Testing for bioequivalence of highly variable drugs from TR-RT crossover designs with heterogeneous residual variances

Christopher I. Vahl, PhD
Department of Statistics
Kansas State University

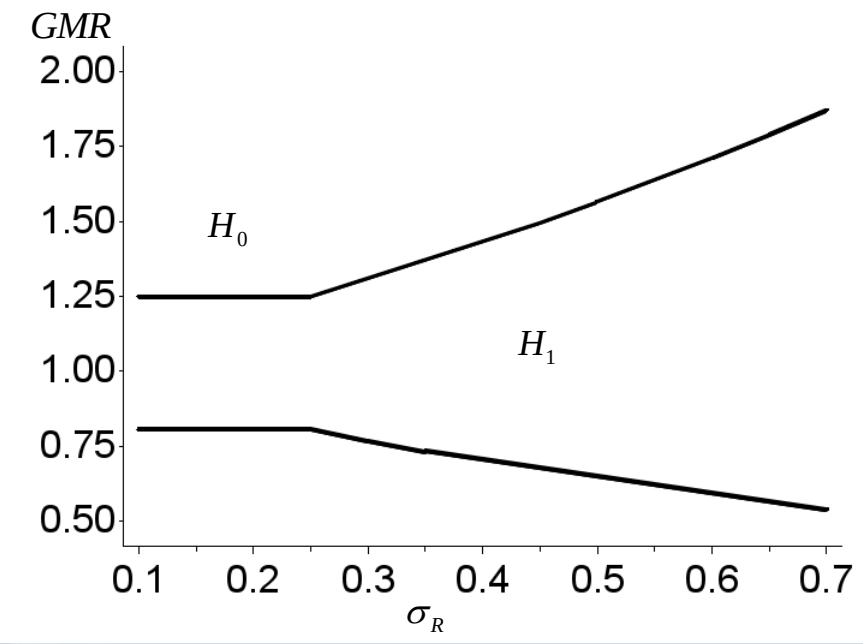
Qing Kang, PhD
Chief Statistical Scientist
The Statistical Intelligence Group, LLC

Introduction

Traditional bioavailability (BA) studies assess average bioequivalence (ABE) between the test (T) and reference (R) products under the TR-RT crossover design.

With highly variable (HV) drugs whose within-subject coefficient of variation (CV) in pharmacokinetic measures is ≥30%, assertion of ABE becomes difficult.

In 2011, the FDA adopted a BE criterion with mixed-scaling. For HV drugs, the equivalence limits for the geometric mean ratio (GMR) are scaled to within-subject variance of the reference product. This is known as reference scaled ABE (RSABE).



$$H_0 : (\mu_T - \mu_R)^2 \geq k \max(0.25^2, \sigma_R^2) \quad \text{vs.}$$
$$H_1 : (\mu_T - \mu_R)^2 < k \max(0.25^2, \sigma_R^2)$$

$$\mu_T - \mu_R = \ln(\text{T:R GMR});$$
$$\sigma_R^2 = \ln(CV_R{}^2 + 1);$$
$$k = [\ln(1.25)]^2 / 0.25^2 .$$

Introduction

The FDA extended the statistical methods for assessing individual BE (IBE) to testing RSABE. The recommended procedure operates exclusively under TRR-RTR-RRT and TRTR-RTRT designs.

Testing IBE calls for separate estimation of subject×formulation variance and within-subject variances, which could only be achieved by replicate crossover designs. In 2003, the FDA discontinued the IBE criterion due to the lack of evidence confirming the existence of subject×formulation.

Designs with more than 2 periods are not always feasible.

- The volume of blood taken from each subject may exceed the acceptable limit.
- They tend to have a large amount of missing data and a high dropout rate.
- To avoid carryover effect, the washout period lasts for ≥5 half lives. A 2-period design is more practical for drugs with a long half life.
- Subject's physiological changes affect the variability of systemic drug concentration. A lengthy study could not be conducted on growing animals or those susceptible to stress under prolonged confinement and repeated dosing.

Goal: To investigate how to evaluate HV drugs under the TR-RT design.

Model

Consider a TR-RT design where the washout time between periods is sufficient to eliminate any carryover effect.

Sequence (i)	No. of Subjects*	Treatment	
		Period 1	Period 2
1	n_1	T	R
2	n_2	R	T

*$n_1 + n_2 = N$

Subject *j* in sequence *i* provides a vector of ln-transformed responses (Y_{ij1} ,Y_{ij2}).

Alternative BE criteria under the TR-RT design were assessed using the model with heterogeneous residual variance.

$$\begin{pmatrix} Y_{ij1} \\ Y_{ij2} \end{pmatrix} = \begin{cases} \eta_i \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} + \begin{pmatrix} \mu_T \\ \mu_R \end{pmatrix} + \boldsymbol{\tau}_{ij} + \begin{pmatrix} \varepsilon_{ijT} \\ \varepsilon_{ijR} \end{pmatrix} & i = 1 \\ \eta_i \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} + \begin{pmatrix} \mu_R \\ \mu_T \end{pmatrix} + \boldsymbol{\tau}_{ij} + \begin{pmatrix} \varepsilon_{ijR} \\ \varepsilon_{ijT} \end{pmatrix} & i = 2 \end{cases}$$

η_i : sequence effect;

π_j : period effect;

μ_T & μ_R : product effects;

$\boldsymbol{\tau}_{ij}$: subject effect, $\boldsymbol{\tau}_{ij} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_S^2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right)$;

ε_{ijT} & ε_{ijR} : residual errors, $\varepsilon_{ijT} \sim N(0, \sigma_T^2), \varepsilon_{ijR} \sim N(0, \sigma_R^2)$;

$\boldsymbol{\tau}_{ij}$, ε_{ijT} & ε_{ijR} are mutually indept.

