

Sample Size for Agreement Studies on Quantitative Variables

Bruno Mario Cesana⁽¹⁾ , Paolo Antonelli⁽²⁾

(1) Associate Professor of Medical Statistics. Retired from the Section of Medical Statistics and Biometry, Department of Molecular and Translational Medicine, Faculty of Medicine and Surgery, University of Brescia, V. le Europa 11, 25123 Brescia, Italy and Department of Clinical Sciences and Community Health, Unit of Medical Statistics, Biometry and Bioinformatics “Giulio A. Maccacaro” Faculty of Medicine and Surgery, University of Milan, Via Celoria 22, 20133, Milan, Italy.

(2) Retired as a Professor of Calculus of Probabilities, Statistics and Operative Research at the State Industrial Technical Institute (ITIS) Benedetto Castelli, Brescia, Via Antonio Cantore, 9, 25128 Brescia, Italy.

CORRESPONDING AUTHOR: Bruno Mario Cesana Phone. +39 02503 - 20854 Fax 02503 – 20855. Email: brnmrcesana@gmail.com

SUMMARY

We reviewed the statistical assessments of the agreement between two measurement methods of continuous variables together with their recent contributions about the sample size calculation based on the “two one side t-tests (TOST) extensions to the individual equivalence. We generalized a restricted null hypothesis that constitutes a particular case in finding the supremum of the probability of rejecting the equivalence under the null hypothesis (H_0) and which, obviously, limits its applicability.

Particularly, we devise and propose an exact procedure for calculating the sample sizes for individual equivalence, as an expression of the agreement between two measurement methods, by using a size a test (that is, with adequate control of Type I error), based on the non-central bivariate t distribution with correlation equal to 1 and to the related functions for calculating a and 1-b probabilities.

Furthermore, our devised procedure allows to calculate the sample sizes by choosing between two most suitable formulations of the global parameters space of the null and alternative hypotheses; indeed, they are based on the portion of the distribution of the differences between the two measurement methods or on appropriately chosen agreement thresholds.

Thereafter, we compared our theoretical results with the recently published proposals of the sample size calculation for the Bland and Altman agreement analysis by means also of simulation studies.

Finally, a program written in the open-source R language to perform sample size calculations according to our procedure is available upon request.

Keywords: Measurement Methods comparison; Quantitative variables; Bland-Altman analysis; Sample size calculation; Individual equivalence.

INTRODUCTION

Let’s define the statistical model of an agreement study between two measurement methods of quantitative variables without replicates, according to the usual model of study carried out in clinical and, above all, laboratory settings. It has to be noted that in the context of an agreement study the two measurement methods under comparison are expected to be of equivalent precision. Particularly, there is an “Old measurement method” that can be defined as “Standard” (but without the connotation of a “Gold Standard”) and a “New measurement method” that can be defined as “Experimental” or “Test” (thereafter,

Test for simplicity). In addition, the Test method has some advantages over the current Standard (lower cost, greater simplicity of execution, for example), but to replace the “Old/ Standard measurement method” it must be proved to be “essentially equivalent” in the context of an agreement study.

Let’s assume that the variable X , Gaussian distributed $[X \sim G(\mu_X, \sigma_X^2)]$, represents the values of the Standard and that the variable Y , Gaussian distributed $[Y \sim G(\mu_Y, \sigma_Y^2)]$, represents the values of the Test.

Thus, a measurement value of the Standard can be expressed as: $X_i = \beta_1 \mu_i + \xi_1 + \varepsilon_{iX}$ and a measurement value of the Test can be expressed as: $Y_i = \beta_2 \mu_i + \xi_2 + \varepsilon_{iY}$. From the above definitions, it is

assumed that each value is obtained by the sum of the true value (μ_i with $i = 1, \dots, n$) of the i -th subject/laboratory sample multiplied by a fixed constant (b_1 or b_2), of the effect of the measurement method (ξ_j , with $j = 1, 2$ for the Standard and the Test, respectively), and of the measurement error ξ_{ij} , considered individually and independently distributed for the i -th subject/laboratory sample. This error component, is assumed (as a usual assumption) to be Gaussian and independently distributed: $[\varepsilon_X \sim G(0, \sigma_{\varepsilon_X}^2)]$ and $[\varepsilon_Y \sim G(0, \sigma_{\varepsilon_Y}^2)]$, respectively.

