Getting the Size Right

Finite-Sample Corrections for Univariate & Multivariate Bioequivalence Testing

Younes Boulaguiem, PhD

Roche · Biostatistics
Translational and Exploratory Biomarkers Group

BBS Suisse Romande
Webinar series 2026
13 November 2026

View Publications

Disclaimer

Views expressed are **my own** and do not necessarily reflect those of **my employer**.

Product names are trademarks of their respective owners.

This Talk in a Nutshell

What is Bioequivalence?

- Bioequivalence is used when we want to assess whether two drugs are **sufficiently similar** (e.g., in their pharmacokinetic behaviour) so that they can be used **interchangeably**.
- We are not looking for *"exactly the same"*, but *"so close that any difference is clinically irrelevant"*.

The added value of our work:

- The standard methods to assess bioequivalence are **too conservative**, meaning that **we fail to detect** bioequivalent drugs more often than necessary.
- Our work helps **correct this conservativeness** while **increasing power** over the whole parameter space.

2024-2029: Increased Reliance on Equivalence Frameworks

Multiple high-revenue drugs are losing exclusivity.

- Increased biosimilar activity.
- Increase in lifecycle management strategies (e.g. subcutaneous Keytruda, Opdivo, Ocrevus).

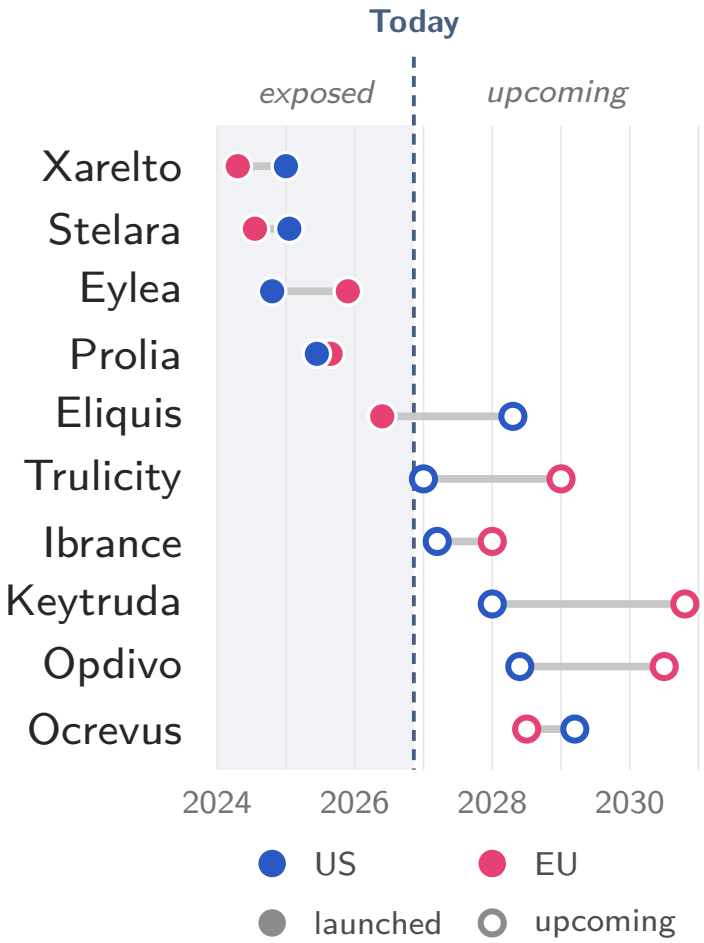
⇒ **Greater use of non-inferiority and equivalence designs.**

Statistical inefficiencies can be consequential.

- Larger required sample sizes.
- Higher development costs.
- Delays in competitive positioning.

How do we assess bioequivalence?

The Patent Cliff



First generic/biosimilar entry (incl. settlements).

Recap and Terminology

- H_0 is the hypothesis we are cautious about rejecting; the burden of proof lies with H_1.
- We build a test and choose a rejection region, bounded by *critical values*, so that the Type I error rate equals α.
- Power can be increased without compromising Type I error rate by reducing uncertainty, which incurs additional costs and efforts.
- α is the *nominal significance level*; the actual Type I error rate is called the size of the test.

size $= \alpha$ $\Rightarrow$ **size-α**, i.e., exact test (ideal),

size $\leq \alpha$ $\Rightarrow$ **level-α**, i.e., conservative test (second best),

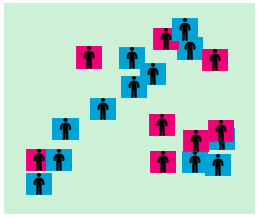
size $> \alpha$ $\Rightarrow$ liberal (we don't like...)

Statistical Hypotheses

$$H_0 : \theta = 0$$

$$H_1 : \theta \gtrless 0$$

Sample



Data

$$\mathbf{y}_T = (y_{T,1}, y_{T,2}, ..., y_{T,m})$$
$$\mathbf{y}_R = (y_{R,1}, y_{R,2}, ..., y_{R,n})$$

Point estimation

$$\widehat{\theta} = g(\mathbf{y}_T, \mathbf{y}_R)$$
$$\widehat{\sigma} = h(\mathbf{y}_T, \mathbf{y}_R)$$

Inference

Deduce information about θ

with statistical tests, confidence intervals

Purpose:

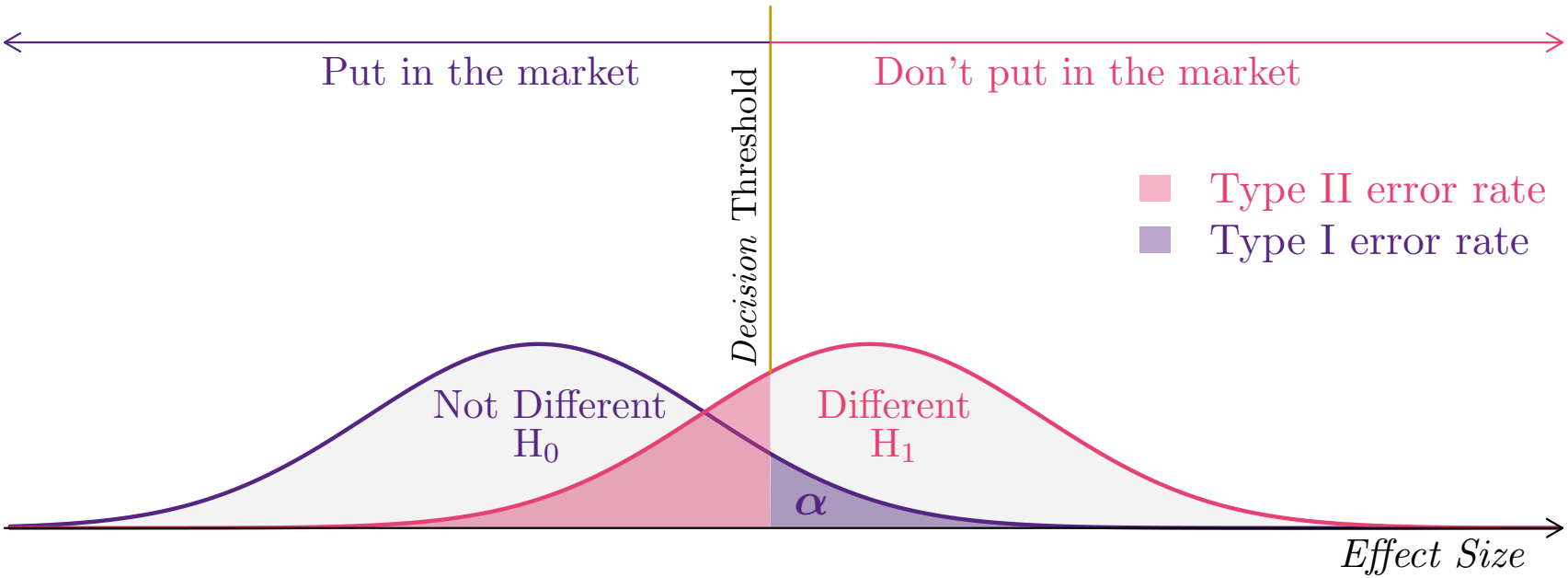
Compare mean drug concentration of reference (R) and test (T)

Limitation of the Classical Approach I

Drawbacks of the Classical Approach:

1. **Not rejecting the null does not mean no difference** $(\theta = 0)$.
   → Many times, just a lack of statistical power!

H_0 : Not Different vs. H_1 : Different

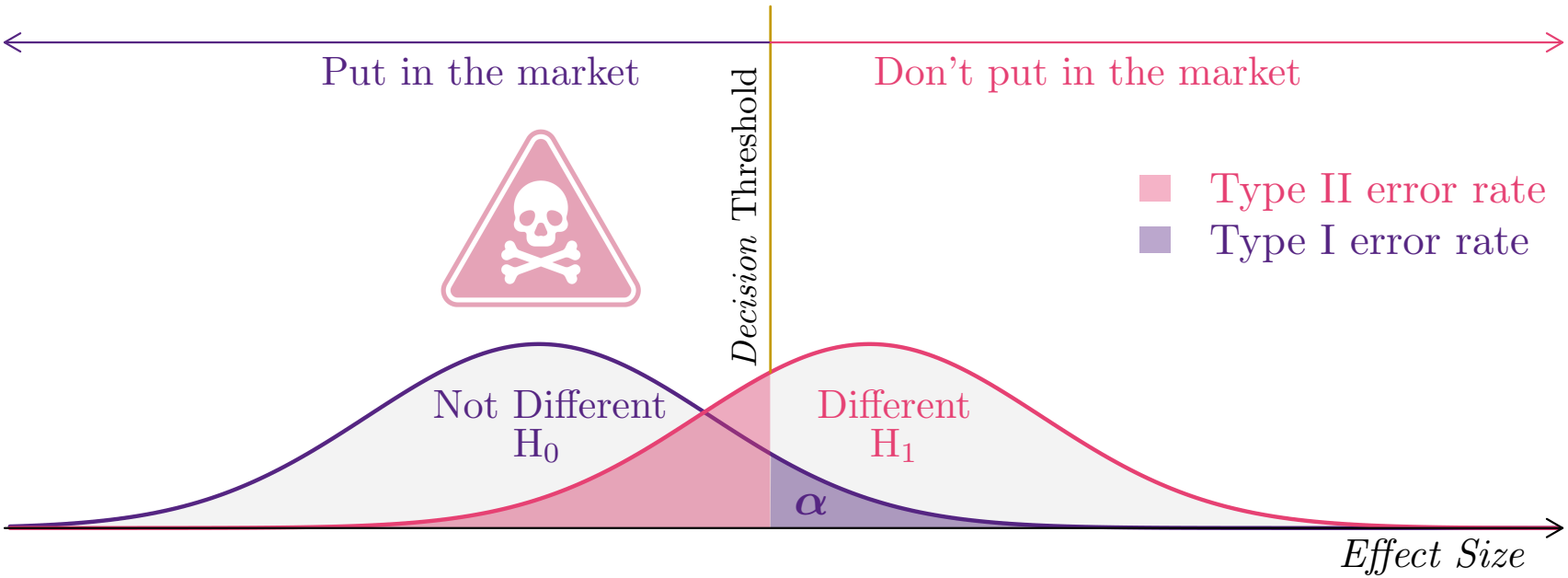


Limitation of the Classical Approach I

Drawbacks of the Classical Approach:

1. **Not rejecting the null does not mean no difference** $(\theta = 0)$.
   → Many times, just a lack of statistical power!

