ALGORITHMS FOR EVALUATING REFERENCE SCALED AVERAGE BIOEQUIVALENCE: POWER, BIAS AND CONSUMER RISK

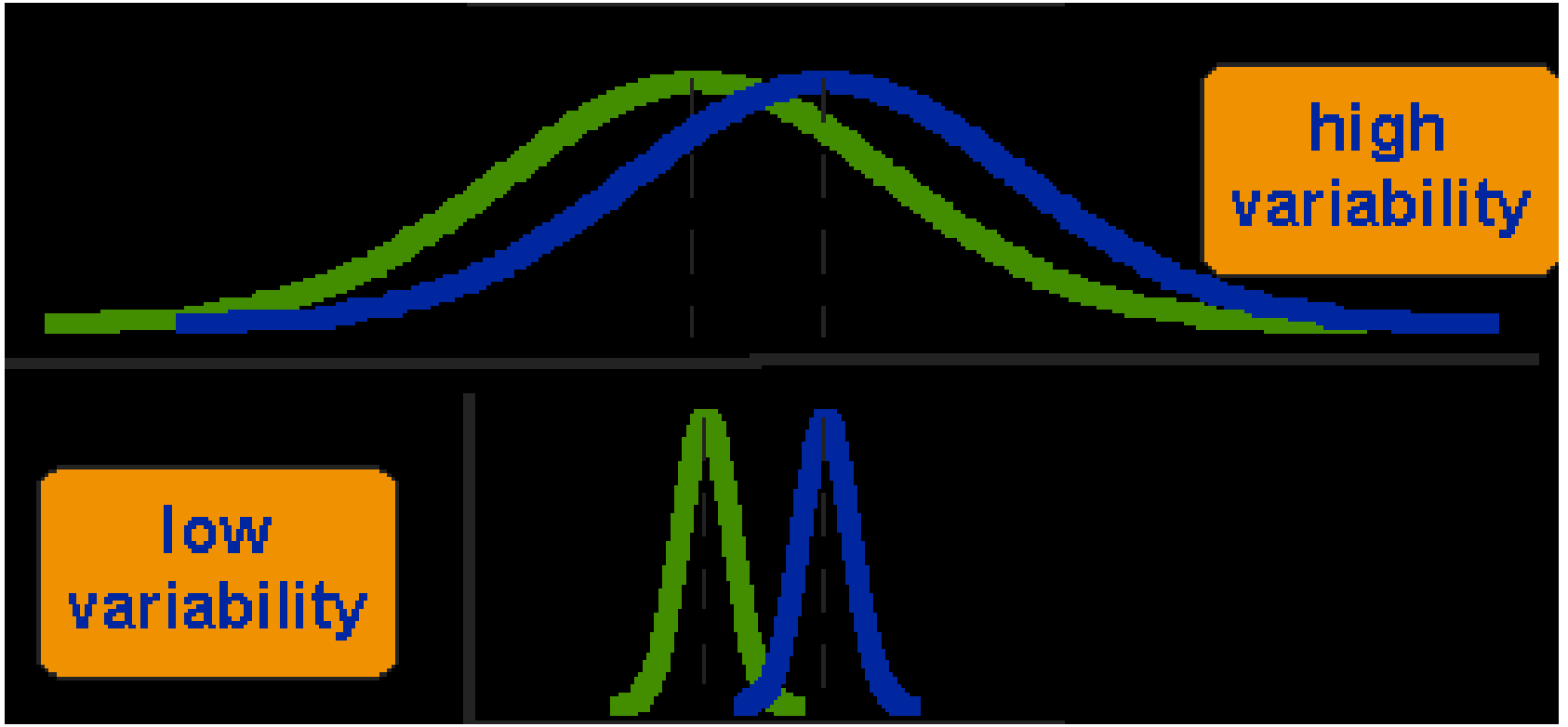
Laszlo Tothfalusi[1] and Laszlo Endrenyi[2]

[1]Semmelweis University, Department of Pharmacodynamics, Budapest, Hungary
[2]University of Toronto, Department of Pharmacology and Toxicology, Toronto, ON, Canada

Notation

- Population values are denoted with Greek letters. The population means of Test (T) and reference (R) formulations are μ_T and μ_R
- The sample estimators of these parameters are random variables and are represented with capital letters such $\bar{Y}_T$, $\bar{Y}_R$, $\bar{Y}_D$, S_{WR}, SE_D. $\bar{Y}_T$, $\bar{Y}_R$ and S_{WR} are the sample estimators of μ_T, μ_R and s_{WR}, respectively. $\bar{Y}_D$ is the $\bar{Y}_T$- $\bar{Y}_R$ difference, and the standard error of this difference is SE_D. The true population value of SE_D is σ_D
- Realization of the random values (i.e. numbers) are $\bar{y}_T$, $\bar{y}_R$, $\bar{y}_D$, s_{WR} and se_D.
- θ – the regulatory constant (0.76 in the EU)