Model

The FDA's model for replicate crossover designs allows $\boldsymbol{\tau}_{ij}$ to have a covariance matrix of any positive definite structure.

$$\boldsymbol{\tau}_{1j} \sim \mathrm{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\sigma_{ST}^2 & \rho\sigma_{ST}\sigma_{SR}\\ \rho\sigma_{ST}\sigma_{SR} & \sigma_{SR}^2\end{pmatrix}\right), \boldsymbol{\tau}_{2j} \sim \mathrm{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\sigma_{SR}^2 & \rho\sigma_{ST}\sigma_{SR}\\ \rho\sigma_{ST}\sigma_{SR} & \sigma_{ST}^2\end{pmatrix}\right)$$

The variance of subject×formulation is

$$\sigma_{Int}^2 = (\sigma_{ST} - \sigma_{SR})^2 + 2(1-\rho)\sigma_{ST}\sigma_{SR}$$

The TR-RT design confounds σ_{Int}^2 with within-subject variances. But when there is no subject×formulation,

$$\sigma_{Int}^2 = 0 \Leftrightarrow \begin{matrix}\sigma_{ST}^2 = \sigma_{SR}^2 = \sigma_S^2\\ \rho = 1\end{matrix} \Leftrightarrow \boldsymbol{\tau}_{ij} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \sigma_S^2\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\right)$$

It is then plausible to assess RSABE from the TR-RT design.

Summary Statistics

Convert (Y_{ij1} , Y_{ij2}) into within-subject sum and T-R difference.

$$Y_{ij+} = Y_{ij1} + Y_{ij2} \qquad Y_{ij-} = \begin{cases} Y_{ij1} - Y_{ij2} & i = 1 \\ Y_{ij2} - Y_{ij1} & i = 2 \end{cases}$$

$$Var\begin{pmatrix} Y_{ij+} \\ Y_{ij-} \end{pmatrix} = \begin{pmatrix} 4\sigma_S^2 + \sigma_{TR}^2 & \beta\sigma_{TR}^2 \\ \beta\sigma_{TR}^2 & \sigma_{TR}^2 \end{pmatrix}$$

$$\beta = (\sigma_T^2 - \sigma_R^2)/(\sigma_R^2 + \sigma_T^2) \qquad \sigma_{TR}^2 = \sigma_T^2 + \sigma_R^2$$

Define the summary statistics with respect to (Y_{ij+} , Y_{ij-}) as

$$\bar{Y}_{i+} = \sum_{j=1}^{n_i} Y_{ij+} \Big/ n_i \qquad \bar{Y}_{i-} = \sum_{j=1}^{n_i} Y_{ij-} \Big/ n_i \qquad \bar{D} = (\bar{Y}_{1-} + \bar{Y}_{2-})/2$$

$$S_+ = \sum_{i=1}^{2}\sum_{j=1}^{n_i} (Y_{ij+} - \bar{Y}_{i+})^2 \qquad S_- = \sum_{i=1}^{2}\sum_{j=1}^{n_i} (Y_{ij-} - \bar{Y}_{i-})^2 \qquad S_{-+} = \sum_{i=1}^{2}\sum_{j=1}^{n_i} (Y_{ij-} - \bar{Y}_{i-})(Y_{ij+} - \bar{Y}_{i+})$$

$$B = S_{-+} \Big/ S_- \qquad S_{+|-} = S_+ - S_{-+}^2 \Big/ S_-$$

Note :

- $\bar{D}$ and $S_-/[2(N-2)]$ correspond the LS mean difference and MSE in the classical ANOVA model;
- B and $S_{+|-}$ represent the slope and the SSE when regressing Y_{ij+} on Y_{ij-} with the same slope but different intercept for i=1,2.

Summary Statistics

By Rao(1973), these summary statistics have the distributional properties of

$$\frac{\sqrt{4n_1n_2}}{\sqrt{N\sigma_{TR}^2}}[\bar{D}-(\mu_T-\mu_R)] \equiv Z_\mu \sim N(0,1) \qquad \frac{S_-}{\sigma_{TR}^2} \equiv U_- \sim \chi^2(N-2)$$

$$\frac{1}{\sigma_{+|-}}(B-\beta)\sqrt{S_-} \equiv Z_\beta \sim N(0,1) \qquad \frac{S_{+|-}}{\sigma_{+|-}^2} \equiv U_{+|-} \sim \chi^2(N-3)$$

$$\sigma_{+|-}^2 = 4\sigma_S^2 + \sigma_{TR}^2(1-\beta^2) : \text{conditional variance of } Y_{ij+} \text{ given } Y_{ij-}$$

Note:

- Z_μ, Z_β, U_-, and $U_{+|-}$ are mutually independent;
- B is not independent of S_- and its marginal distribution is not normal;
- The estimator for $\mu_T-\mu_R$, β and σ_{TR}^2 are $\bar{D}$, B and $S_-/(N\text{-}2)$;
- The estimator for $\sigma_R^2 = 0.5(1-\beta)\sigma_{TR}^2$ is

$$0.5(1-B)\,S_-/(N-2)$$

whose distribution does not follow any classic form.

Testing Procedures

The current test of RSABE employs the modified large sample (MLS) method which is applicable under replicate crossover designs.

The MLS method approximates the confidence limits (CLs) of a linear combination of parameters by restricting it to be exact when only one parameter is unknown. It calls for independent summary statistics with known distributions.

Let b, $s_{\_}$ and $\bar{d}$ be the values of B, $S_{\_}$ and $\bar{D}$ observed from the TR-RT design.