H_0 : Not Different vs. H_1 : Different



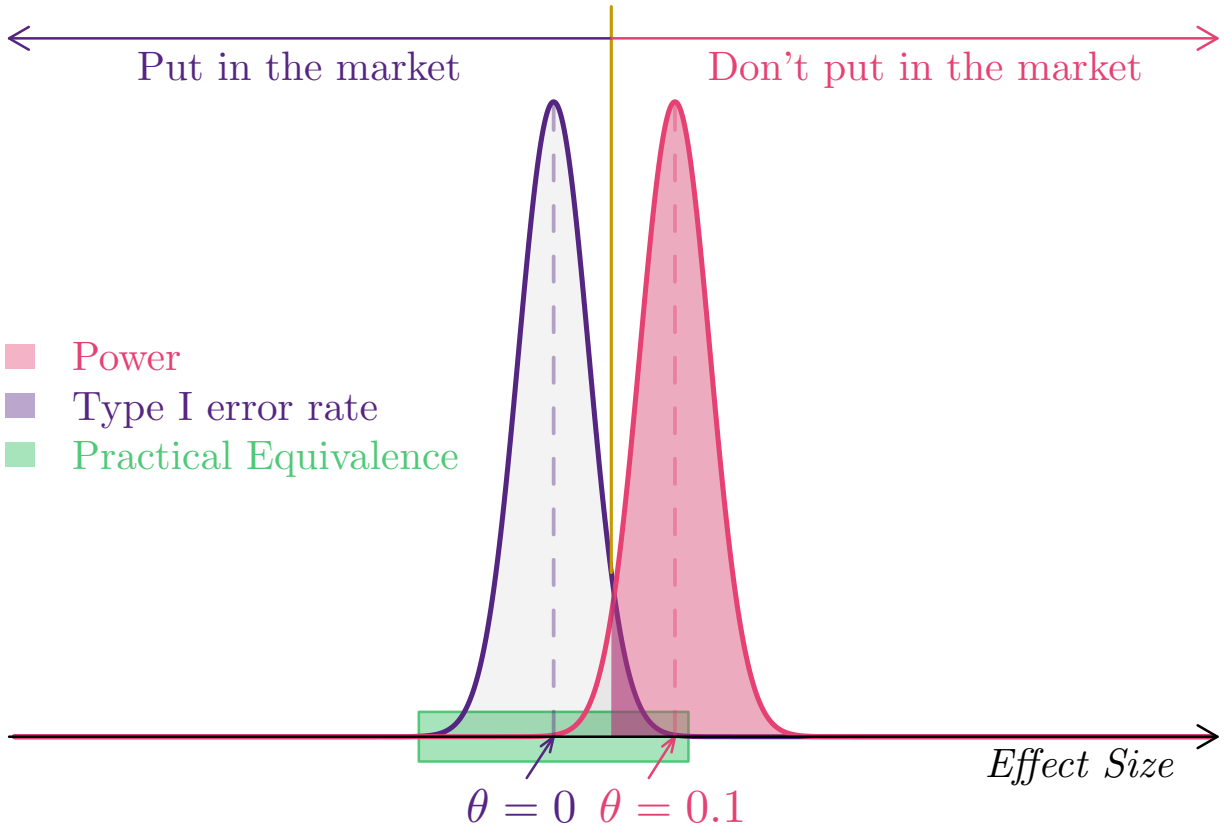
No incentive to pursue power: non-equivalent drugs make it to the market (type II errors).

Limitation of the Classical Approach II

Drawbacks of the Classical Approach:

2. **Rejecting the null does not mean clinically relevant effect size.**
   → Small effects might be clinically irrelevant.

H_0 : Not Different vs. H_1 : Different

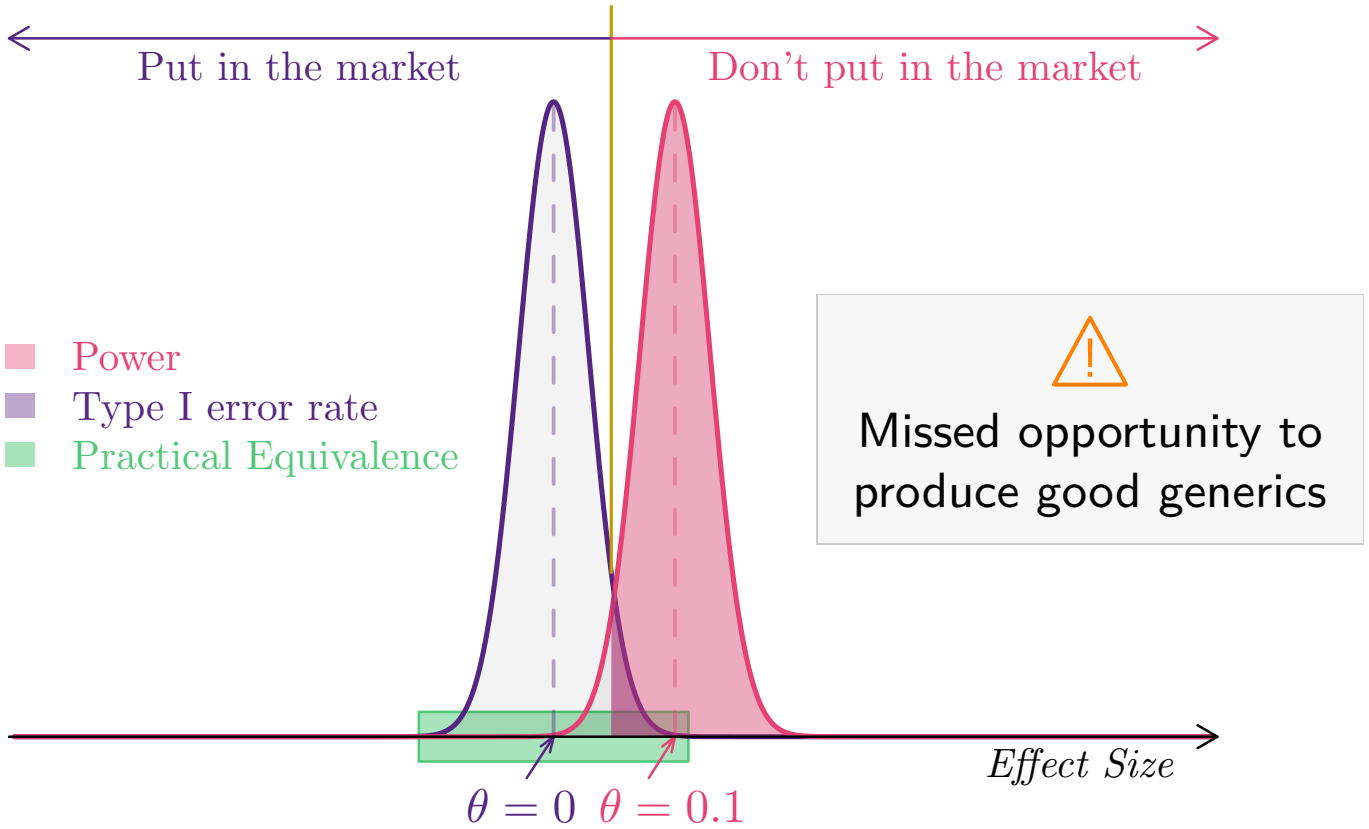


Limitation of the Classical Approach II

Drawbacks of the Classical Approach:

2. **Rejecting the null does not mean clinically relevant effect size.**
   → Small effects might be clinically irrelevant.

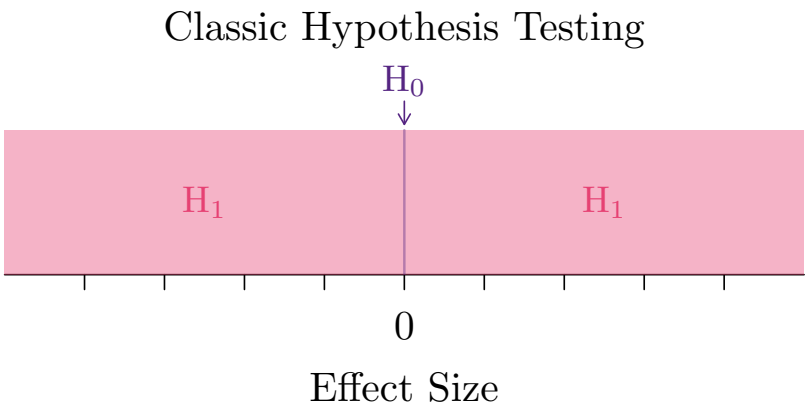
H_0 : Not Different vs. H_1 : Different

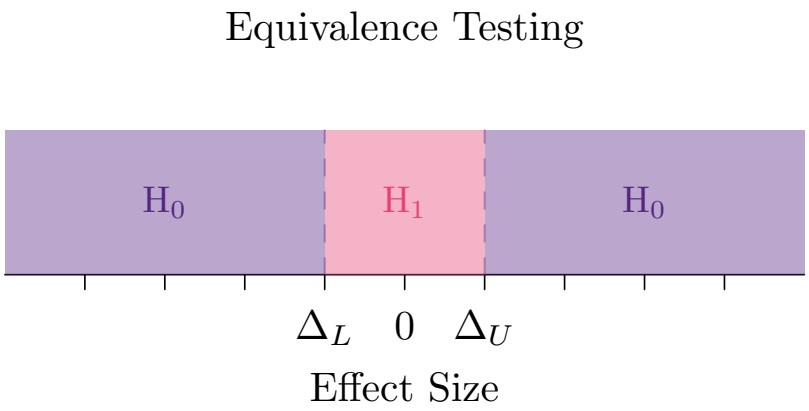


Equivalence Testing

Solution:

1. Put the burden of proof on the drug manufacturer to demonstrate equivalence:
   → Switch the null and the alternative.
2. Define a range for equivalence:
   → An open interval whose bounds correspond to what would be considered the "smallest" clinically relevant effect sizes (in absolute value).



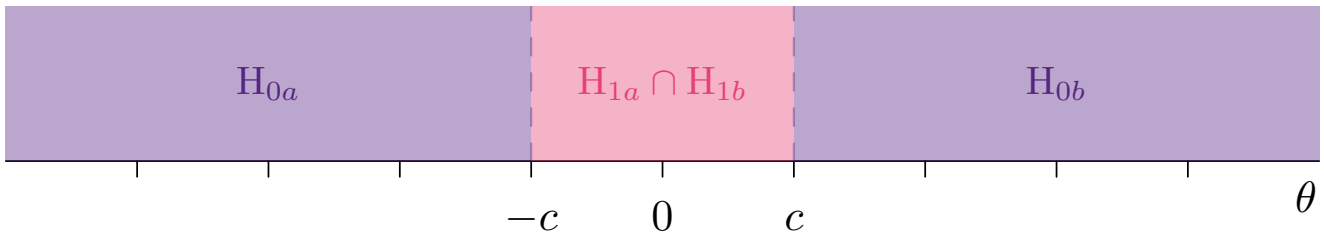


- In practice, equivalence bounds are often taken to be symmetrical around 0, so we define $c := \Delta_U = -\Delta_L$

Equivalence Testing Formally

Formally:

$$\begin{array}{lll} H_{0a} : \theta \leq -c & \textit{or} & H_{0b} : \theta \geq c \\ H_{1a} : \theta > -c & \textit{and} & H_{1b} : \theta < c \end{array}$$



Decision rule by means of confidence intervals (interval inclusion principle): [1]

- Accept $H_{1a} : \theta > -c$ **if** $I^{1-\alpha}_{\theta,U} \not\ni -c$ (Upper one-sided CI),
- Accept $H_{1b} : \theta < c$ **if** $I^{1-\alpha}_{\theta,L} \not\ni c$ (Lower one-sided CI),

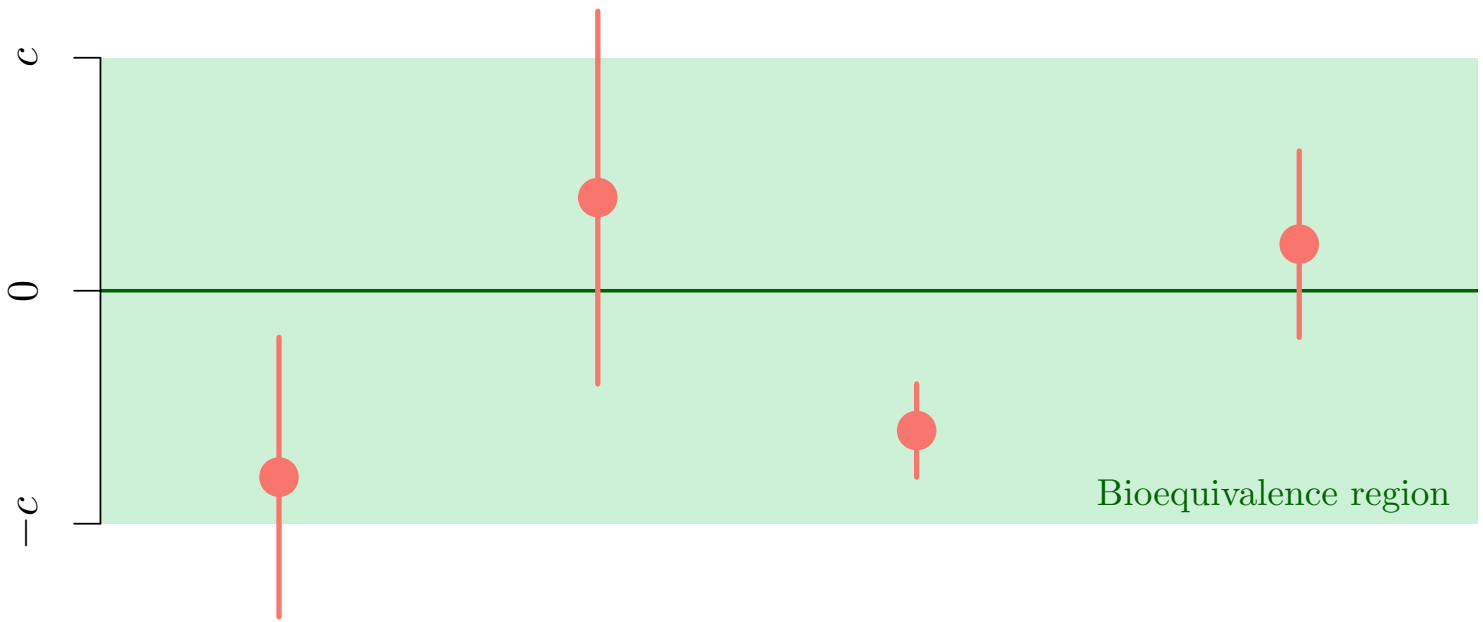
$\rightarrow$ Accept $H_1 := H_{1a} \cap H_{1b}$ **if** $I^{1-2\alpha}_{\theta} := I^{1-\alpha}_{\theta,L} \cap I^{1-\alpha}_{\theta,U} \subset (-c, c)$.