The types of equivalence -graphical interpretation of the concept



Relative difference or overlap

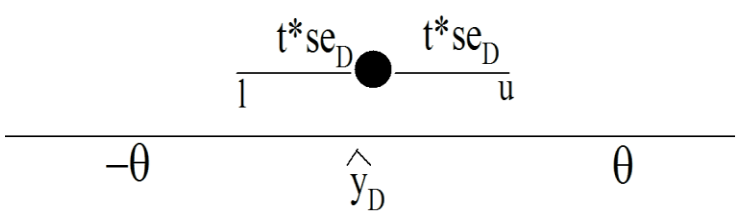
Difference between means

ABE, SABE, RSABE - definition and interpretation

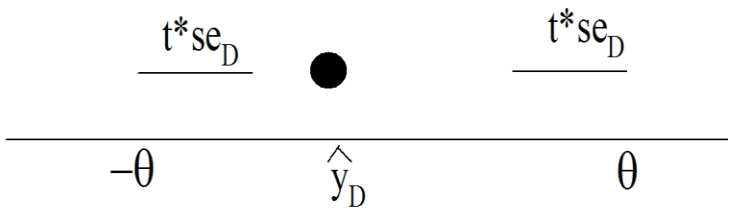
	Equivalence Test Hypothesis	Interpretation, comment
Average bioequivalence (ABE)	$-\theta \leq (\mu_T - \mu_R) \leq \theta$	Difference between means
Scaled Average Bioequivalence (SABE)	$-\theta \leq (\mu_T - \mu_R)/\sigma_W \leq \theta$	Standardized mean (β), overlap, effect size (Cohen's d)
Reference Scaled Average Bioequivalence (RSABE)	$-\theta \leq (\mu_T - \mu_R)/\sigma_{WR} \leq \theta$	Overlap relative to the reference, effect size (Glass' d)

Flavors of the ABE algorithm – The confidence interval based and the TOST variant

Confidence interval



TOST



1- Estimate the l and u interval limits
2- If they are in the $[-\theta,\theta]$ range then BE

H01: Test that $-\theta$ is equal or larger than the point estimate
H02: Test that θ is equal or smaller than the point estimate
If both tests reject H0 then BE.

Visually: the two tests also define a range where the point estimate should be.

RSABE algorithms: The ABEL method

- S_{WR} is treated as if it were a constant. Essentially, the ABEL approach traces the RSABE problem back to the well-established method evaluating ABE
- Rearranging the RSABE test criterion gives

1.	$l = \bar{y}_D - t(0.95,df)*se_D$	same as in ABE
2.	$u= \bar{y}_D + t(0.95,df)*se_D$	same as in ABE
3.	$s_{WR}\,\theta \le l < u \le s_{WR}\,\theta$	instead fixed θ we have random limit

Advantages and disadvantages of the ABEL method

- As the name suggests it is not a RSABE test but „Average Bioequivalence with Expanding Limit”
- Pro
  - Easy to calculate
  - Easy to interpret
  - Official in the EU, Australia, Canada.
- But the consumer error is higher than the nominal 5%
  - 7%-9%
  - This was always clear starting from Boddy et al.(1995)

RSABE algorithms: The Hyslop's method

- Hyslop et al., proposed for IBE (2000)
- We simplified it for RSABE (2003), currently recommended by the FDA
- The starting point is the rearranged and squared criterion
  - $(\mu_T - \mu_T)^2 - \theta^2 * \sigma_{WR}^2 \leq 0$
  - Background: Let W and Y two random variables. Their estimates $\hat{w}$ and $\hat{y}$. Let u_W and l_w of W, and u_y and l_y the estimated lower and upper confidence interval limits of W and Y. Howe showed that

$$l = \hat{w} - \hat{y} + \sqrt{(\hat{w} - l_W)^2 + (u_y - \hat{y})^2}$$

$$u = \hat{w} - \hat{y} + \sqrt{(u_W - \hat{w})^2 + (\hat{y} - l_y)^2}$$

Applying Howe's approximation we get the Hyslop's algorithm for RSABE

Hyslop's algorithm for RSABE

- Pro
  - Theoretically correct
  - FDA backed
- But
  - Cl has no meaning (what does it mean Cl =0.03?)
  - Requires programming- not that simple
  - Consistent but biased test (Tothfalusi and Endrenyi, AAPS ,2016)