The estimate of σ_R^2 is

$$\tilde{\sigma}_R^2 = \frac{1}{2N-4}(1-b)s_{\_}$$

and the CLs for $\mu_T - \mu_R$ is

$$\bar{d} \pm t_{0.95,N-2}\sqrt{\frac{N}{4n_1n_2(N-2)}s_{\_}}$$

The MLS method is inappropriate under the TR-RT design because σ_R^2 is not estimated from independent summary statistics with classic marginal distributions.

Testing Procedures

The generalized pivotal quantity (GPQ) method is used to test RSABE under the TR-RT design. The distribution of the GPQ produces a fiducial-type inference which, in many situations, meets frequentists' standards.

The distributional properties of the summary statistics suggest that the GPQ for $\mu_T-\mu_R$, σ_{TR}^2, and β are

$$T_1 = \bar{d} + \frac{\sqrt{4n_1n_2}}{\sqrt{N\sigma_{TR}^2}}(\mu_T - \mu_R - \bar{D})\sqrt{\frac{N}{4n_1n_2}\sigma_{TR}^2\frac{s_-}{S_-}} = \bar{d} + Z_\mu\sqrt{\frac{N}{4n_1n_2}\frac{s_-}{U_-}}$$

$$T_2 = \frac{s_-}{S_-}\sigma_{TR}^2 = \frac{s_-}{U_-}$$

$$T_3 = b - \frac{1}{\sqrt{\sigma_{+|-}^2}}(B-\beta)\sqrt{S_-}\sqrt{\frac{s_{+|-}}{S_{+|-}}\frac{\sigma_{+|-}^2}{s_-}} = b - Z_\beta\sqrt{\frac{s_{+|-}}{s_-\,U_{+|-}}}$$

The GPQ for $(\mu_T-\mu_R)^2 - k\sigma_R^2 = (\mu_T-\mu_R)^2 - 0.5k\sigma_{TR}^2(1-\beta)$ is then assembled as

$$T_1^2 - 0.5kT_2(1-T_3)$$

Testing Procedures

The distribution of this GPQ is obtained via a resampling algorithm.

Step 1. Independently sample z_μ, u_-, z_β and $u_{+|-}$ and from $N(0,1)$, χ^2 $(N-2)$, $N(0,1)$ and χ^2 $(N-3)$.

Step 2. Replace Z_μ, U_-, Z_β, and $U_{+|-}$ in T_1, T_2 and T_3 with z_μ, u_-, z_β and $u_{+|-}$. This yields t_1, t_2 and t_3.

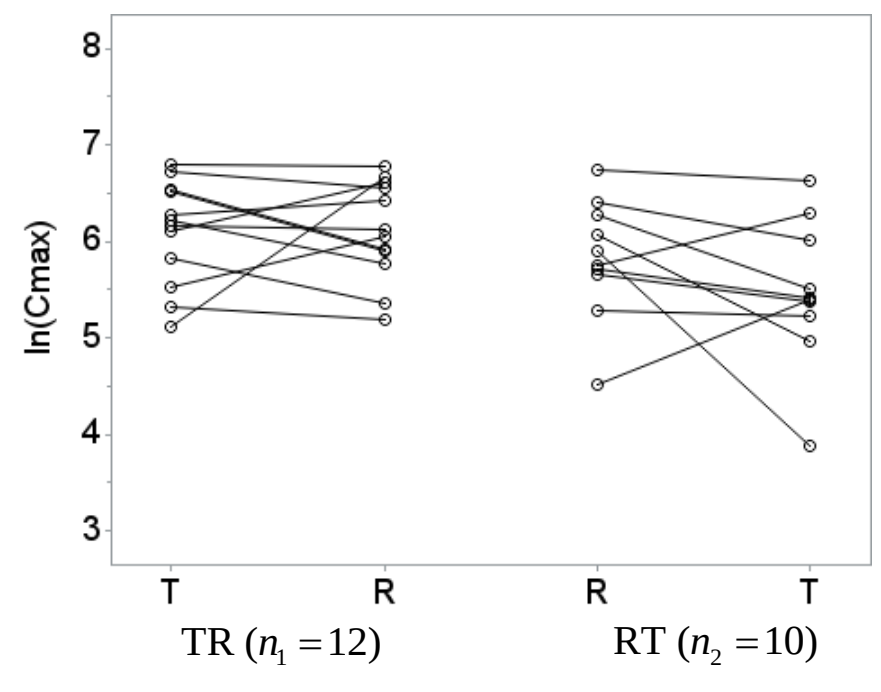
Step 3. Calculate $t_1^2 - 0.5kt_2^2(1-t_3)$ and accumulate its value by repeating steps1 and 2 many times.

The 95% upper CL of $(\mu_T - \mu_R)^2 - k\sigma_R^2$ is given by the 95th percentile of the resampling distribution. The test product passes RSABE when this percentile is ≤ 0.

Analysis of Example Dataset

Source: Dataset #7 in FDA's databank for BA studies.

Data: Cmax measured in the first two periods of a TRTR-RTRT-RTTR-TRRT design.