Furthermore, it has to be outlined that $b_1 \neq 1$ and $b_2 \neq 1$ cause the occurrence of a proportional error which can be of different magnitude for the two measurement methods under comparison. In addition, if $\xi_1 \neq 0$ and $\xi_2 \neq 0$, there will be a systematic error for the two measurement methods under comparison. Indeed, the paired differences between measurement methods in perfect agreement lie on the bisecting line of the first cartesian plane with intercept equal to 0 and slope equal to 1.

Otherwise, the systematic and proportional error between the two measurement methods will be considered when, in the agreement analysis the differences will be regressed on their means according to Bland and Altman [4,5]. Obviously, the difference between the population means (ξ_1 and ξ_2) of the two measurement methods can be the source of a possible "systematic error".

Taking into account the agreement between the two measurement methods under comparison, we focus on their paired sampling differences:

$$D_i = X_i - Y_i = \beta_1 \mu_i + \xi_1 + \varepsilon_{iX} - (\beta_2 \mu_i + \xi_2 + \varepsilon_{iY})$$

$$= \beta_1 \mu_i - \beta_2 \mu_i + \xi_1 - \xi_2 + \varepsilon_X - \varepsilon_Y = \xi_1 - \xi_2 + \varepsilon_X - \varepsilon_Y$$

These differences (D_i) are given by the difference between the values of the two measurement methods and of their measurement errors, since, under the agreement assumption, the multiplicative constant of the true values are assumed to be equal ($b_1 = b_2 = \beta$, even if they are not equal to 1) and the true values (μ_i), equal components of the two above reported expressions, are removed together with the biological variability of each subject/laboratory sample.

Finally, D_i , being the difference of two independent Gaussian distributed variables with measurement error, is Gaussian distributed with mean μ_D , and variance $\sigma_D^2 = \sigma_{\varepsilon_X}^2 + \sigma_{\varepsilon_Y}^2$, both unknown. However, under the agreement situation, μ_D is expected to be almost zero and $\sigma_D^2 \approx 2\sigma_{\varepsilon_X}^2$ or $\approx 2\sigma_{\varepsilon_Y}^2$, since it is expected that $\sigma_{\varepsilon_X}^2 \approx \sigma_{\varepsilon_Y}^2$ with correlation equal to zero. Thereafter, μ_D and σ_D will be indicated as μ and σ for simplicity.

It should be emphasized that agreement studies carried out in clinical settings are not always preceded by the pertinent assessments of accuracy, precision, reliability, repeatability and reproducibility of the

new (Test) measurement method as practically always happens in laboratory environments where there is a greater knowledge of the measurement aspects and a stricter adherence to the pertinent guidelines. Therefore, it is possible that a new clinical measurement method is compared with the current standard without having satisfactory metric properties.

Statistical analysis and Sample size for agreement studies

It is widely recognized that biomedical researches have to be adequately powered in order to have a satisfactorily high probability to achieve their (primary) objectives.

Of course, also agreement studies between two (or more) measurement methods need to be adequately powered, although there does not seem to be adequate attention to this aspect, as reported by a recent review by Han et al. [1]. Particularly, Han et al. [3] wrote that: "only 27 studies out of 82 (33%) gave justification for their sample size".

Furthermore, also Kottner et al. [4] and Gerke et al. [5] have pointed out that formal justifications for sample size were rarely reported in agreement studies.

We are interested in sample size calculations for agreement studies on quantitative variables carried out according to the Bland and Altman approach [1,2] (thereafter B&A, for simplicity).

Indeed, Han et al. [3] reported that agreement studies on continuous variables are very frequently carried out by calculating the limits of agreement (LoAs) according to B&A's method [1,2] that "enables us to separate systematic and random error, which r (the correlation coefficient: authors' note), combines into a single measure" as B&A wrote [6].

However, it has to be stressed that researchers often consider and use the B&A's procedure [1,2] simply as a "graphical approach" without paying due attention to its assumptions and conditions which have to be fulfilled in order to draw valid conclusions. Particularly, the Gaussian distribution of the differences, the equality of the variances of the two measurement methods, the absence of relevant systematic bias and, above all, the absence of relevant proportional bias, as Taffè [7] has recently reiterated. Furthermore, it has also to be said that, unfortunately, these assumptions are unlikely to hold in practice.