1. $E_m = (\hat{y}_T - \hat{y}_R)^2$
2. $E_s = \theta^2 * s_{WR}^2$
3. $C_m = [Abs(\hat{y}_T - \hat{y}_R) + t * se_D]^2$
4. $C_s = \theta^2 * df\ s_{WR}^2/\chi^2$
5. $L_m = (C_m - E_m)^2$
6. $L_s = (C_s - E_s)^2$
7. $Cl = E_m - E_s + (L_m + L_s)^{½}$
8. Bioequivalent if $Cl < 0$.

Exact algorithms for RSABE

- Hyslop's algorithm is based on approximation
- Can we do without it ?

RSABE definition in terms of population parameters

Idea: $-\theta \le (\mu_T - \mu_R)/\sigma_{WR} \le \theta$

Working definition →

RSABE definition in terms of estimators (random variables)

$\text{Prob}(-\theta \le \bar{Y}_D / S_{WR} \le \theta) > 0.9$

Ratio of a normally ditributed and a χ^2 distributed random $vv - s$
Similar to noncentral t…

The distribution of $\bar{Y}_D / S_{WR}$

Assume that there is k constant that $\sigma_D = k^* \sigma_{WR}$

Substitute the definion of k

$$\bar{Y}_D \sim N(\mu_T - \mu_R, k * \sigma_{WR})$$

$$S_{WR} \sim \sigma_{WR} \sqrt{\chi^2/df}$$

$$\frac{\bar{Y}_D}{S_{WR}} \sim \frac{N(\mu_T - \mu_R, k)}{\sqrt{\chi^2/df}}$$

After divison simplify the right side with σ_{WR}

$$k^{-1} \frac{\bar{Y}_D}{S_{WR}} \sim \frac{N(\lambda,1)}{\sqrt{\chi^2/df}}$$

Multiply with k^{-1} both sides and define λ as = $k^{-1} (\mu_T - \mu_R)/\sigma_{WR}$.

Noncentral t by definition

So we showed that $k^{-1}\bar{Y}_D / S_{WR}$ is distributed as noncentral t with $\lambda = k^{-1}(\mu_T - \mu_R)/\sigma_{WR}$. But λ is unknown.

Idea: Construct an equivalence test for the $\lambda = k^{-1}(\mu_T - \mu_R)/\sigma_{WR}$ noncentrality parameter. If we know that λ is in $[-k^{-1}\theta, k^{-1}\theta]$ range with 90% confidence then also $k\lambda = (\mu_T - \mu_R)/\sigma_{WR}$ will be in the $[\theta, \theta]$ range.

$$\text{Prob}(-\theta \leq \bar{Y}_D / S_{WR} \leq \theta) \geq 0.9$$

They are equivalent

$$\text{Prob}(-k^{-1}\theta \leq k^{-1}\bar{Y}_D / S_{WR} \leq k^{-1}\theta) \geq 0.9$$

Seems that we have a test but actually we do not…. Hedges' correction.

The point is that ABEL and Hyslop tests are equivalence tests for the population mean (δ) of the $\bar{Y}_D / S_{WR}$ statistics. Hedges showed that the λ noncentrality parameter is not equal with the δ. The following relationship holds:

$\delta = H_f^{-1}\lambda$ where $H_f = 1 - 3/(4df-1)$ where df is the degree of freedom of S_{WR}.

For example if df = 6 then $\delta = 1.15\lambda$.

Idea: Multiply θ with H_f . A test for λ is also a test for δ.

$Conf(-H_f k^{-1}\theta \leq \lambda \leq H_f k^{-1}\theta) \geq 0.9$ —Multiply with H_f^{-1}→ $Conf(-k^{-1}\theta \leq \delta \leq k^{-1}\theta) \geq 0.9$

Only and only If this is true

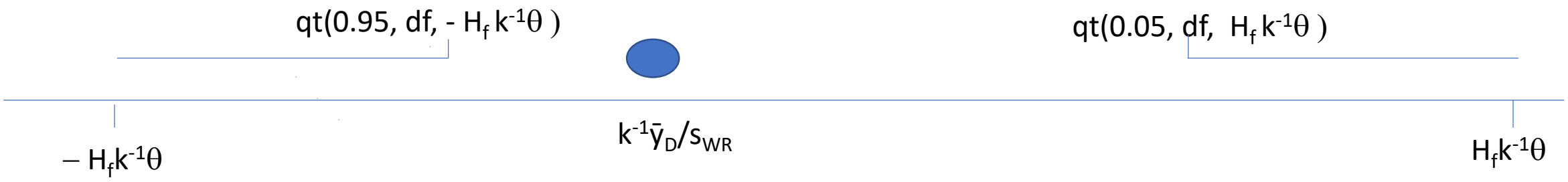
this will be also true

Conclusion: The RSABE test was derived back to an equivalence test for a noncentrality parameter (λ) of a noncentraly distributed t variable. If the original range was $[-\theta, \theta]$ then we should test that λ is in $[-H_f k^{-1}\theta, H_f k^{-1}\theta]$

What about „k”?