$\bar{d} = 0.1724$ $\quad s_+ = 19.5285$

$s_- = 10.1230$ $\quad s_{-+} = -1.4336$

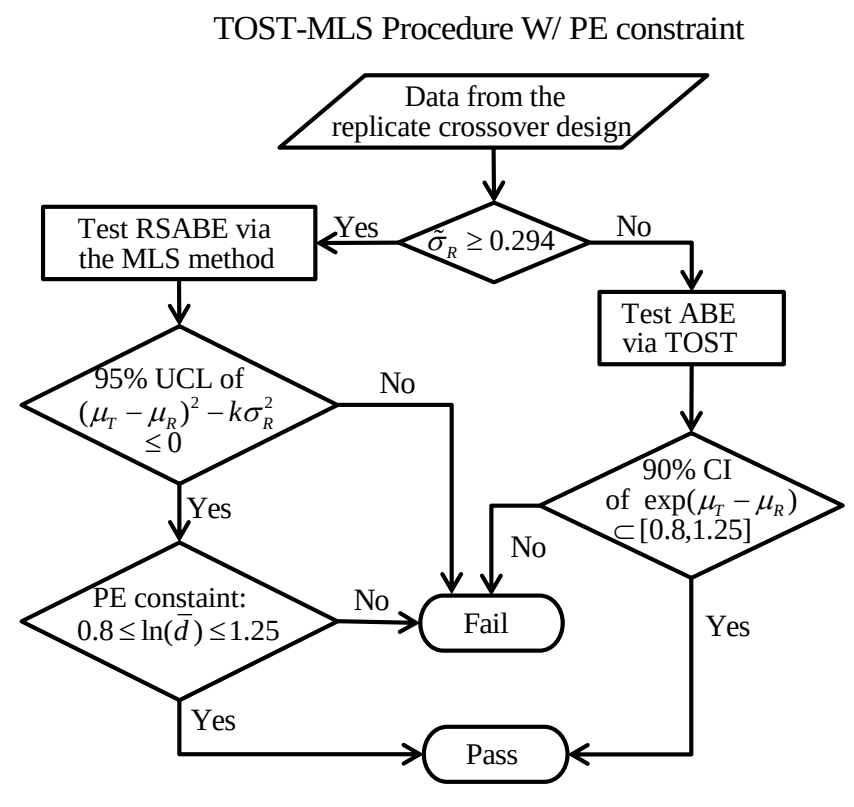
$b = -0.1416$ $\quad s_{+|-} = 19.3255$

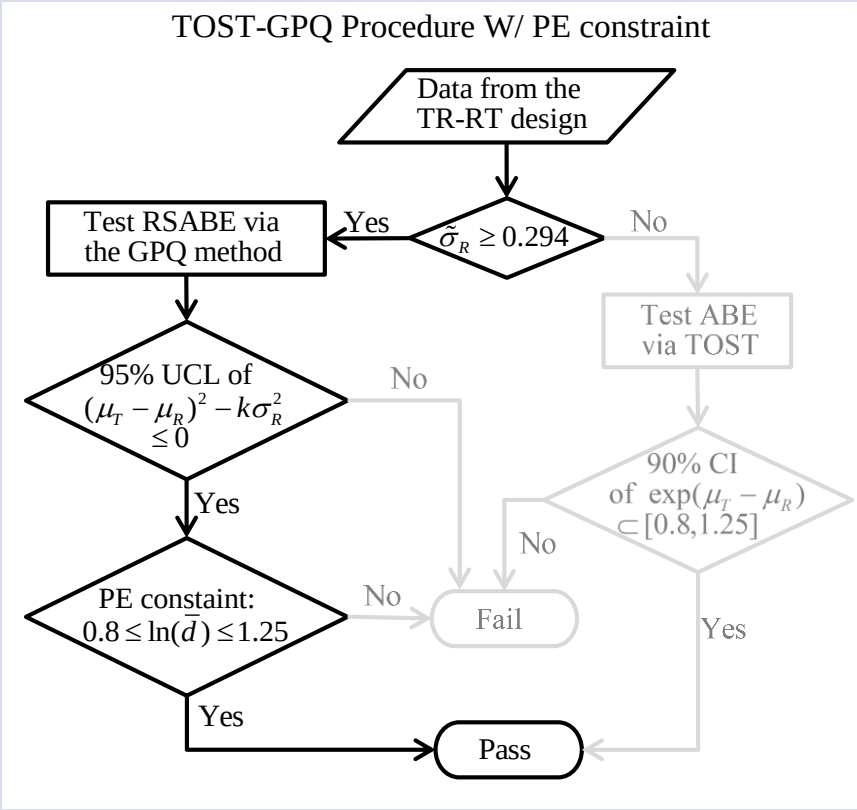
Parameter	Point Estimate (**PE**)	Conf. Level	Conf. Interval (**CI**)
σ_R	0.5375	--	--
$\exp(\mu_T - \mu_R)$	1.1882	90%	(0.9148, 1.5434)
$(\mu_T - \mu_R)^2 - k\sigma_R^2$	-0.2004	95%	(-∞, -0.0113)

Analysis of Example Dataset

The FDA's decision rule for testing RSABE under replicate crossover designs applies TOST when $\tilde{\sigma}_R < 0.296$.

To ensure public confidence and harmonize with various regulatory agencies, the FDA requires PE of the estimated T:R geometric mean ratio to lie within 0.8~1.25.