Difference Testing vs. Equivalence Testing

Decision rule by means of confidence intervals (Interval Inclusion Principle):[1]

Case II: (a) Not different
(b) Nonequivalent

Case IV: (a) Not different
(b) Equivalent



Case I: (a) Different
(b) Nonequivalent

Case III: (a) Different
(b) Equivalent

RESEARCH ARTICLE

Statistics in Medicine WILEY

Finite sample corrections for average equivalence testing

Younes Boulaguiem[1] | **Julie Quartier**[2,3] | **Maria Lapteva**[2,3] | **Yogeshvar N. Kalia**[2,3] | **Maria-Pia Victoria-Feser**[1] | **Stéphane Guerrier**[1,2,3] | **Dominique-Laurent Couturier**[4,5]

[1]Geneva School of Economics and Management, University of Geneva, Switzerland

[2]School of Pharmaceutical Sciences, University of Geneva, Switzerland

[3]Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Switzerland

[4]Medical Research Council Biostatistics Unit, University of Cambridge, England

[5]Cancer Research UK – Cambridge Institute, University of Cambridge, England

Correspondence
Dominique-Laurent Couturier, Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge, UK.
Email:

Average (bio)equivalence tests are used to assess if a parameter, like the mean difference in treatment response between two conditions for example, lies within a given equivalence interval, hence allowing to conclude that the conditions have "equivalent" means. The *two one-sided tests* (TOST) procedure, consisting in testing whether the target parameter is respectively significantly greater and lower than some pre-defined lower and upper equivalence limits, is typically used in this context, usually by checking whether the confidence interval for the target parameter lies within these limits. This intuitive and visual procedure is however known to be conservative, especially in the case of highly variable drugs, where it shows a rapid power loss, often reaching zero, hence making it impossible to conclude for equivalence when it is actually true. Here, we propose a finite sample correction of the TOST procedure, the α-TOST, which consists in a correction of the significance level of the TOST allowing to guarantee a test size (or type-I error rate) of α. This new procedure essentially

Two One-Sided Tests (TOST)

Canonical form:

$$\widehat{\theta} \sim \mathcal{N}\left(\theta, \sigma_\nu^2\right) \quad \text{and} \quad \frac{\nu\widehat{\sigma}_\nu^2}{\sigma_\nu^2} \sim \chi_\nu^2,$$

where $\sigma_\nu^2 := \sigma^2 / n$ and σ^2 denotes the asymptotic variance.

The TOST[2] is based on two test statistics:

$$T_L := \frac{\widehat{\theta} + c}{\widehat{\sigma}_\nu} \sim t_\nu\left(\frac{\theta + c}{\sigma_\nu}\right) \quad \text{and} \quad T_U := \frac{\widehat{\theta} - c}{\widehat{\sigma}_\nu} \sim t_\nu\left(\frac{\theta - c}{\sigma_\nu}\right).$$

Two One-Sided Tests (TOST)

Decision rule by means of the TOST:

- Accept $H_{1a} : \theta > -c$ **if** $T_L \geq t_{\alpha,\nu}$
- Accept $H_{1b} : \theta < c$ **if** $T_U \leq -t_{\alpha,\nu}$,

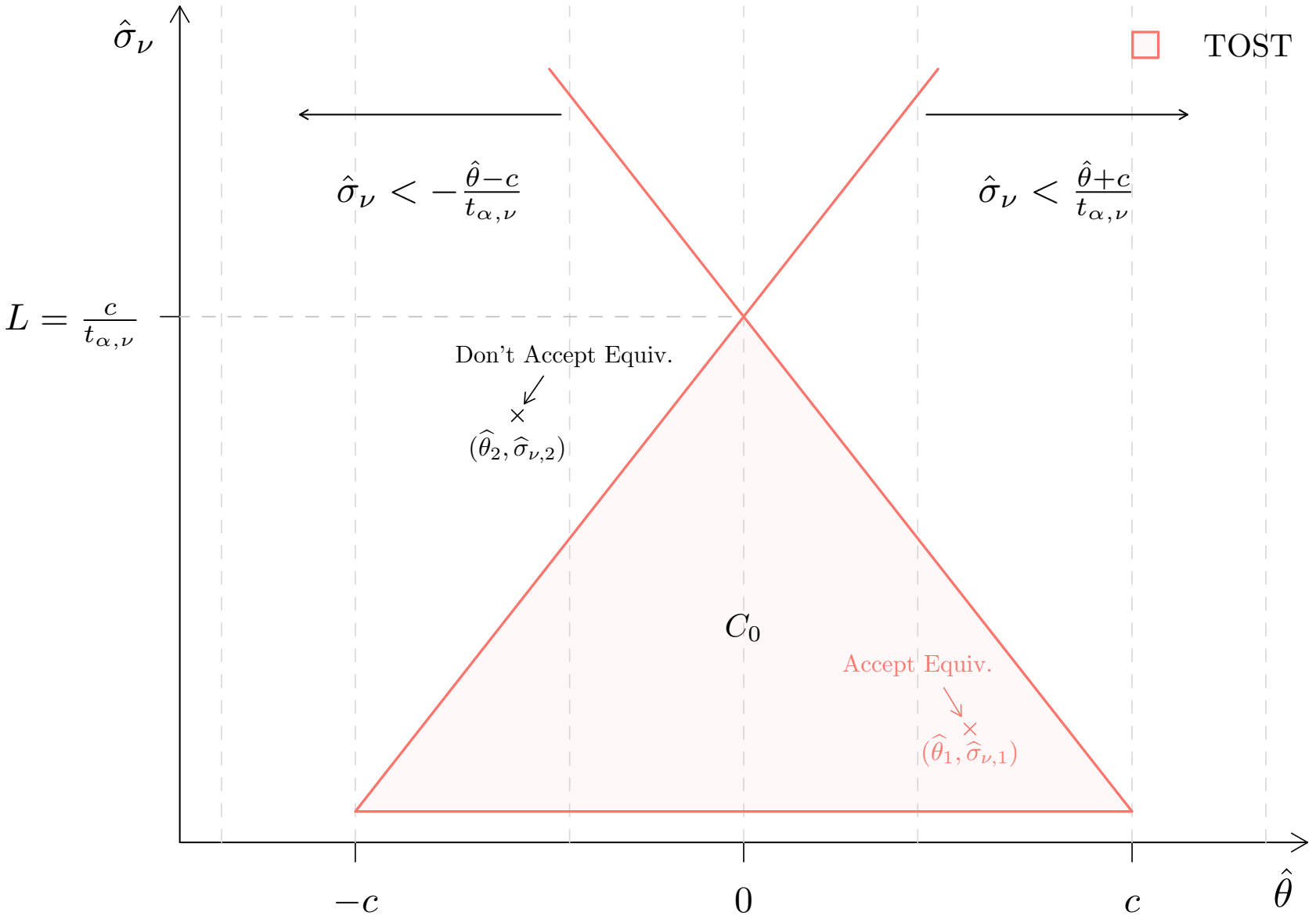
$\rightarrow$ Accept $H_1 := H_{1a} \cap H_{1b}$ **if** both tests reject their marginal nulls.

Rejection Region by rearranging the terms:

$$C_0 := \left\{ \widehat{\theta} \in , \widehat{\sigma}_\nu > 0 \,\middle|\, |\widehat{\theta}| \leq c - t_{\alpha,\nu}\widehat{\sigma}_\nu \right\}.$$

Accept equivalence **if** $(\widehat{\theta}, \widehat{\sigma}_\nu) \in C_0$.

TOST: Rejection Region



Size of the TOST

Power function:

$$p(\alpha, \theta, \sigma_\nu, \nu, c) := \Pr(\text{ Reject } H_0 \text{ }) = \Pr\left(|\hat{\theta}| \leq c - t_{\alpha,\nu}\hat{\sigma}_\nu\right)$$

$$= \mathbb{E}_{\hat{\sigma}_\nu|\sigma_\nu}\left[\Pr\left(|\hat{\theta}| \leq c - t_{\alpha,\nu}\hat{\sigma}_\nu \mid \hat{\sigma}_\nu\right)\right].$$

Size function:

- The size is defined as **the largest probability** achieved under the **space of the null**.
- The **further outside** the equivalence region θ is, **the less likely** we are to reject the null (non-equivalence). Therefore the sup is achieved at $-c$ or c (p is even in θ).

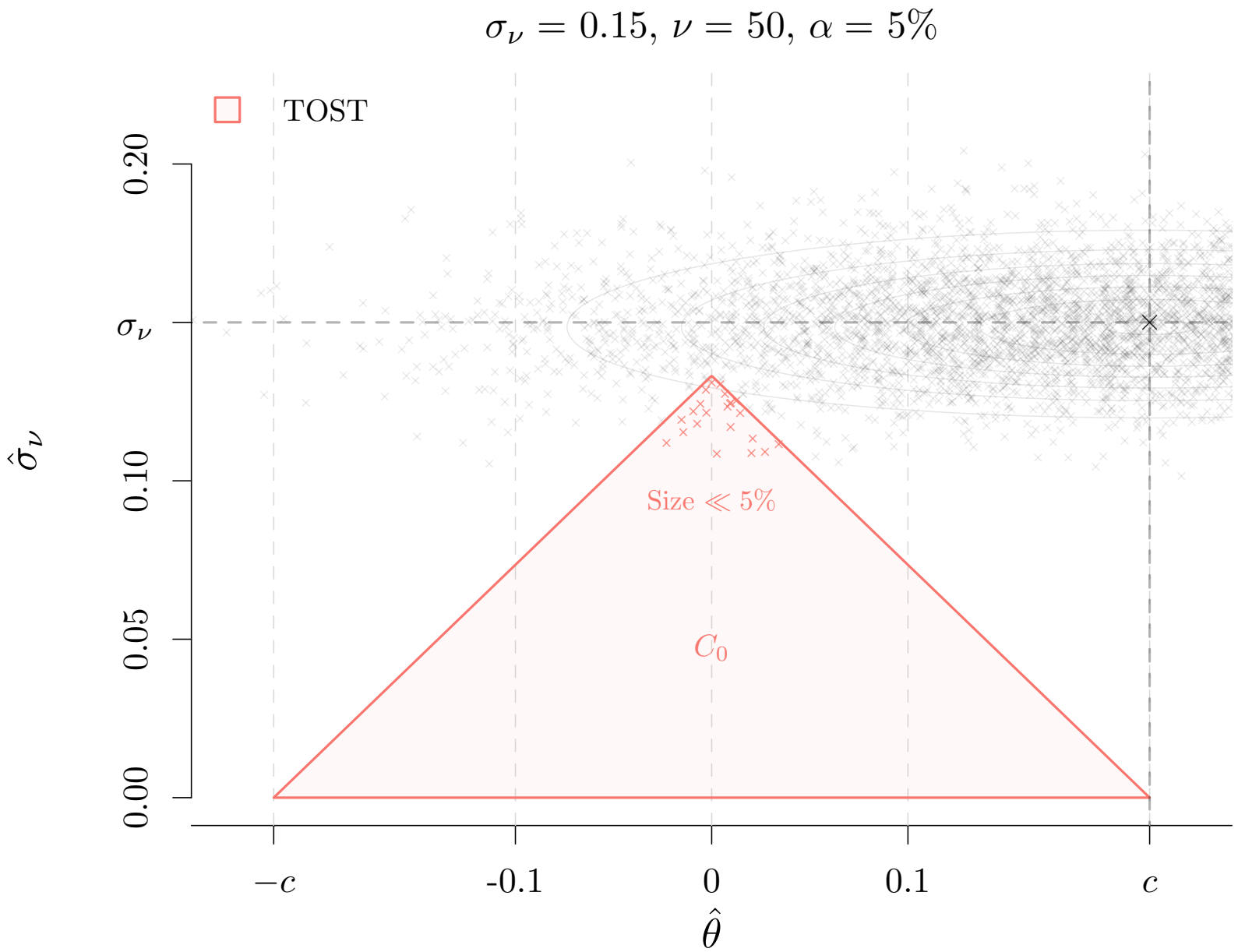
$$\omega(\alpha, c, \sigma_\nu, \nu) := \Pr(\text{ Reject } H_0 \mid H_0 \text{ })$$

$$= \sup_{\theta \notin (-c,c)} \Pr(\text{ Reject } H_0 \text{ }) = p(\alpha, \text{ } c \text{ }, \sigma_\nu, \nu, c)$$

$$< \Pr\left(T_\nu \leq -t_{\alpha,\nu}\right) = \alpha$$

$\Rightarrow$ The TOST is conservative

Conservativeness of the TOST



Corrective Approaches

Recalling the size function, we can operate on two parameters:

$$\omega\Big(\ \alpha\ ,\ c\ ,\sigma_\nu,\nu\Big)$$

α-TOST:

$$\alpha^* := \underset{\gamma\in[\alpha,0.5]}{\text{argzero}}\ \big[\omega(\ \gamma\ ,c,\sigma_\nu,\nu)-\alpha\big]$$