It has to be stressed that the sample size calculation for agreement studies has progressively been moved to the sample size for equivalence studies. Furthermore, from equivalence studies focussed on the equivalence of means (ABE: Average BioEquivalence), the recent methodological contributes deal with the Population BioEquivalence (PBE) aspect in which the agreement limits must include a relevant part of the central population of the differences between two measurement methods, and, finally with the individual bioequivalence (IBE) where the assessment is made on the differences between values

measured on the same unit and, consequently, paired. Particularly, the usual 95% limits of agreement (LoA) encompass the 2.5th percentile and 97.5th percentile of the distribution of the paired differences between the values recorded by the measurement methods on the same specimen/subject. Accordingly, sample sizes procedures have been proposed.

Therefore, we considered outdated the proposed sample sizes calculations based on the precision of the 95% Confidence Interval (CI) of the LoAs according to a generic recommendation given by Bland on his website [8]: “I usually recommend 100 as a good sample size, which gives a 95% CI about $\pm 0.34s$, being “s” the standard deviation of the differences between the measurements of the two methods. A sample of 200 subjects is even better, giving a 95%CI about $\pm 0.24s$. As with all estimation, to determine the appropriate sample size the researchers must decide what accuracy is required.”

Moreover, Bland’s suggestion [8] cannot be considered completely shareable owing to the fact that the required precision of the LoAs can be obtained only about the 50% of the cases as firstly highlighted by Cesana and Antonelli [9] and recognized also by Lu et al. [10].

Likewise, we do not consider the sample sizes calculations approaches by Shieh [11] and Shieh [12] with the exact centile calculation, shown by Carkeet [13] and by Carkeet and Goh [14], for having an adequate probability, defined as “assurance probability”, of obtaining 95%CI intervals width of the LoA less than a required width. Similarly, we do not consider the sample size calculation proposed by Jan and Shieh [15].

Bland-Altman’s method and its sample size calculation

We focus our attention on the sample size calculation proposed by Shieh [16] together with the TOST procedures for assessing the agreement proposed by Liu and Chow [17], Lin et al. [18], and Lu et al. [10], considered in Shieh’s paper [16], together with the B&A’s approach [1,2] based on the approximate confidence intervals of normal centiles.

Considering that “to establish the agreement between two methods, the central portion of the distribution of paired differences needs to be within a close range around zero”, Shieh [16] calculated the supremum of the Type I error (α) under a null hypothesis of no equivalence attained “when the two centiles coincide with the boundary values $\{\hat{\theta}_{1-p}; \hat{\theta}_p\} = \{-\Delta; \Delta\}$ ”. This sentence leads to the simultaneous equality of $\hat{\theta}_{1-p}$ with $-\Delta$ and $\hat{\theta}_p$ with Δ .

Rather curiously, also in the Krishnamoorthy and Mathew’s book [19] (on page 35, equation 2.3.11) there is a similar unproved sentence “note that the supremum in the above equation is attained at $L_l = \mu - z_{(1+p)/2}\sigma$ and $L_u = \mu + z_{(1+p)/2}\sigma$ ” where the

$(1+p)/2$ subscript corresponds to the population proportion, defined p in our notation. It has to be stressed that the affirmation of Krishnamoorthy and Mathew [19], followed by Shieh [16], is valid only in a particular subset of H_0 as we will demonstrate and, consequently, it has to be considered unsuitable because it is not general.

Moreover, consequently to his definition of the supremum under H_0 , Shieh [16] calculated the critical value $\gamma_{1-\alpha}$ such that the statistical test is of size α . Then, the statistical agreement test rejects the null hypothesis if $\gamma_{1-\alpha} < T_L$ and $T_U < -\gamma_{1-\alpha}$ where $T_L = (\bar{D} + \Delta) / \sqrt{(S^2 / n)}$ and $T_U = (\bar{D} - \Delta) / \sqrt{(S^2 / n)}$ where \bar{D} and S^2 are the sample mean and the variance of the differences, respectively, n is the sample size and Δ is the equivalence threshold.

Furthermore, consistently with his calculated critical value $\gamma_{1-\alpha}$, Shieh [16] claimed that the Type I errors of the considered TOST approaches by Liu and Chow [17], B&A [2], Lin et al. [18] and Lu et al. [10] turned out to be too much conservative.

Aim of the paper

Aim of our paper is to show the appropriate sample size calculation approach for agreement studies carried out according to the “individual equivalence” model. It has to be noted that the appropriate approach must consider all the possible pair $(\mu; \sigma)$ under H_0 and H_A and, consequently, it should not be limited to particular cases.

Indeed, we have found that Shieh’s proposal [16] is valid only in a particular context and, consequently we propose an exact general procedure for calculating the sample size for individual equivalence by using a size α test (that is, with adequate control of Type I error), as defined by Casella and Berger [20], based on the non-central bivariate t distribution, with correlation equal to 1 and Owen’s Q functions [21] for calculating a and $1-\beta$ probabilities.