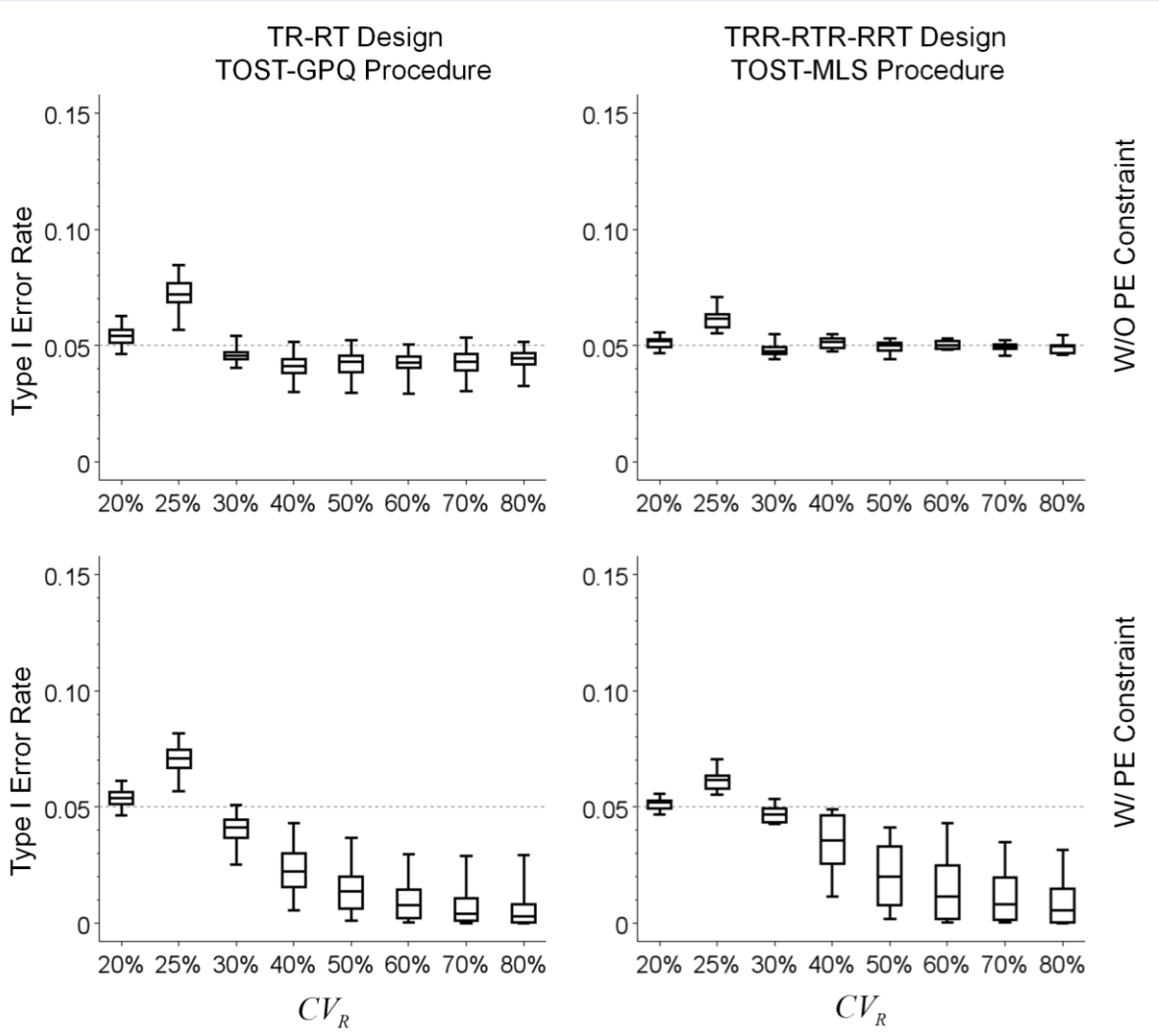


Simulation Study

Simulation parameters were chosen based on historical publications on BA studies.

- Total Sample size
  TR-RT designs: N =24, …, 72 with $n_1=n_2=N/2$;
  TRR-RTR-RRT designs: N=18,…,48 with $n_1=n_2=n_3=N/3$;
- Variation of the reference product
  High: CV_R= 40%, …,80%;
  Borderline high: CV_R= 25% (the changeover point) and 30%;
  Regular: CV_R= 20% .
- Within-subject variance
  Heterogeneous: σ_T/σ_R=0.5 and 2;
  Homogeneous: σ_T/σ_R=1.
- σ_S =0.2, … ,1.
- Zero sequence and period effects.
- 10,000 simulations at each sample size and parameter setting.
- 5, 000 resamples within each simulated dataset.

Simulation Study



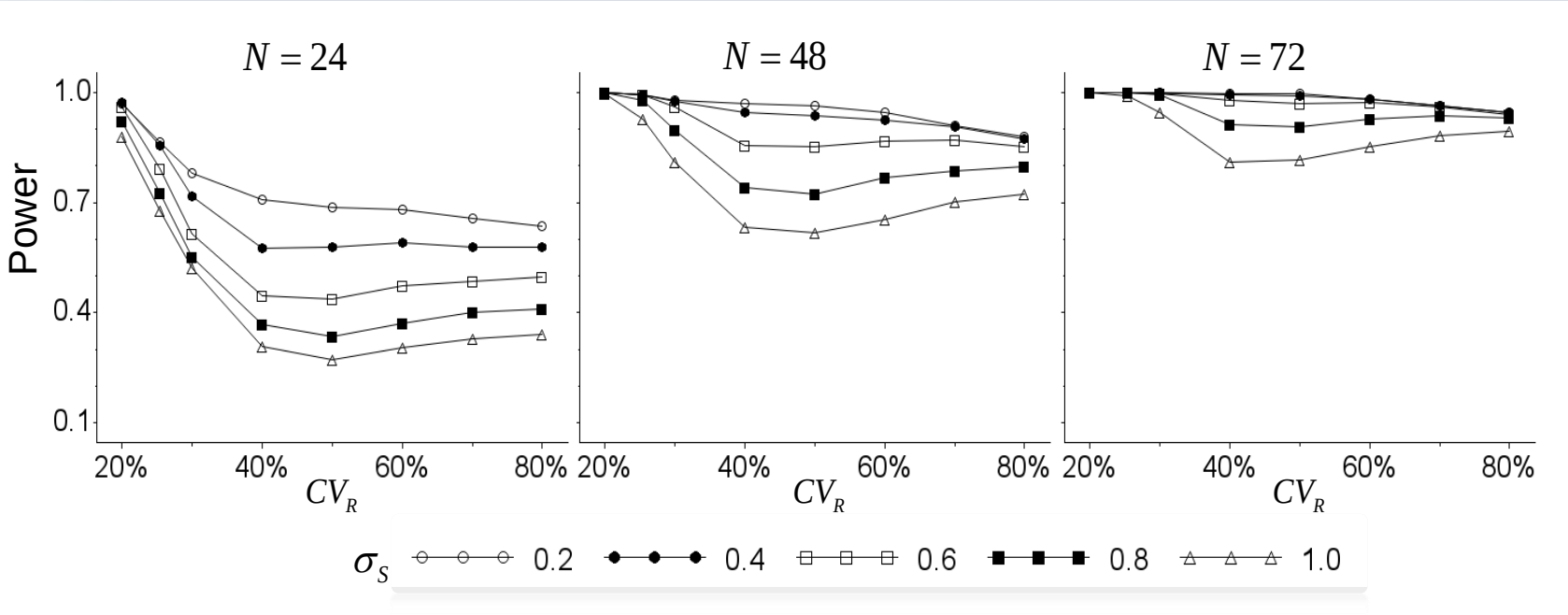
Type I error rate at $(\mu_T-\mu_R)^2=k\max(0.25^2,\sigma_R^2)$. Each schematic boxplot summarizes results of 75 combinations of simulation setting.