- The constant „k” is the ratio of the standard deviation of $\bar{Y}_D$ and σ_{WR}.
- If we assume that σ_{WR}= σ_{WT} then it depends only on the design. For example for the TRTR-RTRT design it is $n^{-0.5}$.
- Using contrast we can give formula even for the heteroscedastic case. But of course we can be never sure that σ_{WT}/ σ_{WR}= 1.
- We initially proposed (AAPS,2016) first estimate the σ_{WT}/ σ_{WR} ratio and using that to calculate k. This step was difficult and was practically impossible for partial replicate design (TRR-RTR-RRT).
- But we „rediscovered” paper of Schall (1995) who gave a very simple solution: k= se_D /s_{WR}--→ Just use the estimates of the output !

The RSABE problem: Given the point estimate $\bar{y}_D/s_{WR}$ make a statement that $E(\bar{Y}_D/S_{WR})$ is in the $[-\theta, \theta]$ range. The TOST solution. (The NcTOST algorithm)



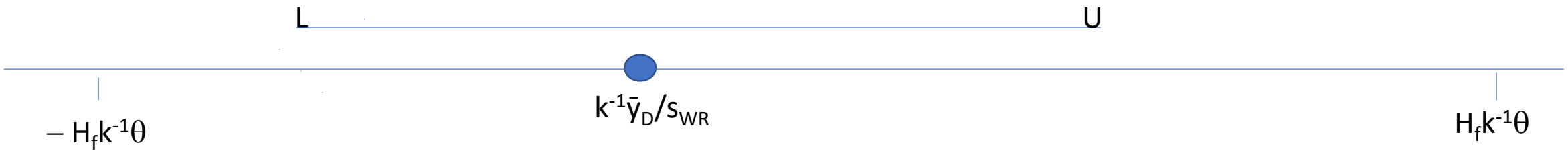
Step 1. Compute $k = se_D/s_{WR}$

Step 2. Compute L_θ = qt(0.95, df, - $H_f k^{-1}\theta$) and U_θ = qt(0.05, df, $H_f k^{-1}\theta$)

Step 3. If $L_q < k^{-1}\bar{y}_D/s_{WR} < U_q$ accept the RSABE hypothesis else reject

qt = Quantile function of a noncentral t variables (see R base documentation)

The RSABE problem: Given the point estimate $\bar{y}_D/s_{WR}$ make a statement that $E(\bar{Y}_D/S_{WR})$ is in the $[-\theta, \theta]$ range. The classical confidence interval solution. (The NcConf algorithm)



