)
  - Sample size re-estimation method must include some adjustment to preserve overall type I error (Stein's method)
- Adaptive Sample Size Sequential Methods B, C, D
  - All methods control Type I error ($\leq$ 0.052)
  - Method B is more conservative than method C & D for small CVs (10-30%).
  - All methods maintain power reasonably ($\geq$ 70%)
    - Greatest loss occurs for small n1 and larger CVs
  - Methods C & D less likely to go to stage 2 (for reasonable CVs)
  - There is generally some cost to using a 2-stage design when n1 is small

Potvin et al (2008) - Conclusions

- PQRI goal was to identify and validate a method for adaptive sample size sequential method in BE.
- Not to find the best one
- Methods B, C, D met our criteria of $\alpha \leq 0.052$
- Method D was more conservative (average total n larger) as compared to B and C.
- For sponsor, small power advantage of Method C over B.
- Method C is coming back to a one-stage BE study if adequate power at first stage.
- The intention of using such a method must be specified in the protocol, a priori.

Potvin et al – informal feedback

- Can the results be extended to a wider range of cases than covered in Potvin et al (2008)?

- Potvin et al (2008): For purposes of estimating the power and the sample size, the true ratio of T/R geometric mean (GMR) was assumed to be 0.95

Montague et al (2011)

- Explores the operating characteristics of the three group sequential designs (Methods B, C & D) assuming the **GMR (θ) = 0.90**.
  - Methodology is the same other than simulations are run in R.
    - Code is validated by re-creating results from Potvin et al 2008.
- Results
  - Type I error inflated (> 0.052) for some scenarios for methods B and C when CV $\leq$ 50%
  - Type I error controlled for method D ($\leq$ 0.052).
  - Other design characteristics are similar to those when θ = 0.95
    - Although average sample sizes are slightly larger
- Conclusions
  - Method B and C may result in excessive inflation of Type I error rate in some cases when θ = 0.90. Thus should be avoided
  - Method D could be used for θ between 0.90 and 0.95, but has not been validated for values beyond this range.
  - How these methods behave when desired power is 90% rather than 80% has not been explored and thus is unknown.

Acknowledgments

- Diane Potvin
- Charles E. DiLiberti
- Walter W. Hauck
- Alan F. Parr
- Donald J. Schuirmann
- Robert A. Smith
- PQRI
- Jean Lavigne
- Tingting Zhan

References

1. Food and Drug Administration, Center for Drug Evaluation and Research (CDER). Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products — General Considerations. March 2003.
2. Health Canada, Ministry of Health – Health Products and Food Branch. Guidance for Industry: Conduct and Analysis of Bioavailability and Bioequivalence Studies—Part A: Oral Dosage Formulations Used for Systemic Effects, 1992.
3. World Health Organization Expert Committee on Specifications for Pharmaceutical Preparations. Fortieth Report, Annex 7, Multisource (generic) pharmaceutical products: guidelines on registration requirements to establish interchangeability. WHO Technical Report 937 2006; 347-390.
4. Hauck WW, Preston PE and Bois FY. A group sequential approach to crossover trials for average bioequivalence. *Journal of Biopharmaceutical Statistics* 1997; **7:**87-96.
5. Lan KKG and Demets DL. Discrete sequential boundaries for clinical trials. *Biometrika* 1983; **70:**659-663.
6. O'Brien PC and Fleming TR. A multiple testing procedure for clinical trials. *Biometrika* 1979; **35:**549-556.
7. Pocock, SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977; **64:**191-199.
8. Gould, AL. Group sequential extensions of a standard bioequivalence testing procedure. *Journal of Pharmacokinetics and Biopharmaceutics* 1995; **23**:57-86.
9. Jennison C, Turnbull BW. Sequential equivalence testing and repeated confidence intervals, with application to normal and binary responses. *Biometrics* 1993; **40**:225-230.
10. Birkett MA and Day SJ. Internal pilot studies for estimating sample size. *Statistics in Medicine* 1994; **13**:2455-2463.
11. Coffey CS and Muller KE. Exact test size and power of a Gaussian error linear model for an internal pilot study. *Statistics in Medicine* 1999; **18**:1199-1214.

References

12. Coffey CS and Muller KE. Controlling test size while gaining benefits of an international pilot design. *Biometrics* 2001; **57**, 625-631.
13. Denne JS and Jennison C. Estimating the sample size for a *t*-test using an internal pilot. *Statistics in Medicine* 1999; **18**:1575-1585.
14. Kieser M and Friede T. Re-calculating the sample size in internal pilot designs with control of type I error rate. *Statistics in Medicine* 2000; **19**:901-911.
15. Stein C. A two-sample test for a linear hypothesis whose power is independent of the variance. *Annals of Mathematical Statistics* 1945; **16**:243-258.
16. Wittes J and Brittain E. The role of internal pilot studies in increasing the efficacy of clinical trials. *Statistics in Medicine* 1990; **9**:65-72.
17. Wittes J, Schabenberger O, Zucker D, Brittain E and Proschan M. Internal pilot studies I: Type I error rate of the naïve *t*-test. *Statistics in Medicine* 1999; **18**:3481-3491.
18. Zucker DM, Wittes JT, Schabenberger O and Brittan E. Internal pilot studies II: Comparison of various procedures. *Statistics in Medicine* 1999; **18**:3493-3509.
19. Cui L, Hung MJ, and Wang S-J. Modification of sample size in group sequential clinical trials. *Biometrics* 1999; **55:**853-857.
20. Jennison C and Turnbull BW. Mid-course sample size modification in clinical trials based on the observed treatment effect. *Statistics in Medicine* 2003; **22**:971-993.
21. Chen YHJ, DeMets DL and Lan KKG. Increasing the sample size when the unblinded interim result is promising. *Statistics in Medicine* 2004; **23**:1023-1038.
22. Schuirmann DJ A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm* 1987; **15**:657-80.
23. Hauschke D, Steinijans VW, Diletti E And Burke M. Sample size determination for bioequivalence assessment using a multiplicative model. *Journal of Pharmacokinetics and Biopharmaceutics* 1992; **20**:557-561.