Simulation Study

Note:

- The TOST-GPQ procedure was moderately liberal at CV_R =25% but held a desirable type I error rate at all other examined values of CV_R.
- The TOST-MLS procedure preserved its nominal level when $CV_R \neq 25\%$. In comparison to the TOST-GPQ procedure, it had a slightly lower inflation of type I error at CV_R =25% .
- The PE constraint led to ultra conservativeness when $CV_R \geq 40\%$ but had little impact when $CV_R \leq 25\%$.

Simulation Study

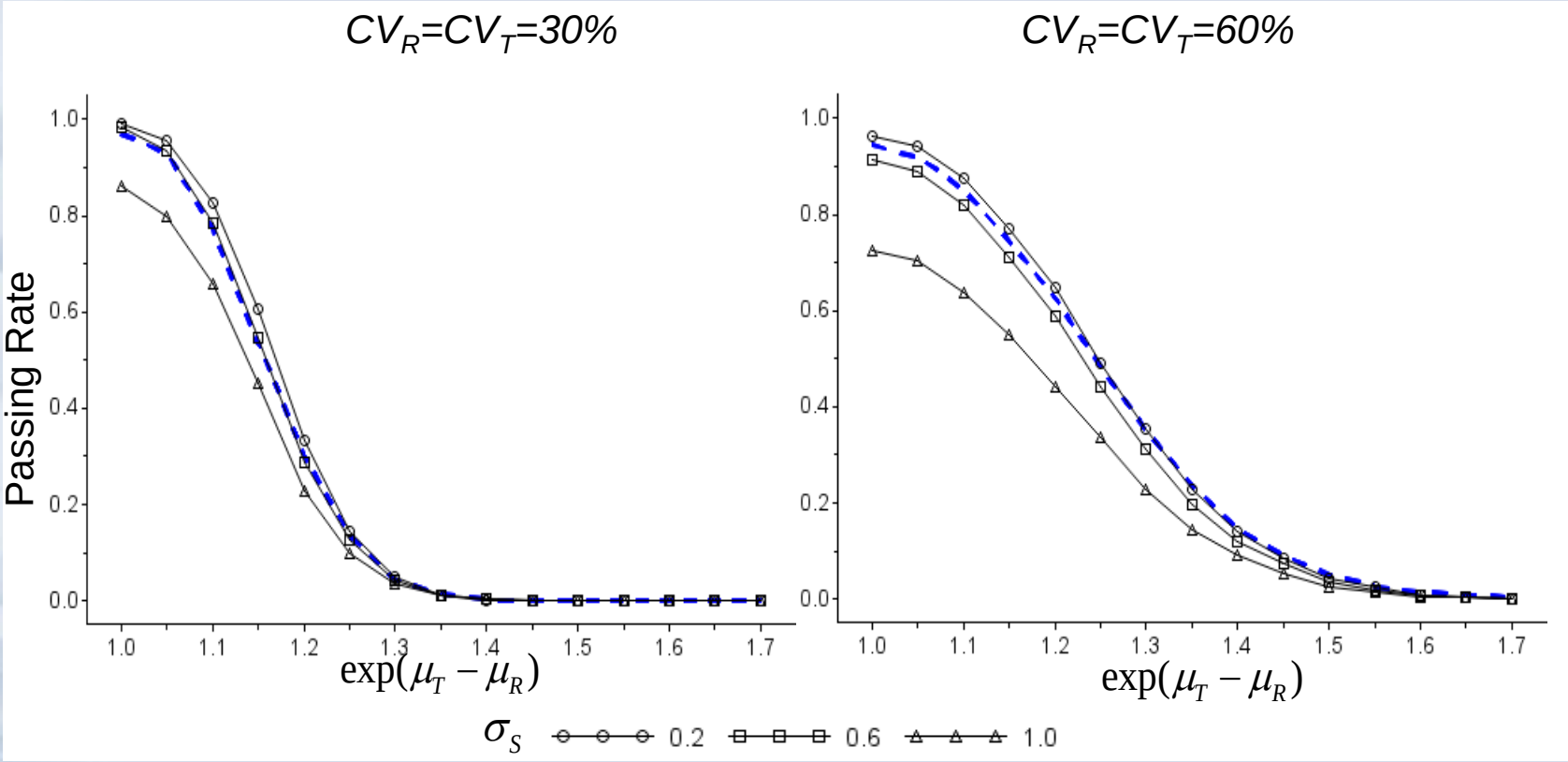


Power for the TOST-GPQ procedure with PE constraint at $\mu_T - \mu_R = 0$ & $\sigma_T/\sigma_R = 1$.