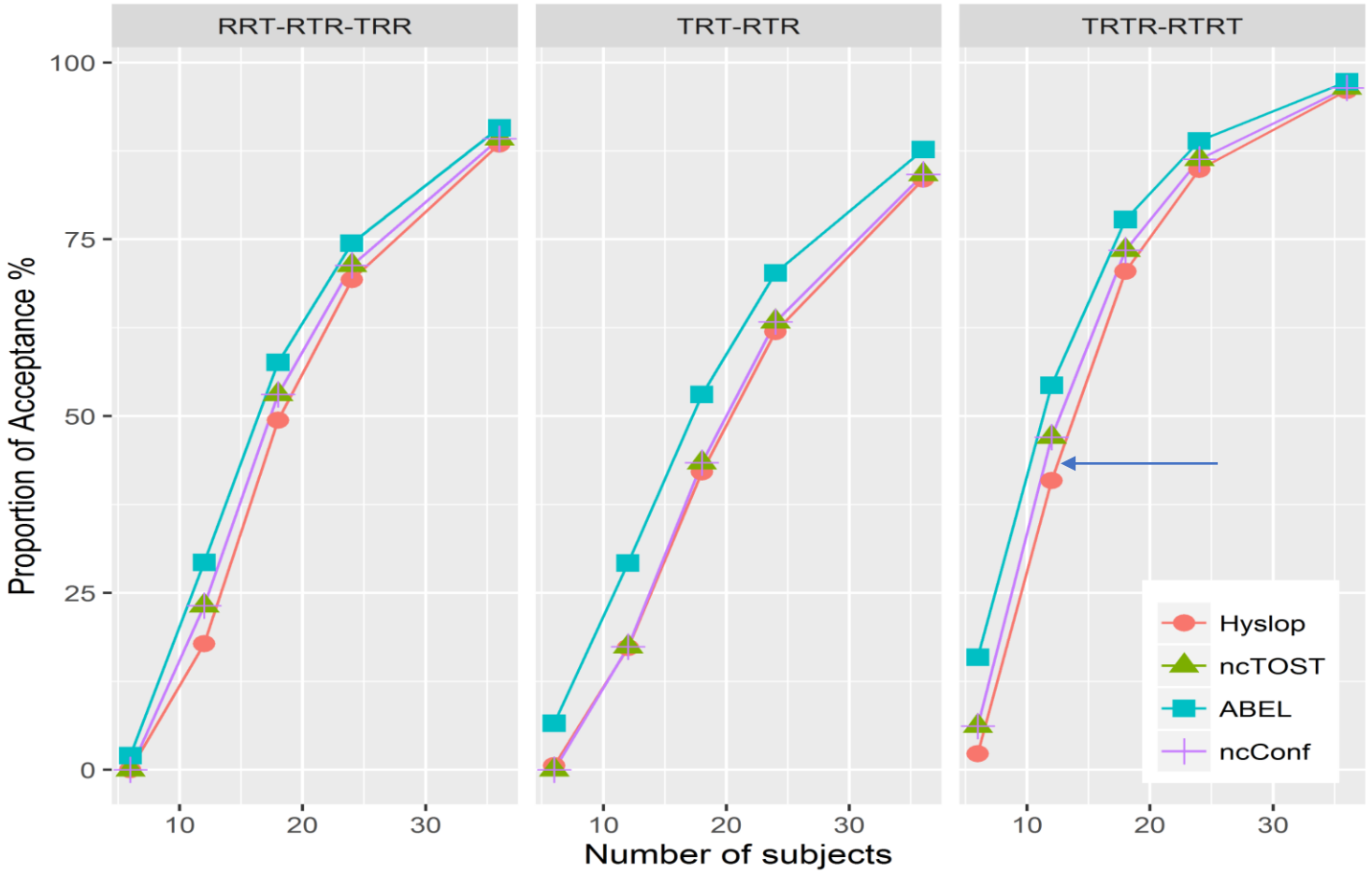
Step 1. Compute $k = se_D/s_{WR}$

Step 2. Estimate the lower (L) and upper (U) confidence limits for $k^{-1}(\mu_T - \mu_R)/\sigma_{WR}$ The calculation requires solving two nonlinear equations, solving $qt(0.95, df, L) = k^{-1}\bar{y}_D/s_{WR}$ and $qt(0.05, df, U) = k^{-1}\bar{y}_D/s_{WR}$ for L and U.

Step 3. RSABE is established if $L \geq -H_f k^{-1}\theta$ és $U \leq H_f k^{-1}\theta$.