- A unique solution exits provided:

$$\sigma_\nu < \frac{2c}{\Phi^{-1}(\alpha+0.5)},$$

- Converges exponentially fast:

$$\left|\alpha^{*(k+1)}-\alpha^*\right| < \frac{1}{2}\exp(-bk).$$

- Yields the rejection region:

$$C_1 := \left\{\hat\theta\in\ ,\ \widehat\sigma_\nu>0\ \middle|\ |\hat\theta|\le c-t_{\alpha^*,\nu}\widehat\sigma_\nu\right\}.$$

δ-TOST:

$$\delta^* := \underset{\delta\in[c,\infty)}{\text{argzero}}\ \big[\omega(\alpha,\ \delta\ ,\sigma_\nu,\nu)-\alpha\big]$$

- A unique solution always exist:

$$\omega(c)<\alpha,\qquad \lim_{\delta\to\infty}\omega(\delta)=1,$$

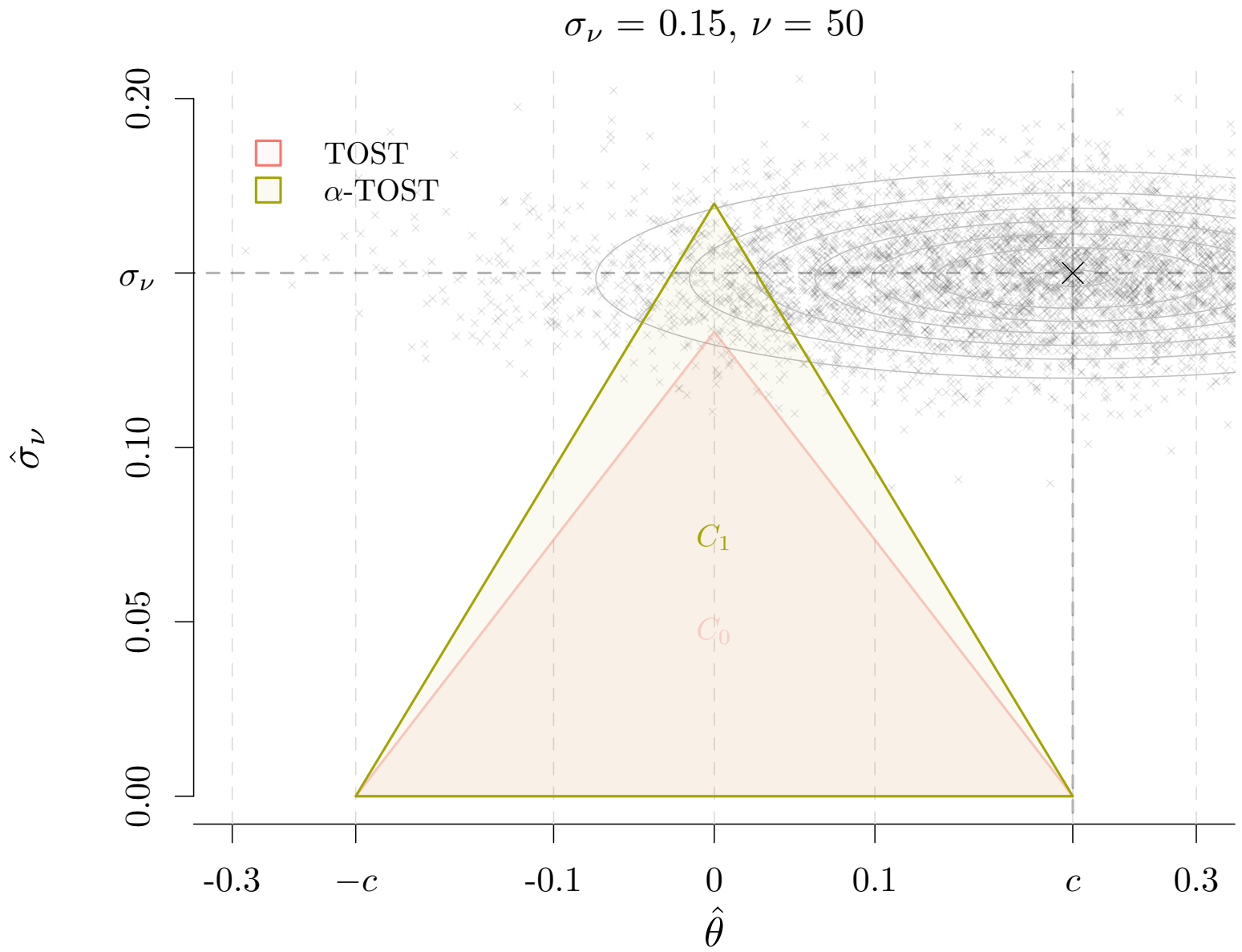
- No contraction.

→ Can be solved using general-purpose optimization methods.

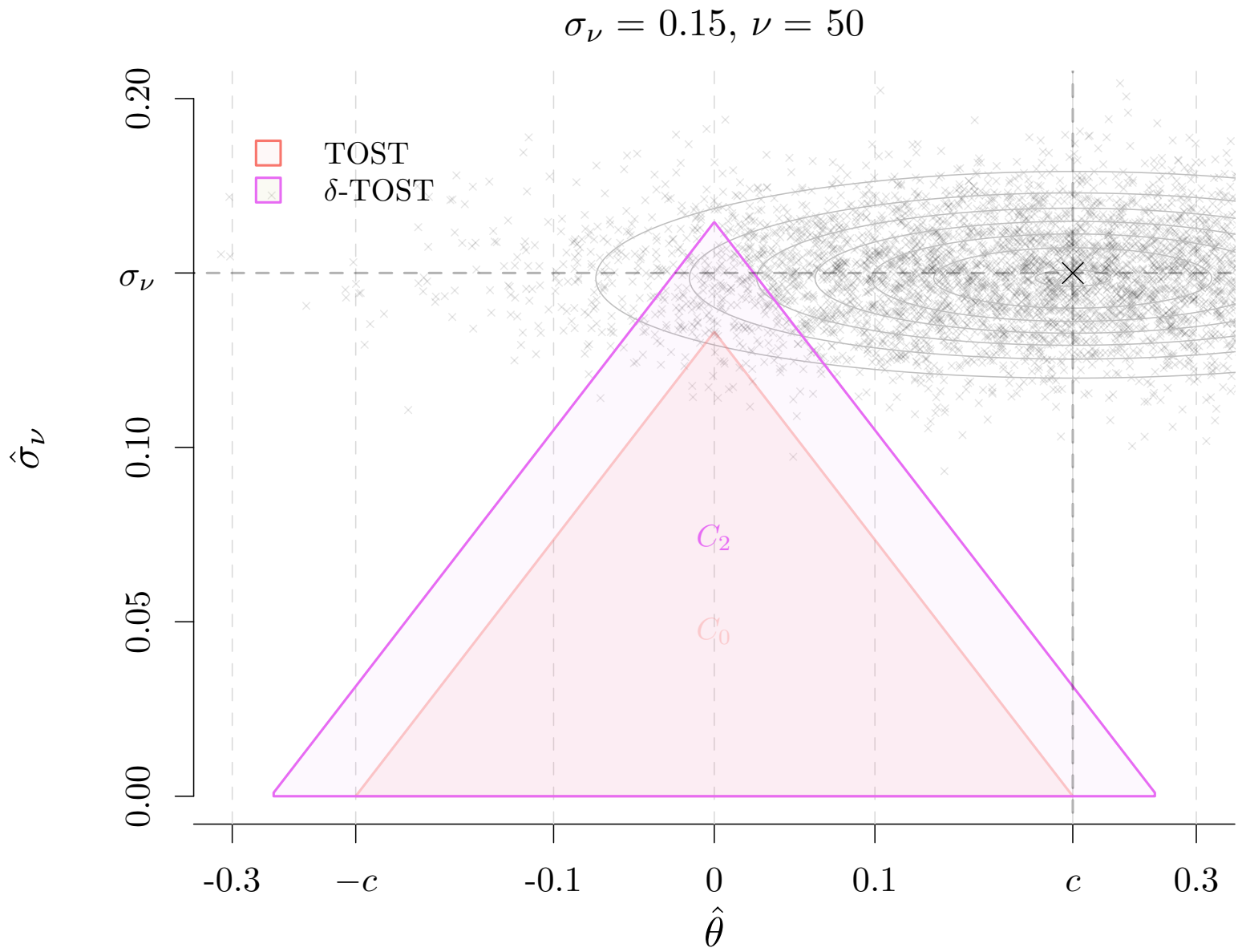
- Yields the rejection region:

$$C_2 := \left\{\hat\theta\in\ ,\ \widehat\sigma_\nu>0\ \middle|\ |\hat\theta|\le \delta^*-t_{\alpha,\nu}\widehat\sigma_\nu\right\}.$$

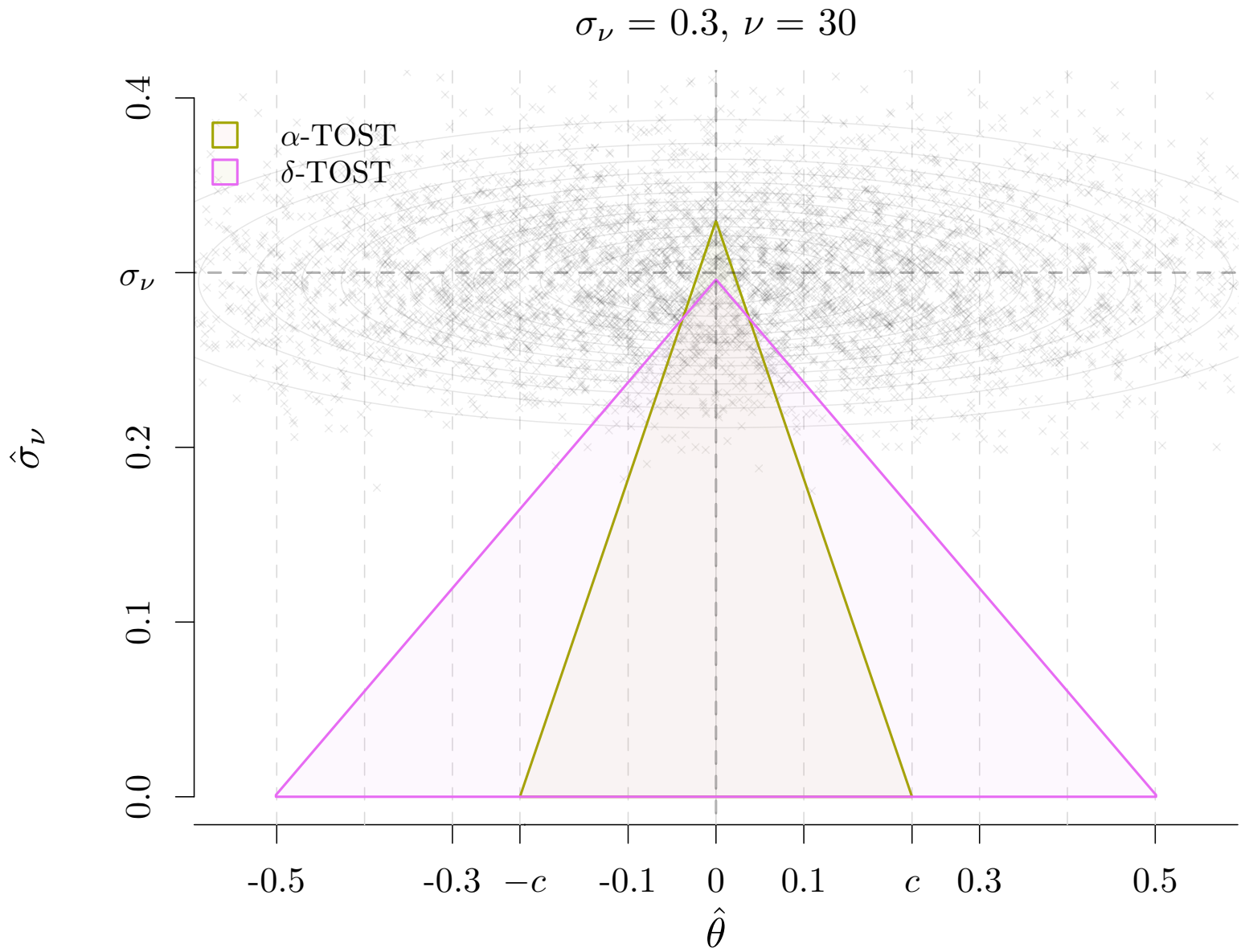
Rejection Region: α-TOST



Rejection Region: δ-TOST



Power at $\theta = 0$: α-TOST vs. δ-TOST



The α-TOST Empirically

When σ_ν is unknown:

- Replace it by its empirical $\widehat{\sigma}_\nu$ and solve

$$\widehat{\alpha}^* := \alpha^*(\widehat{\sigma}_\nu) = \underset{\gamma \in [\alpha, 0.5]}{\text{argzero}} \left[\omega(\gamma, c, \widehat{\sigma}_\nu, \nu) - \alpha \right].$$

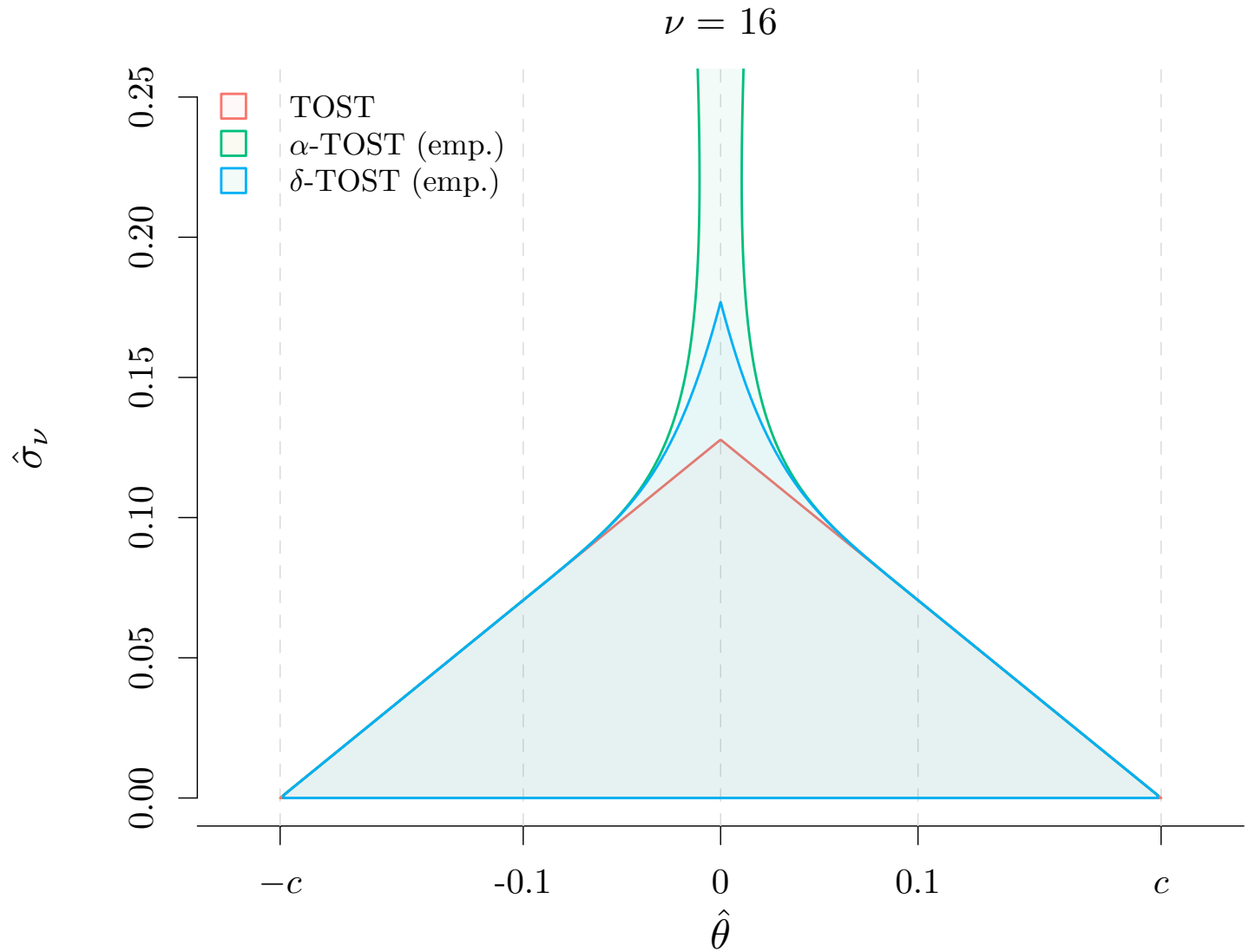
- Conditions for existence and computational efficiency remain unchanged.

Large sample behaviour:

$$\widehat{\alpha}^* = \alpha^* + o_p\left(\nu^{-1}\right), \quad \widehat{\sigma}_\nu = \sigma_\nu + \mathcal{O}_p\left(\nu^{-1}\right), \quad \widehat{\theta} = \theta + \mathcal{O}_p\left(\nu^{-1/2}\right).$$

$\rightarrow$ Uncertainty related to $\widehat{\alpha}^*$ is asymptotically negligible compared to that of $\widehat{\theta}$ and $\widehat{\sigma}_\nu$.

Rejection Areas: empirical α-TOST vs. empirical δ-TOST



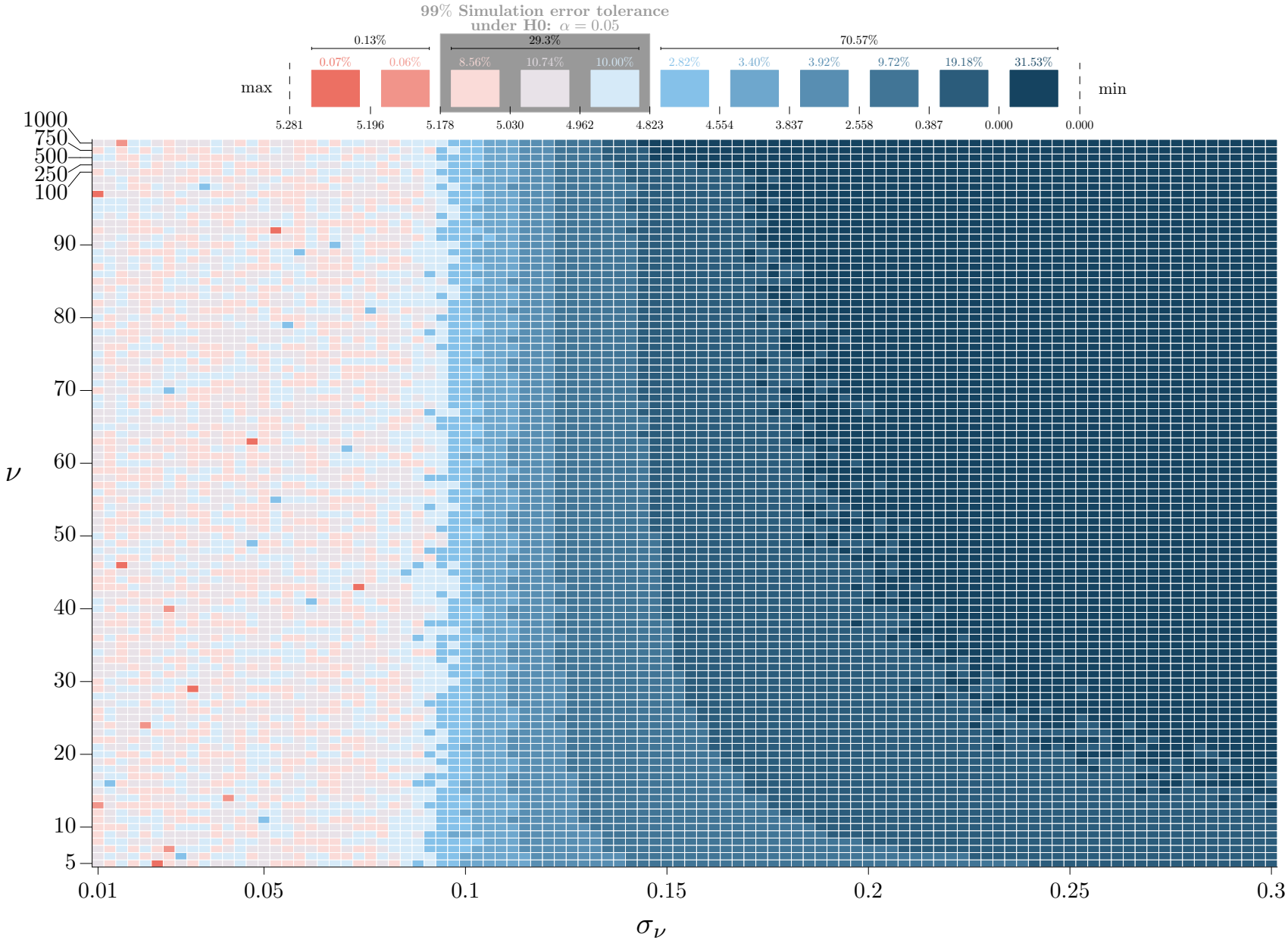
The α-TOST empirical rejection region **entirely includes** the other two.

$\rightarrow$ The α-TOST is **uniformly more powerful**.

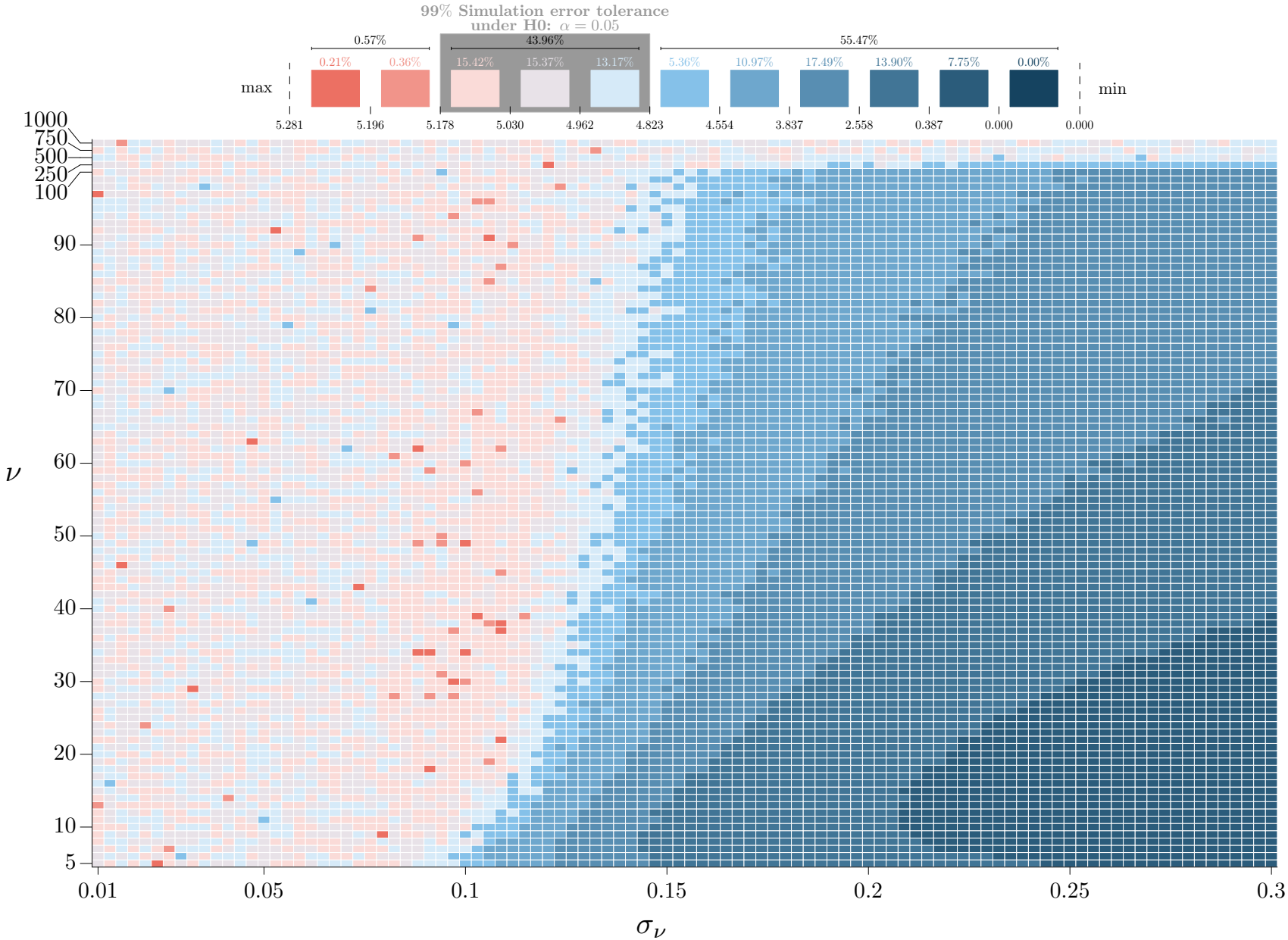
Simulation Study

	Simulation settings	
	Size	**Power**
c	$\log(1.25) \approx 0.2231$	
ν	5, 6, 7, . . . , 100, 250, 500, 750, 1000 (100 values)	
σ_ν	100 evenly spaced values between 0.01 and 0.3	
θ	c	0
α	0.05	
B	10^5	
Design	General (canonical form)	
Methods	TOST, α-TOST and δ-TOST	