Note:

- Power increased as σ_S decreased and as *N* increased.
- There was no monotonic relationship between power and CV_R. For a given σ_S & *N*, power was always the highest when CV_R is around 20% and dipped to the bottom when CV_R is around 50%.

Simulation Study



Passing rate of the TOST-GPQ procedure under the TR-RT design of size 54 (—) and the TOST-MLS procedure under the TRR-RTR-RRT design of size 36 (- - -).

Note:

- These hybrid procedures performed similarly well at σ_S=0.2 & 0.6.
- The TOST-GPQ procedure underperforms at σ_S=1.

Sample Size Estimation

Total sample size needed to establish BE with mixed scaling at $\mu_T - \mu_R = 0$.

$CV_T = CV_R$	80% power			
	TRR-TRT-RRT Design	TR-RT Design		
		σ_S=0.4	σ_S=0.6	σ_S=0.8
30%	21†	28 (1.3‡)	32 (1.5)	38 (1.8)
40%	22	34 (1.5)	44 (2)	56 (2.5)
50%	22	34 (1.5)	44 (2)	56 (2.5)
60%	23	34 (1.5)	42 (1.8)	52 (2.3)
70%	25	36 (1.4)	42 (1.7)	50 (2)

† Provided by Tothfalusi & Endrenyi 2012. ‡ The ratio of sample sizes.

Note:

- The two designs have a trivial difference at σ_S=0.4 .
- The TR-RT design is more expensive at σ_S=0.6 & 0.8.

Discussion

So far, BA studies on HV drugs have all been carried out under replicate crossover designs with 3 or 4 periods. In cases that these designs are infeasible, parallel designs have been considered as a backup choice. The setbacks include

- prohibitive sample sizes;
- difficulty in translating RSABE when the between- and within-subject variability are confounded.

The proposed approach

- utilizes the fact that subject×formulation is negligible;
- operates under the traditional TR-RT design;
- allows continued exercise of the existing BE criterion.

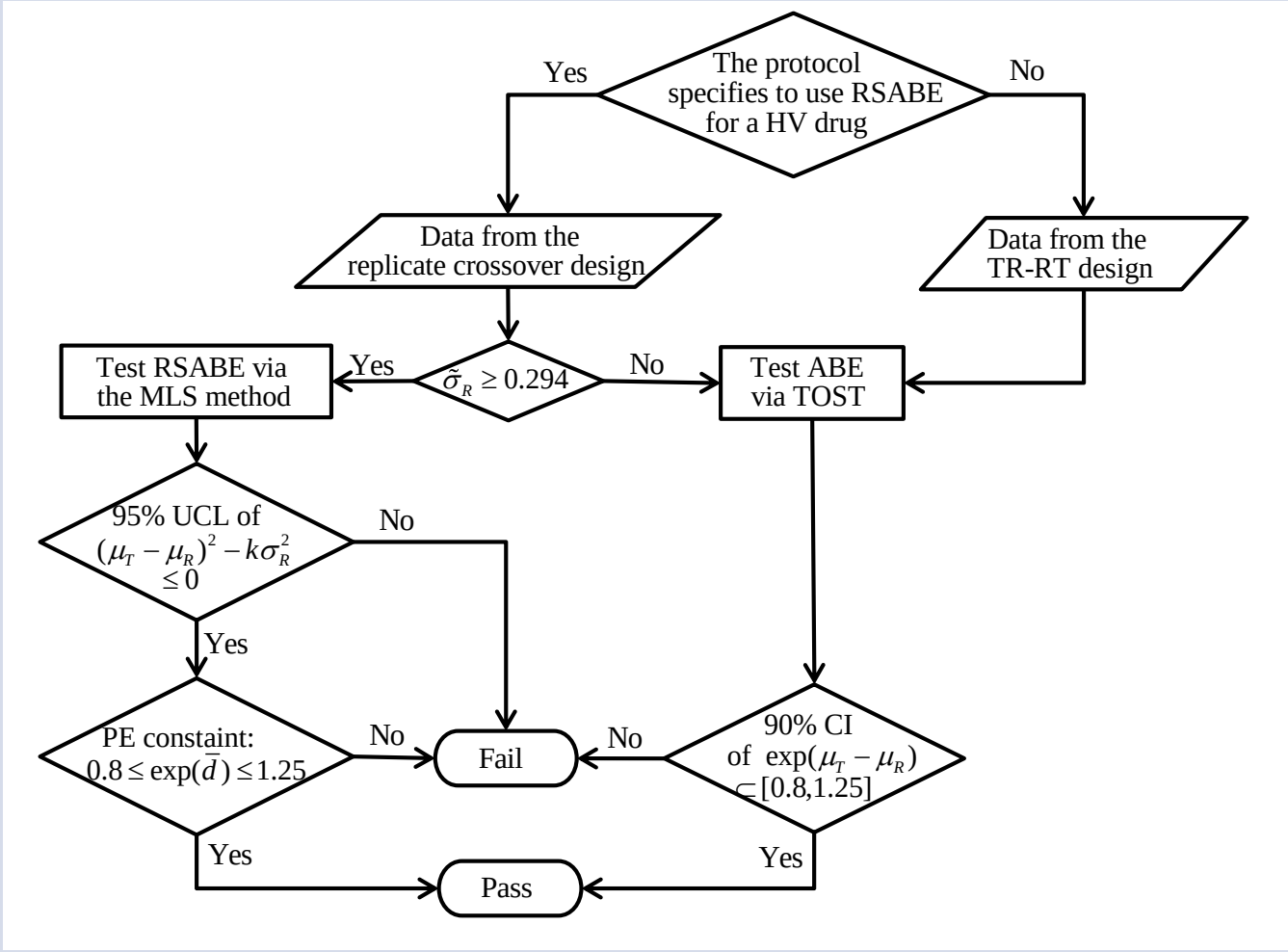
When $\sigma_S \leq 0.4$, the resulting inference is at least as good as that derived from replicate crossover designs. Larger σ_S weakens the inference.