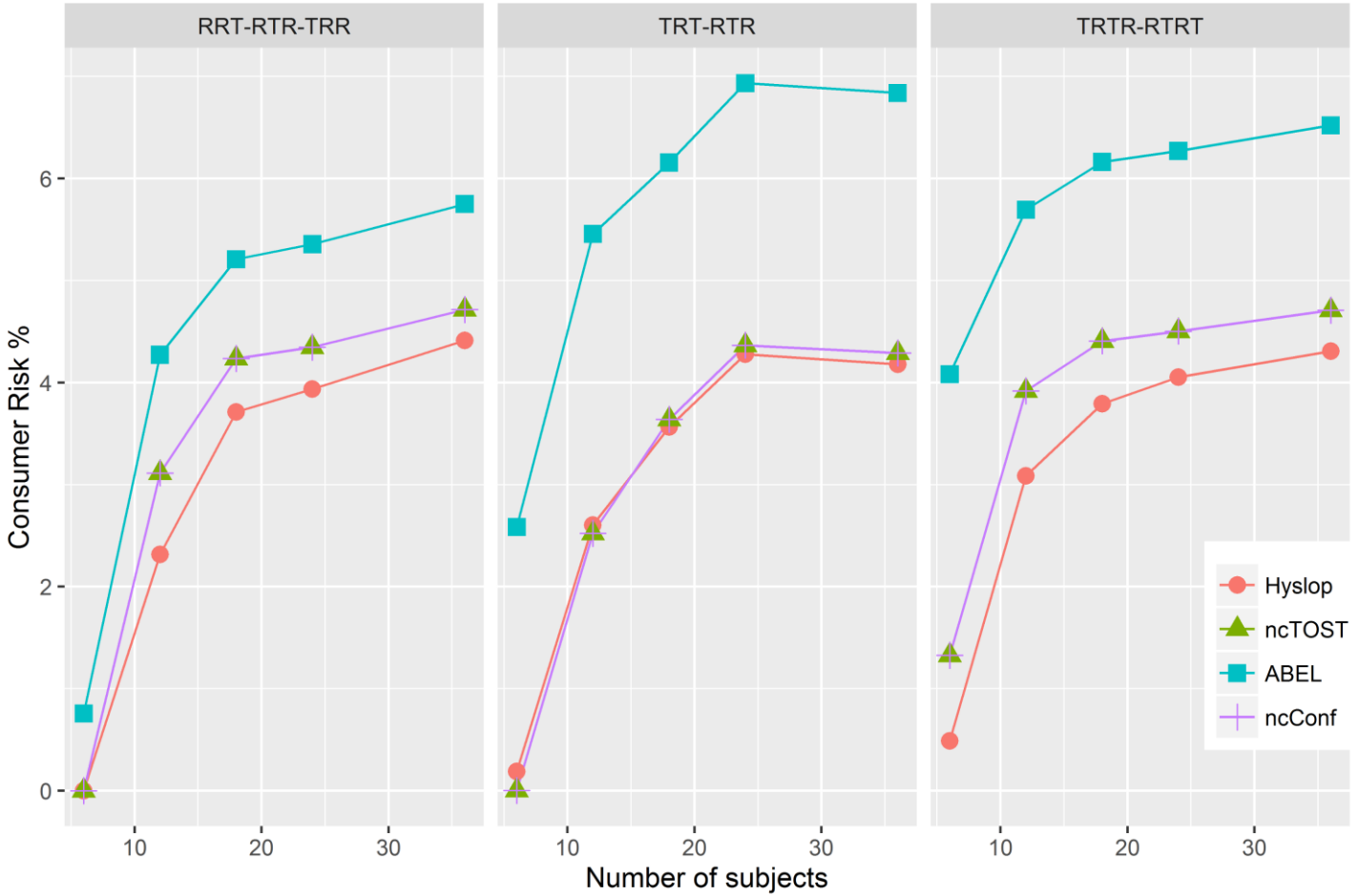
Step 4. (Optional) The confidence limits for $(\mu_T - \mu_R)/\sigma_{WR}$ are kL and kU.