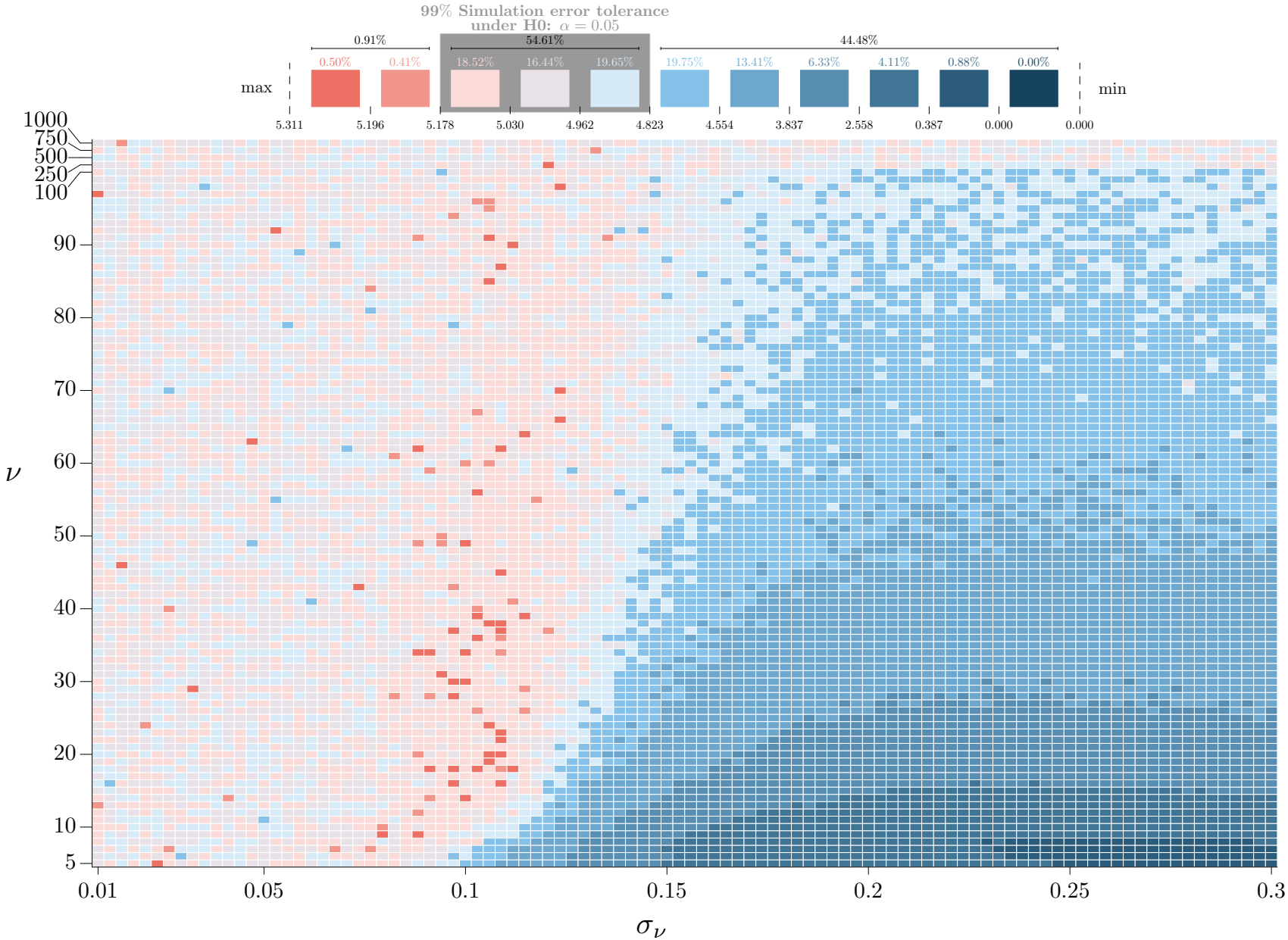
Simulation Study: TOST Empirical Size



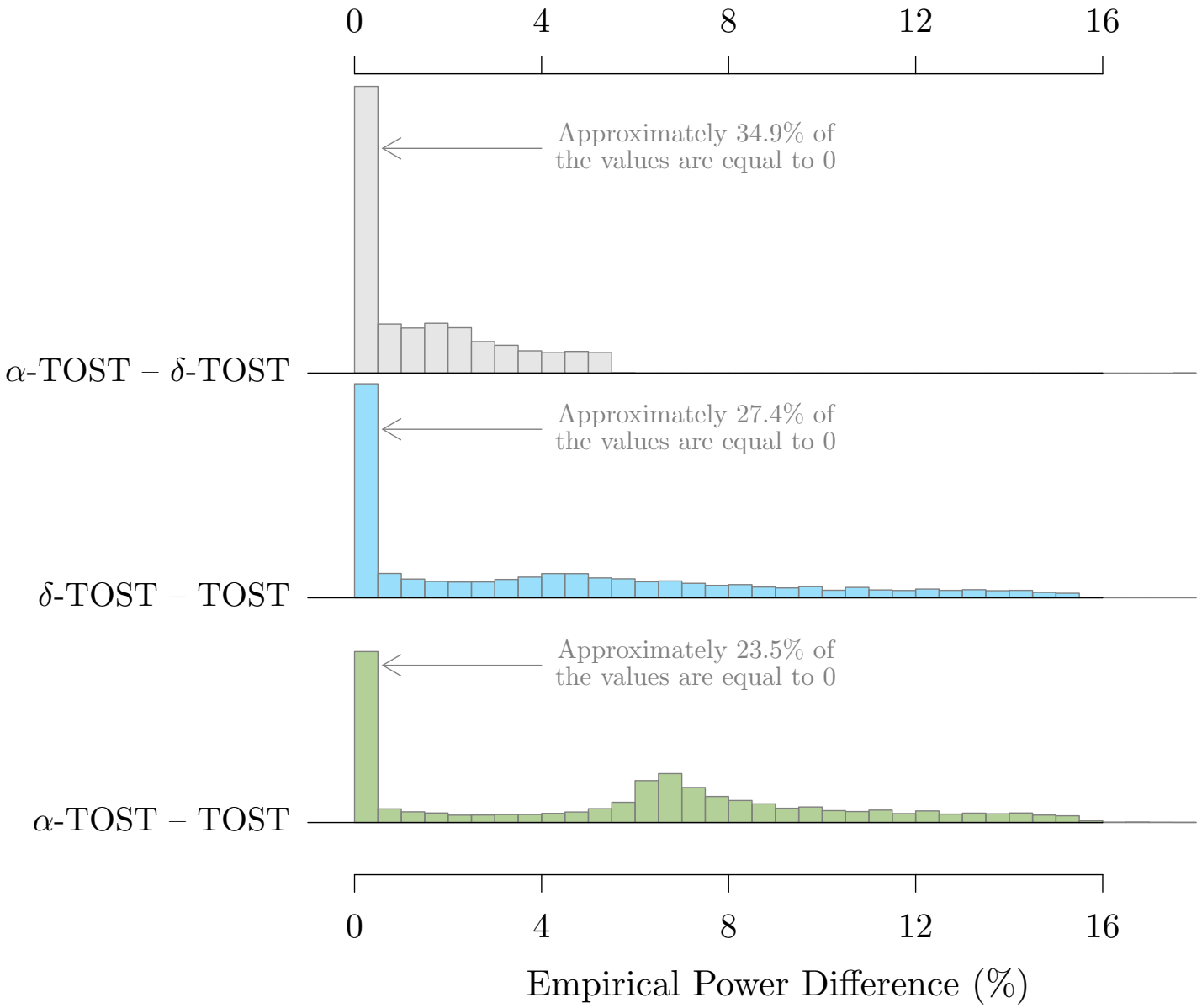
Simulation Study: δ-TOST Empirical Size



Simulation Study: α-TOST Empirical Size

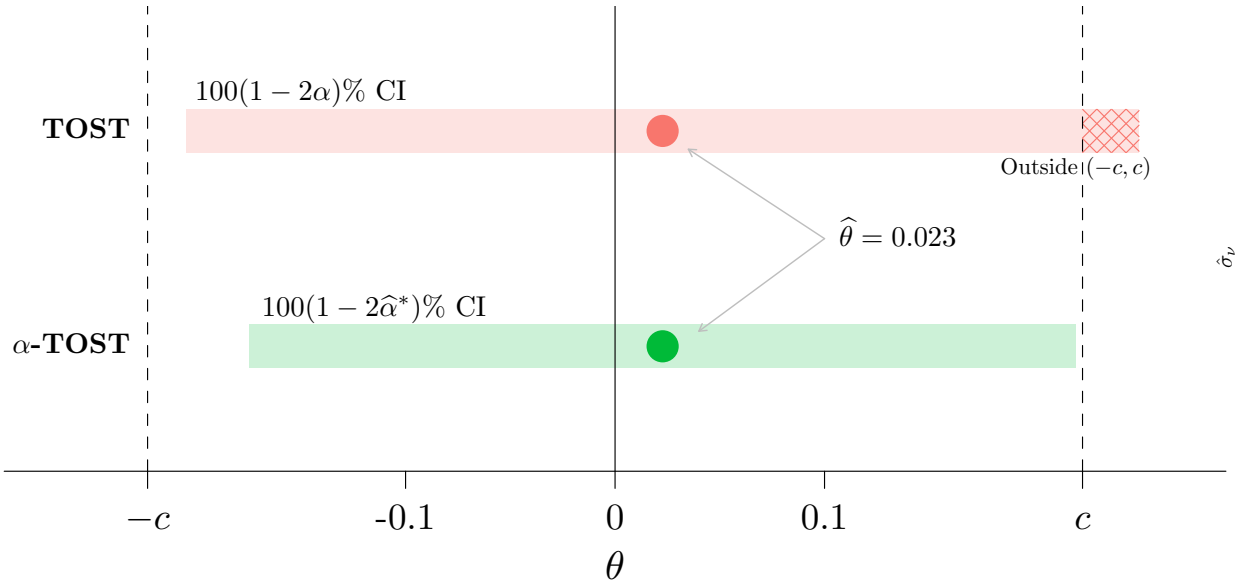


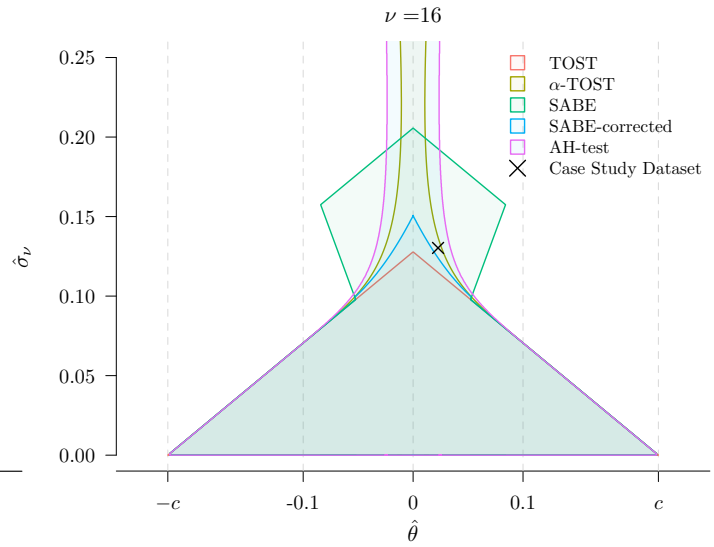
Simulation Study: Empirical Power at $\theta = 0$