Backups

Potvin et al (2008) – Method A results (Type I error)

		Intra-subject CV (%)		
		10	**20**	**30**
n1	**12**		0.0584	0.0575
	24		0.0505	0.0550
	36		0.0497	0.0523
	48		0.0500	0.0502
	60		0.0500	0.0498

Source: Table 1 of Potvin et al (2008)

Potvin et al (2008) – Method B Results (Type I error)

		Intra-subject CV (%)									
		10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
n1	**12**	0.0297	0.0463	0.0437	0.0344	0.0309	0.0297	0.0294	0.0292	0.0289	0.0291
	24	0.0294	0.0320	0.0475	0.0433	0.0338	0.0307	0.0299	0.0298	0.0298	0.0298
	36	0.0294	0.0294	0.0397	0.0485	0.0420	0.0333	0.0306	0.0303	0.0296	0.0298
	48	0.0292	0.0292	0.0324	0.0458	0.0484	0.0399	0.0328	0.0303	0.0297	0.0297
	60	0.0294	0.0297	0.0296	0.0409	0.0483	0.0466	0.0381	0.0318	0.0300	0.0301

Source: Table 1 of Potvin et al (2008)

Potvin et al (2008) – Method C Results (Type I error)

		Intra-subject CV (%)									
		10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
n1	**12**	0.0496	**0.0510**	0.0441	0.0346	0.0311	0.0299	0.0294	0.0292	0.0285	0.0290
	24	0.0500	0.0490	0.0492	0.0435	0.0339	0.0307	0.0298	0.0301	0.0298	0.0295
	36	0.0500	0.0499	0.0477	0.0489	0.0418	0.0331	0.0308	0.0299	0.0296	0.0297
	48	0.0501	0.0495	0.0494	0.0469	0.0480	0.0399	0.0325	0.0302	0.0298	0.0297
	60	**0.0504**	0.0500	0.0502	0.0470	0.0483	0.0472	0.0380	0.0319	0.0301	0.0297

Source: Table 1 of Potvin et al (2008)

Potvin et al (2008) – Method D Results (Type I error)

		Intra-subject CV (%)									
		10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
n1	**12**	0.0498	0.0499	0.0415	0.0330	0.0296	0.0284	0.0280	0.0276	0.0278	0.0274
	24	0.0500	0.0493	0.0475	0.0408	0.0319	0.0290	0.0287	0.0285	0.0284	0.0283
	36	**0.0504**	0.0499	0.0471	0.0464	0.0395	0.0315	0.0290	0.0286	0.0284	0.0282
	48	0.0502	0.0497	0.0495	0.0455	0.0456	0.0373	0.0306	0.0288	0.0287	0.0285
	60	0.0501	0.0500	0.0499	0.0456	0.0461	0.0442	0.0359	0.0301	0.0286	0.0281

Source: Table 1 of Potvin et al (2008)

Montague et al (2011) – Method B Results (Type I error)

		Intra-subject CV (%)									
		10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
n1	**12**	0.0336	**0.0538**	0.0437	0.0340	0.0309	0.0298	0.0297	0.0296	0.0298	0.0295
	24	0.0295	0.0471	**0.0529**	0.0439	0.0333	0.0305	0.0297	0.0299	0.0296	0.0292
	36	0.0295	0.0386	**0.0512**	**0.0529**	0.0427	0.0329	0.0303	0.0297	0.0297	0.0295
	48	0.0295	0.0313	0.0483	**0.0527**	**0.0513**	0.0403	0.0324	0.0300	0.0297	0.0296
	60	0.0294	0.0294	0.0449	**0.0509**	**0.0526**	0.0487	0.0384	0.0318	0.0300	0.0297

Source: Table 1 of Montague et al (2011)

Montague et al (2011)– Method C Results (Type I error)

		Intra-subject CV (%)									
		10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
n1	**12**	0.0499	**0.0547**	0.0437	0.0335	0.0309	0.0301	0.0298	0.0299	0.0298	0.0296
	24	0.0501	0.0497	**0.0531**	0.0438	0.0332	0.0306	0.0295	0.0296	0.0298	0.0293
	36	0.0499	0.0480	**0.0513**	**0.0526**	0.0427	0.0329	0.0299	0.0297	0.0296	0.0300
	48	0.0502	0.0499	0.0485	**0.0522**	**0.0514**	0.0407	0.0323	0.0302	0.0296	0.0295
	60	0.0503	**0.0504**	0.0466	**0.0509**	**0.0530**	0.0486	0.0384	0.0318	0.0298	0.0296

Source: Table 1 of Montague et al (2011)

Montague et al (2011)– Method D Results (Type I error)

		Intra-subject CV (%)									
		10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
n1	**12**	0.0498	**0.0518**	0.0414	0.0322	0.0293	0.0286	0.0286	0.0280	0.0281	0.0282
	24	0.0501	0.0475	**0.0506**	0.0414	0.0316	0.0292	0.0281	0.0283	0.0281	0.0282
	36	**0.0504**	0.0477	0.0489	0.0502	0.0401	0.0313	0.0288	0.0287	0.0282	0.0282
	48	0.0503	0.0499	0.0470	0.0500	0.0484	0.0381	0.0307	0.0284	0.0281	0.0282
	60	0.0498	0.0498	0.0449	0.0492	**0.0509**	0.0462	0.0358	0.0301	0.0285	0.0279

Source: Table 1 of Montague et al (2011)