Simulations: Power



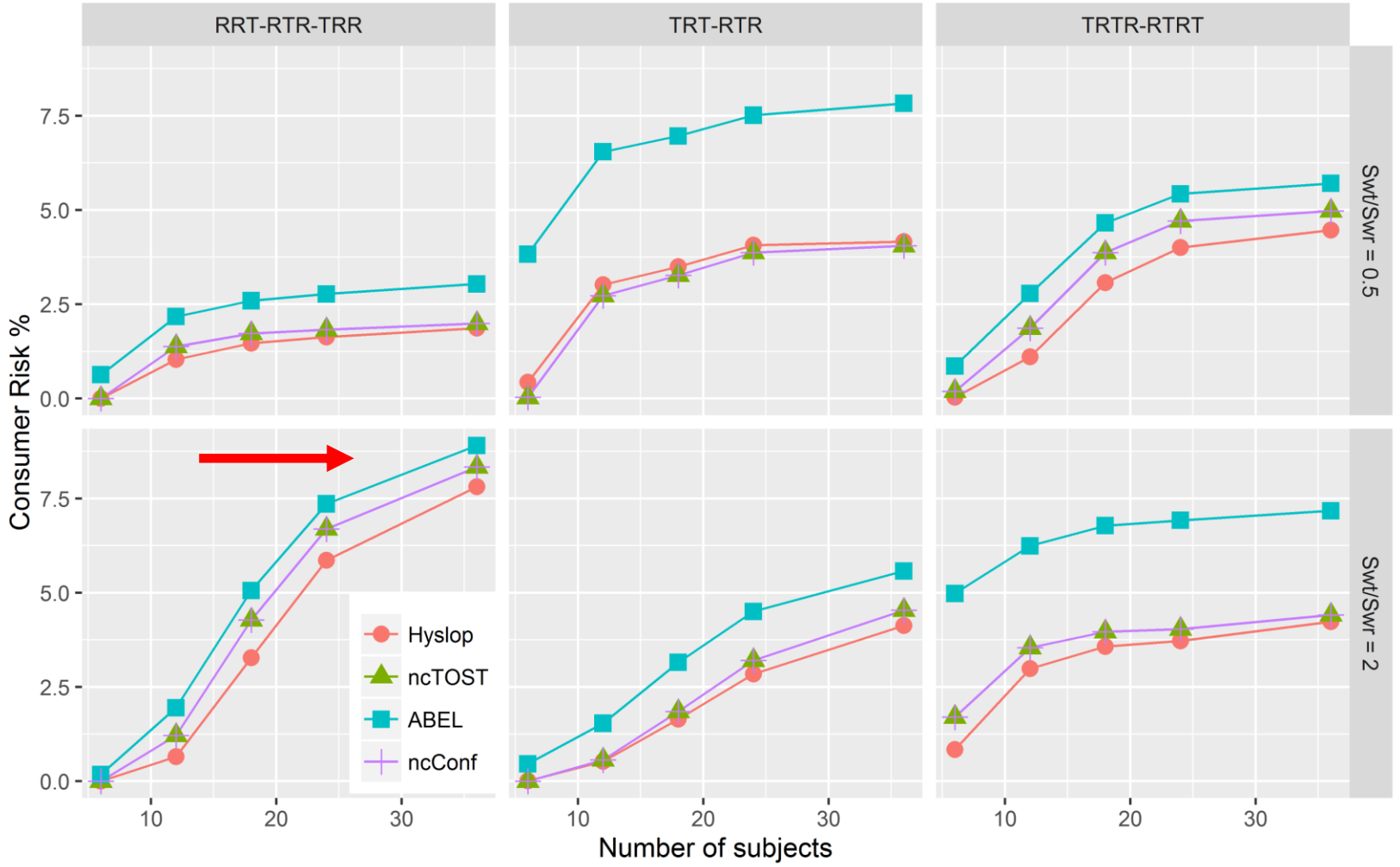
The approximate Hyslop method has noticeably lower power. For example, the acceptances were, with twelve subjects and the TRTR-RTRT design, 54.6, 47.1, 47.1 and 40.9 % with the ABEL, ncTOST, ncConf and Hyslop's algorithms

Simulations: Consumer Risk $\sigma_{WR} = \sigma_{WT}$



The consumer risk remains below 5 % with both variants of the Exact algorithms.

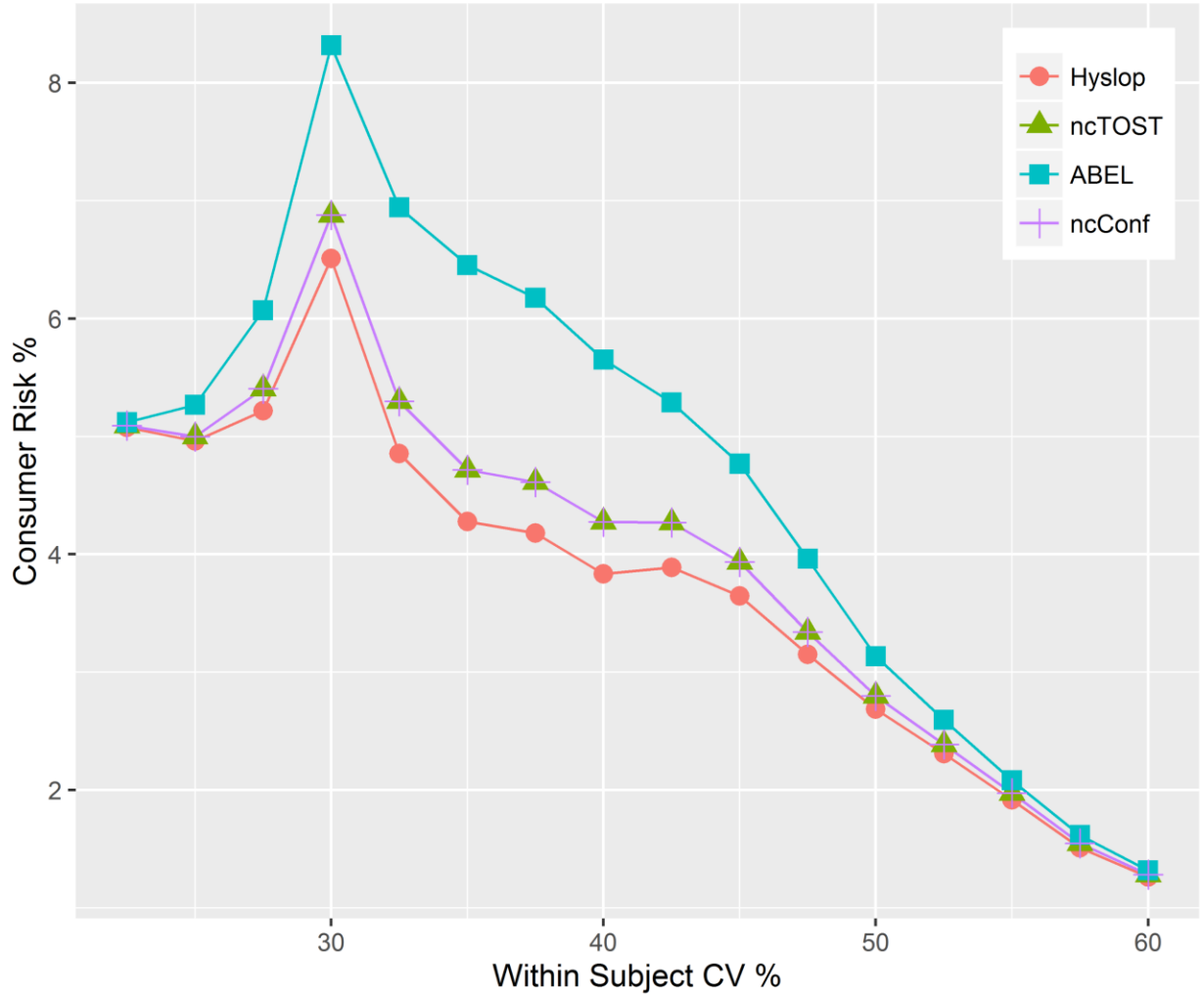
Simulated trials where σ_{WT}/σ_{WR} was either 0.5 (upper row) or 2 (lower row).