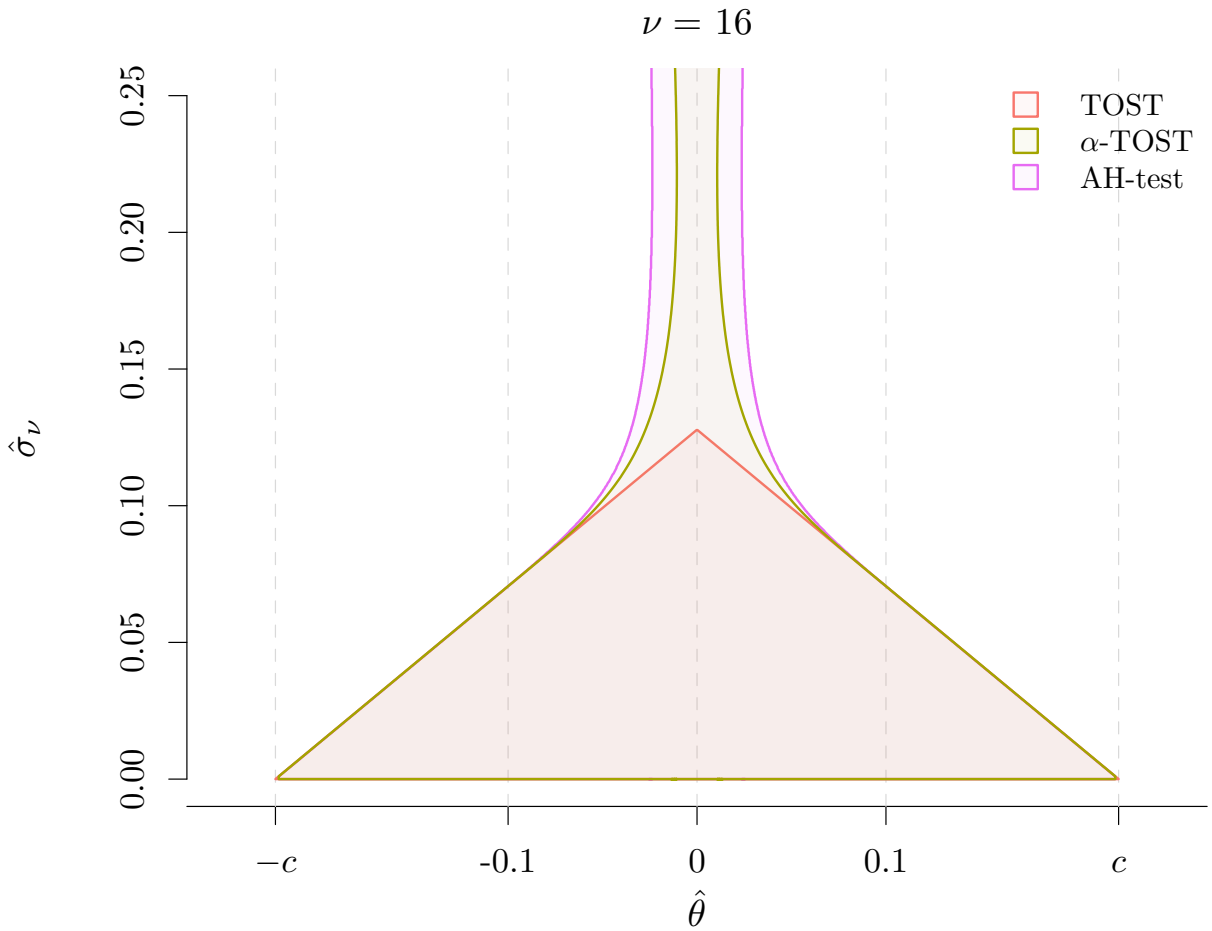
Case Study: Econazole Nitrate Deposition

- **Objective:** Assess bioequivalence of econazole nitrate deposition from a reference and an already approved generic cream.
- **Design:** 17 **paired** porcine skin samples.
- **Purpose:**
  1. Our method is **design-agnostic**, unlike SABE-like methods that require replicate designs.[3]
  2. Our method can produce **confidence intervals**, unlike other methods like the AH-test, Brown et al., and Berger & Hsu.[4]





α-TOST: The Right Compromise



The AH-test is know to be quite liberal. In his TOST paper,[2] *Schuirmann* said the following after comparing the two methods:

> *"[...] The best procedure to use may therefore turn out to be a compromise between the two procedures."*

Multivariate Adjustments for Average Equivalence Testing

Younes Boulaguiem[1] | Luca Insolia[1] | Maria-Pia Victoria-Feser[2] | Dominique-Laurent Couturier[3,4] | Stéphane Guerrier[1,5,6]

[1]Geneva School of Economics and Management, University of Geneva, Geneva, Switzerland | [2]Department of Statistical Sciences, University of Bologna, Bologna, Italy | [3]Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge, UK | [4]Cancer Research UK, Cambridge Institute, University of Cambridge, Cambridge, UK | [5]School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland | [6]Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Geneva, Switzerland

Correspondence: Stéphane Guerrier (stephane.guerrier@unige.ch)

Received: 24 April 2024 | **Revised:** 2 October 2024 | **Accepted:** 7 October 2024

Funding: Dominique-Laurent Couturier is supported by UK Medical Research Council grant number MC_UU_00040/03 and MC_UU_00002/14, and by the Cancer Research UK grant C9545/A29580. Luca Insolia and Stéphane Guerrier are supported by the SNSF Grants #176843 and #211007 as well as by the Innosuisse Grants #37308.1 IP-ENG and #53622.1 IP-ENG. Younes Boulaguiem and Maria-Pia Victoria-Feser are partially supported by the SNSF Grant #182684.

Keywords: finite-sample adjustments | hypothesis testing | interval-inclusion principle | multivariate bioequivalence | two one-sided tests

ABSTRACT

Multivariate (average) equivalence testing is widely used to assess whether the means of two conditions of interest are "equivalent" for different outcomes simultaneously. In pharmacological research for example, many regulatory agencies require the generic product and its brand-name counterpart to have equivalent means both for the AUC and $C_{\max}$ pharmacokinetics parameters. The multivariate Two One-Sided Tests (TOST) procedure is typically used in this context by checking if, outcome by outcome, the marginal $100(1-2\alpha)\%$ confidence intervals for the difference in means between the two conditions of interest lie within predefined lower and upper equivalence limits. This procedure, already known to be conservative in the univariate case, leads to a rapid power loss when the number of outcomes increases, especially when one or more outcome variances are relatively large. In this work, we propose a finite-sample adjustment for this procedure, the multivariate α-TOST, that consists in a correction of α, the significance level, taking the (arbitrary) dependence between the outcomes of interest into account and making it uniformly more powerful than the conventional multivariate TOST. We present an iterative algorithm allowing to efficiently define α^*, the cor-

Extension to Multivariate Setting

Canonical form: $\hat{\boldsymbol{\theta}} \sim \mathcal{N}_m(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ and $\nu\widehat{\boldsymbol{\Sigma}} \sim \mathcal{W}_m(\nu, \boldsymbol{\Sigma})$.

Equivalence region: $\boldsymbol{\Theta}_1 := \{x \in {}^m \mid -c < x_j < c,\, j = 1, \ldots, m\}$.

Rejection region: $C_0(\widehat{\boldsymbol{\Sigma}}) := \bigcap_{j=1}^m \left\{|\hat{\theta}_j| \leq c - t_{\alpha,\nu}\hat{\sigma}_j\right\}$.

The main issue: the TOST's conservativeness is further exacerbated by the increased dimensionality (to the point of flat power curves over the whole parameter space).

Multivariate α-TOST: the argsup $\lambda(\alpha, \boldsymbol{\Sigma}) = \arg\sup_{\boldsymbol{\theta} \notin \boldsymbol{\Theta}_1} p(\alpha, \boldsymbol{\theta}, \boldsymbol{\Sigma}, \nu, c)$ depends on α and $\boldsymbol{\Sigma}$. We prove existence of α^* under mild conditions and solve

$$\alpha^*(\boldsymbol{\Sigma}) = \underset{\gamma \in [\alpha, 0.5]}{\text{argzero}} \left[p(\gamma, \lambda(\gamma, \boldsymbol{\Sigma}), \boldsymbol{\Sigma}, \nu, c) - \alpha\right],$$

which requires iterating between γ-calibration and λ-optimization.

Result: the multivariate α-TOST is uniformly more powerful and more accurate than the mvt TOST and norm-based alternatives empirically.

Equivalence Assessment of Topical Products



2 December 2014
EMA/CHMP/QWP/558185/2014
Committee for Medicinal Products for Human use (CHMP)

Concept paper on the development of a guideline on quality and equivalence of topical products

Challenges:

1. Blood-based PK methods fail: drug concentration often **untraceable in blood** (only side effects remain).

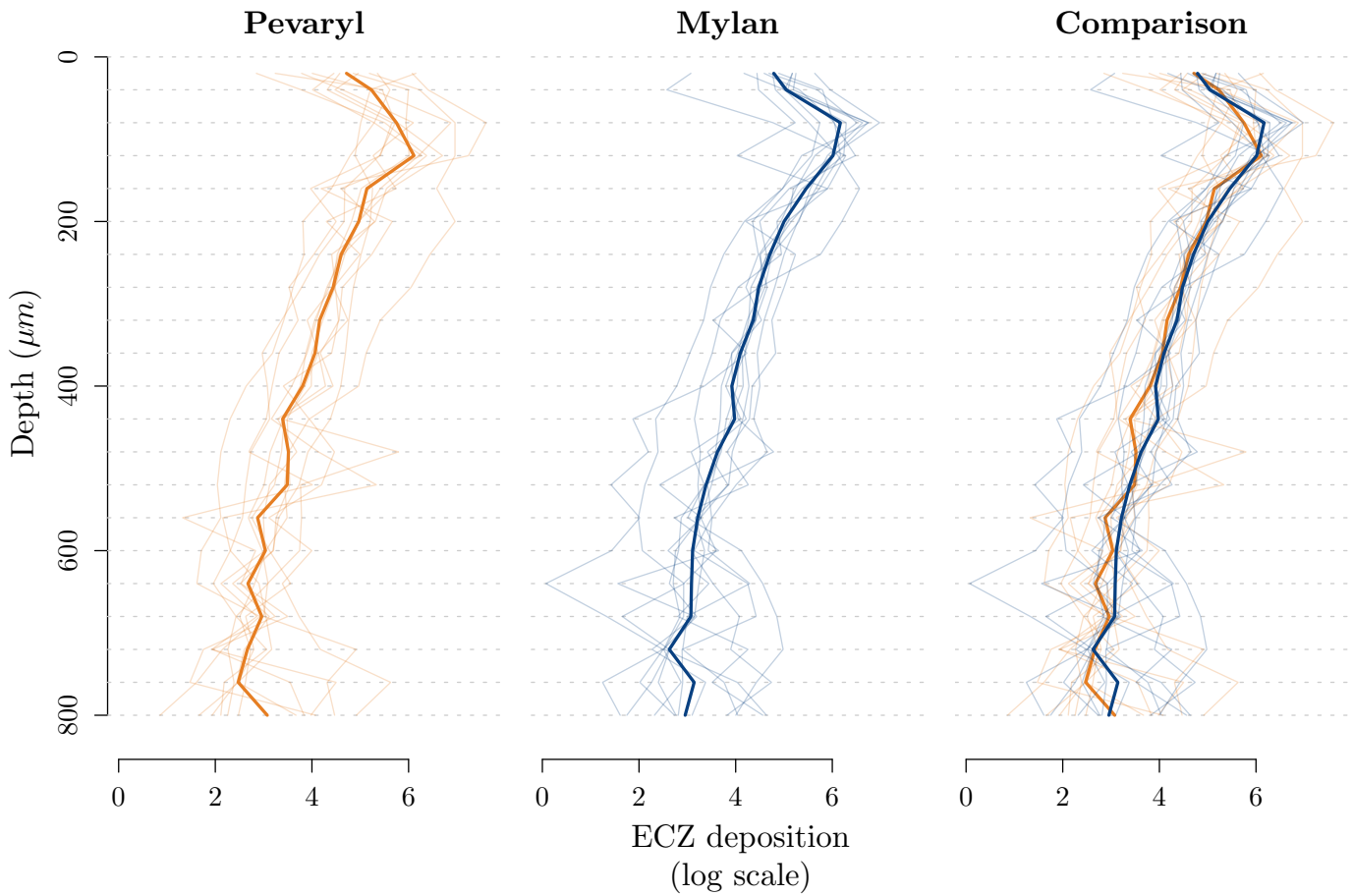
→ Need to assess bioavailability directly across **multiple skin targets**.