Recommendation: Replicate crossover designs are preferred for BA studies on HV drugs. Under practical restrictions, producers are encouraged to implement the TR-RT design and the proposed TOST-GPQ procedure.

Discussion

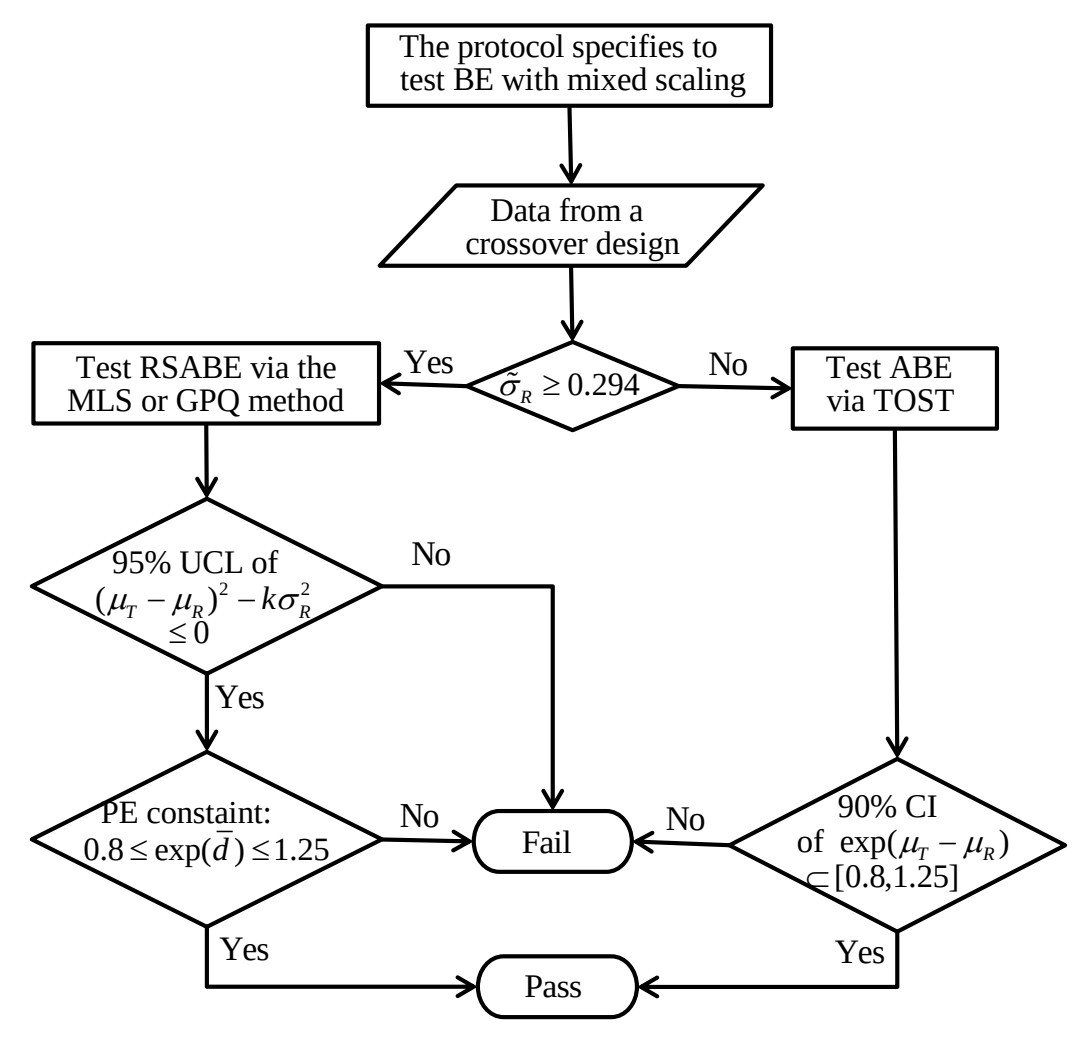
The FDA requires producers to determine, *a priori*, whether the reference drug is HV. Variability classification is a known problem for human and animal drugs.

Dispute arises when producers classify the same reference drug differently.



Discussion

The proposed approach promotes coherence in regulatory evaluation.



References

Claxton R et al. Estimating product bioequivalence for highly variable veterinary drugs. *J. Vet. Pharmaco. Ther.* 2012; 35 (Suppl. 1):11–16.

Davit BM *et al.* Implementation of a reference-scaled average bioequivalence approach for highly variable generic drug products by the US Food and Drug Administration. *AAPS J.* 2012; **14**:915–924.

Draft guidance on progesterone, 2011. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM209294.pdf.

Guilbaud O. Exact inference about the within-subject variability in 2×2 crossover trials. *JASA* 1993; **88**:939-946.

Hannig J *et al*. Fiducial generalized confidence intervals. *JASA* 2006; **101**:254-269.

Haidar *et al.* Evaluation of a scaling approach for the bioequivalence of highly Variable drugs. *AAPS J.* 2008; **10**:450–454.

Rao CR. *Linear statistical inference and its applications* (2nd ed.). Wiley: New York; 1973.

Tothfalusi L & Endrenyi L. Sample sizes for designing bioequivalence studies for highly variable drugs. *J. Pharm. Pharm. Sci.* 2012; **15**:73 – 84.

Weerahandi S. Generalized confidence Intervals. *JASA* 1993; **88**: 899-905.