Observe that with partial replicate design and σ_{WT}/σ_{WR} = 2 condition („bad generic”) , the consumer risk increased steadily and significantly above the 5% limit with **each** algorithm.

Mixed strategy: The „Blip”

The consumer risk is above 5% with all algorithms but returns, around CV_{WR}=35%, below the nominal level with the Hyslop and Exact algorithms. The same pattern can be observed with ABEL except that it remains above the 5% level



Conclusions

- The NcTOST algorithm is a one-liner, can be computed without programing. Compared to the Hyslop's approximation, at small sample size it is more powerful (not biased).
- The NcConf is computationally more complex but it provides exact confidence intervals.
- The consumer risk with the Hyslop's and Exact algorithms are below 5% except in two cases: partial replicate design and around CV_{WR}= 30%.
- Paradoxically the partial replicate design is the recommended design by the EMA.
- The difference between $E(\bar{Y}_D / S_{WR})$ and $(\mu_T-\mu_R)/\sigma_{WR}$ raises new questions.
- Biosimilarity implications:
  - BE studies (including the PD part)
  - Comparative assessment of quality attributes?

History

- Schall provided the first version of the Exact algorithm. (1995)
- The version presented is based on Tothfalusi and Endrenyi (2017). It is an improved version compared to Tothfalusi and Endrenyi (2016).

References

Boddy AW, Snikeris FC, Kringle RO, Wei GC, Oppermann JA, Midha KK. An approach for widening the bioequivalence acceptance limits in the case of highly variable drugs. Pharm Res. 1995 Dec;12(12):1865–8.

Schall R. Assessment of individual and population bioequivalence using the probability that bioavailabilities are similar. *Biometrics* 1995; **51(2)**: 615-626.

Hyslop T, Hsuan F, Holder DJ. A small sample confidence interval approach to assess individual bioequivalence. *Statistics in Medicine* 2000; **19(20)**: 2885-2897

Tothfalusi L, Endrenyi L. Limits for the scaled average bioequivalence of highly variable drugs and drug products. Pharmaceutical Research. 2003;20(3):382-389.

Tothfalusi L, Endrenyi L. An exact procedure for the evaluation of reference-scaled average bioequivalence. *AAPS Journal* 2016; **18(2)**: 476-489.

Tothfalusi L, Endrenyi L.
Algorithms for evaluating reference scaled average bioequivalence: Power, bias, and consumer risk. Statistics in Medicine 2017 Aug 29. doi: 10.1002/sim.7440. [Epub ahead of print]

European Medicines Agency. Questions & Answers: positions on specific questions addressed to the Pharmacokinetics Working Party (PKWP). 19 November 2015 EMA/618604/2008 Rev. 13, Question 19

Questions ?