2. Need for appropriate statistical tools.

Cutaneous Biodistribution

Solutions:

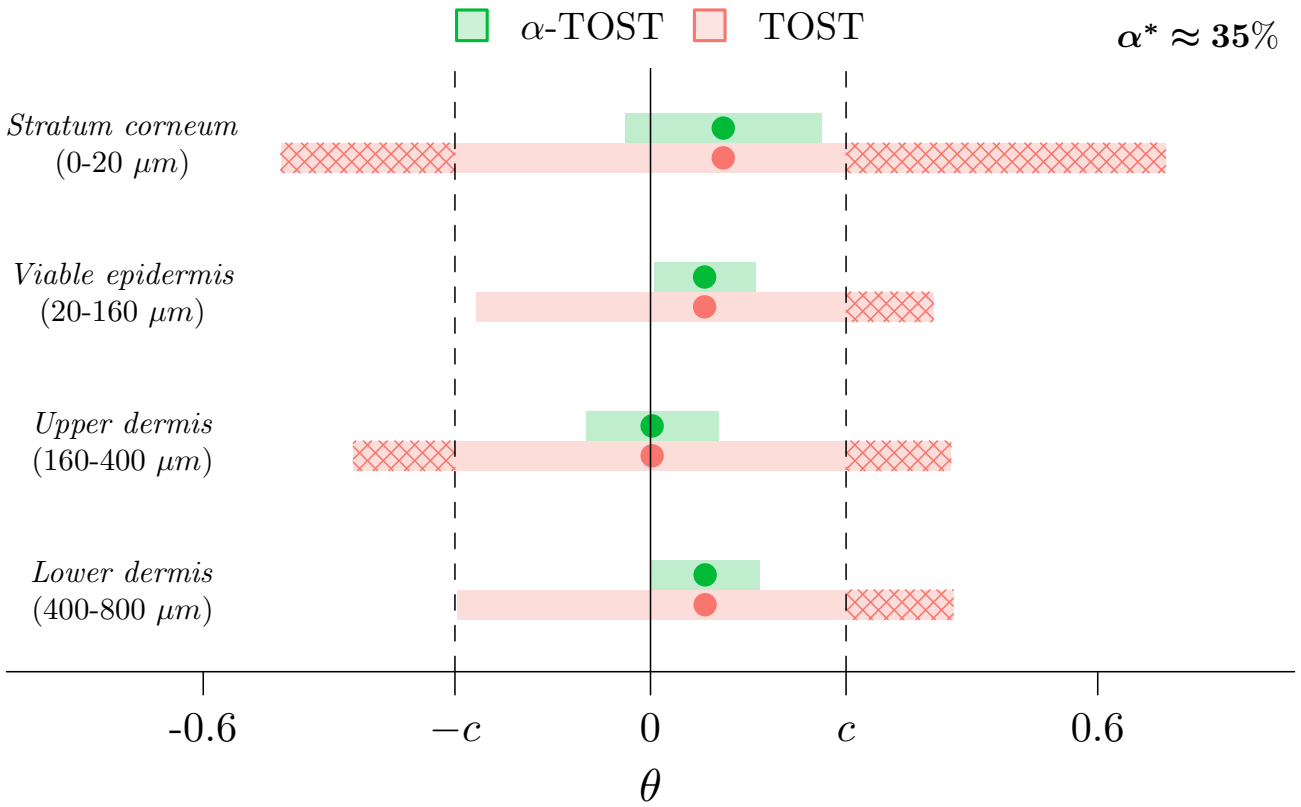
1. **Cutaneous Biodistribution**[5] provides spatial distribution of the drug across the skin.



Econazole Nitrate: Biodistribution + MVT α-TOST

Solutions:

1. **Cutaneous Biodistribution**[5] provides spatial distribution of the drug across the skin.
2. **Mvt. α-TOST** assesses mvt. bioequivalence



Increasingly Relevant: New EMA Draft Guidelines



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

5 October 2026
EMA/220888/2026
Committee for Medicinal Products for Human Use (CHMP)

Guideline on the requirements for demonstrating therapeutic equivalence for nasal products

What this draft asks (in vitro)		**Why the α-TOST helps**
90% CIs within $\pm$**10–15%**, from as few as **3 batches** $\times$ **10 units**.	$\Rightarrow$	TOST is conservative here: **α-TOST restores size α**.
All quality attributes must pass: droplet size, dose, spray pattern...	$\Rightarrow$	Correlated endpoints: **multivariate α-TOST**.
Limits widen by a fixed **SABE** rule (threshold, cap): **can be liberal** (size $> \alpha$).	$\Rightarrow$	Correction **derived from the size**: **continuous, exact size α**, any design.

Tight limits, small n, many endpoints: **exactly where the TOST is weakest.**

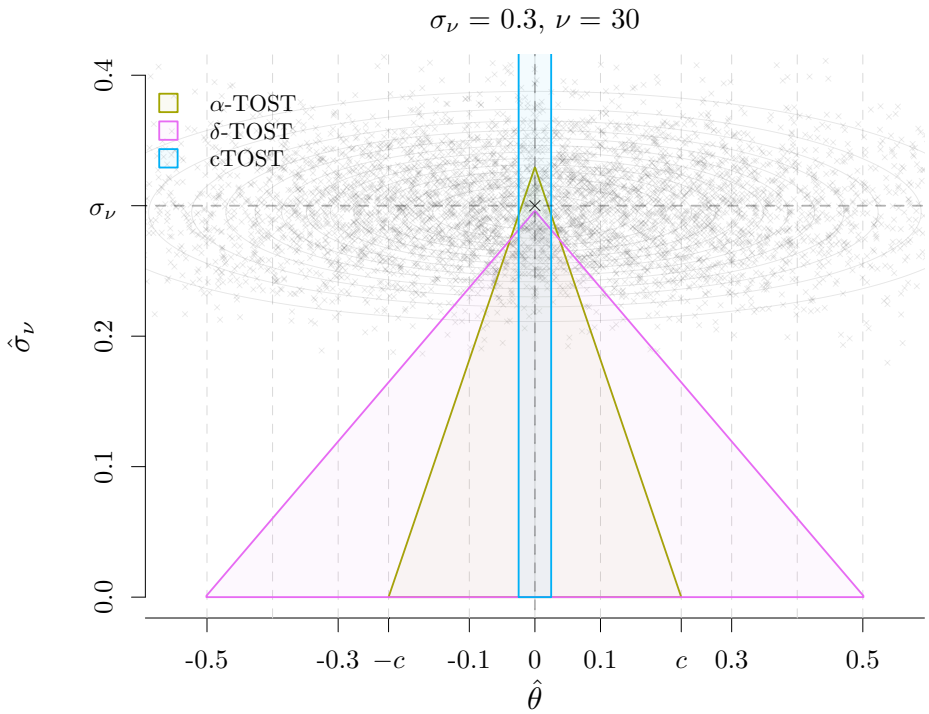
Towards Optimal Equivalence Testing

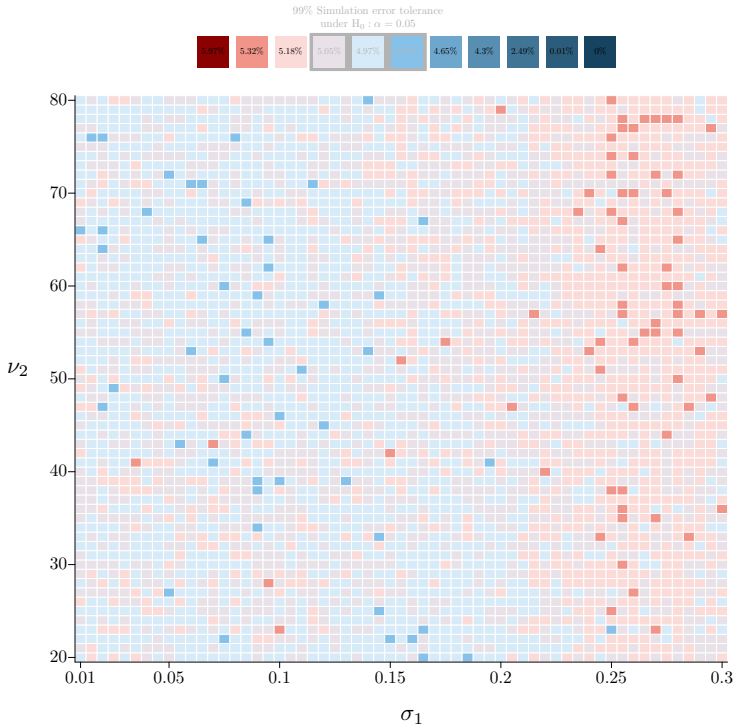
Can we do even better?

- Yes! By simultaneously optimizing for α and c:

$$[\alpha^*, c^*] \in \mathcal{A} = \underset{\substack{\gamma \in (0, 1/2] \\ \delta \geq 0}}{\text{argzero}}\ \omega(\gamma, \delta, \sigma_\nu, \nu) - \alpha.$$

- We show that fixing $\alpha^* = 0.5$ and optimizing for c yields the uniformly most powerful test. We call it cTOST.





Towards Optimal Equivalence Testing

Can we do even better?

- Yes! By simultaneously optimizing for α and c:

$$[\alpha^*, c^*] \in \mathcal{A} = \underset{\substack{\gamma \in (0,1/2] \\ \delta \geq 0}}{\text{argzero}}\ \omega(\gamma, \delta, \sigma_\nu, \nu) - \alpha.$$

- We show that fixing $\alpha^* = 0.5$ and optimizing for c yields the uniformly most powerful test. We call it cTOST.

Bioequivalence Assessment for Locally Acting Drugs: A Framework for Feasible and Efficient Evaluation

Luca Insolia[1-3], Yanyuan Ma[4],

Younes Boulaguiem[5] & Stéphane Guerrier[1-3,*]

[1]School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland; [2]Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Geneva, Switzerland; [3]Department of Earth Sciences, University of Geneva, Geneva, Switzerland; [4]Department of Statistics, The Pennsylvania State University, University Park, Pennsylvania, USA; [5]Geneva School of Economics and Management, University of Geneva, Geneva, Switzerland.

Limitations, Extensions, and Other Useful Applications

Reliance on Canonical form:

- Strong assumption although widely used in practice.

  $\Rightarrow$ **Extension to non-parametric** framework possible.
  e.g., bootstrap + KDE to approximate rejection rate.

Underestimation of variability in planning:

- BE trials may narrowly fail due to underestimated variance.

  $\Rightarrow$ **α-TOST can recover BE** by slightly shrinking CIs.

A decision tool for sponsors:

- At interim: compare **conditional power** under TOST vs α-TOST.

  ▷ **Large gap**: shortfall is TOST **geometry**, **not weak signal** $\rightarrow$ continue or resize.

  ▷ **Both low**: weak signal $\rightarrow$ consider stopping for **futility**.

cTOST R Package is in CRAN and GitHub!



→ Example of output:

```
mtost = tost(theta = theta_hat, sigma = Sigma_hat, nu = nu, delta = log(1.25))
mtost

✖ Can't accept (bio)equivalence
Equiv. Region:   |----------------0----------------|
t_half                (----------x----------)
AUC                 (-------x--------)
AUC_inf             (--------x-------)
C_max             (---------x---------)

CIs:
t_half   (-0.15767 ; 0.12503)     ✔
AUC      (-0.18553 ; 0.00992)     ✔
AUC_inf  (-0.17914 ; 0.01620)     ✔
C_max    (-0.22379 ; 0.02154)     ✖

Method: TOST
alpha = 0.05; Equiv. lim. = +/- 0.22314
```

```
matost = ctost(theta = theta_hat, sigma = Sigma_hat, nu = nu, delta = log(1.25))
matost

✔ Accept (bio)equivalence
Equiv. Region:   |----------------0----------------|
t_half                (----------x----------)
AUC                 (-------x-------)
AUC_inf             (--------x-------)
C_max             (---------x---------)

CIs:
t_half   (-0.15011 ; 0.11747)     ✔
AUC      (-0.18031 ; 0.00469)     ✔
AUC_inf  (-0.17392 ; 0.01097)     ✔
C_max    (-0.21723 ; 0.01498)     ✔

Method: alpha-TOST
alpha = 0.05; Equiv. lim. = +/- 0.22314
Corrected alpha = 0.05908
```

References

[1] Stefan Wellek. *Testing Statistical Hypotheses of Equivalence and Noninferiority*. CRC Press, 2010.

[2] Donald J Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, 15(6):657–680, 1987.

[3] D. Labes and H. Schütz. Inflation of type I error in the evaluation of scaled average bioequivalence, and a method for its control. *Pharmaceutical Research*, 33: 2805–2814, 2016.

[4] R. L. Berger and J. C. Hsu. Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, 11:283–319, 1996.

[5] Julie Quartier, Ninon Capony, Maria Lapteva, and Yogeshvar N Kalia. Cutaneous biodistribution: a high-resolution methodology to assess bioequivalence in topical skin delivery. *Pharmaceutics*, 11(9):484, 2019.