Furthermore, we consider two different sample size calculation approaches: the first is based on the population proportion (p) or on its corresponding normal quantile (z_p), and the second is based on the agreement threshold (Δ).

Particularly, keeping fixed Δ , it is possible to hypothesize a population proportion under H_A greater than the population proportion under H_0 ($p_A > p_0$) or, equivalently, in the terms of the quantiles $z_{p_A} > z_{p_0}$. Vice versa, keeping fixed the population proportion (p) or the normal quantile z_p , it is possible to hypothesize, under H_A , a narrower interval $(-\Delta_A, \Delta_A)$ than under H_0 $(-\Delta_0, \Delta_0)$.

Furthermore, we compare our results with those coming from the recent proposal of Shieh [16] and in particular with those from Liu and Chow [17] procedure.

Finally, we clarify whether the TOST procedure applied to the individual equivalence is of size α or of level α , according to the definition of Casella and Berger [20].

The Methods section details the theoretical approach of the sample size calculation. Even if the more technical aspects have been moved to appropriate appendices so as not to interrupt the flow of presentation and of the reasoning, the following topics will be considered: (i) Null (H_0) and alternative (H_A) hypotheses together with their parameter space; (ii) Outlines of the proposed procedure; (iii) The probability of rejecting H_0 and its calculation. Then, the paragraph "Sample size calculation" shows: (A) the determination of the non-centrality parameters under H_0 and H_A ; (B) the alternative hypothesis according to two proposed different approaches (Case 1: model with fixed population proportion p_0 . Case 2: model with fixed agreement threshold Δ); and (C) the Sample size calculation procedure.

The Results section shows the "Tables of the sample size" for the main common and sensible scenarios of agreement studies, the "Comparisons between the sample sizes calculated with p fixed (Case 1) and Δ fixed (Case 2)" and "a particular approach: the sample size calculation under two simple hypotheses". Then, there are the results of simulation studies focussing on: (A) Check of the fulfilment of the nominal significance level α , and (B) Check of the fulfilment of the nominal power $(1 - \beta)$. A paragraph about the "comparisons among the different considered methods: our new AC procedure, Shieh [16], Liu and Chow [17], Bland and Altman [2], Lin et al. [18] and Lu et al. [10] and a paragraph "Check of the fulfilment of the nominal significance level α of the various methods" follow.

The two paragraphs "Considerations about the canonical null hypothesis H_0 and Shieh's approach [16]" and "Considerations about Liu and Chow's TOST procedure [17] and our AC procedure" precede the Discussion section that concludes the paper.

METHODS

Let's denote the 100p-th centile of the Gaussian distribution $G(\mu, \sigma^2)$ as $\theta_p = \mu + z_p \sigma$, where z_p is the 100p-th centile of the standard Gaussian distribution $G(0, 1)$; it has to be noted that we use the lowercase "p" notation to define the centile of a distribution.

Null (H_0) and Alternative (H_A) Hypotheses together with their Parameter Space

To establish the agreement between two measurement methods, a relevant central portion (defined as a capital "P": $P = 2p - 1$) of the distribution of the paired differences needs to be within a narrow interval (say, $-\Delta, \Delta$) around zero.

Thus, the statistical test is formulated as an equivalence test with the null (H_0) and the alternative (H_A) hypotheses given by:

$$H_0 : (\theta_{1-p} \leq -\Delta) \vee (\theta_p \geq \Delta)$$

$$\text{vs. } H_A : (-\Delta < \theta_{1-p}) \wedge (\theta_p < \Delta) \quad \text{Formula 1}$$

Otherwise, being $\theta_{1-p} = \mu - z_p \sigma$, $\theta_p = \mu + z_p \sigma$ with $p > 0.5$, $\mu \in R$ and $\sigma > 0$,

$$H_0 : (\mu - z_p \sigma \leq -\Delta) \vee (\mu + z_p \sigma \geq \Delta) \text{ vs.}$$

$$H_A : (-\Delta < \mu - z_p \sigma) \wedge (\mu + z_p \sigma < \Delta), \text{ respectively.}$$

The null (H_0) and the alternative (H_A) hypotheses depend on μ and σ , being fixed the other parameters (p and Δ). These hypotheses have an adequate geometrical representation in the R^2 space, by placing μ on the horizontal X axis and σ on the vertical Y axis as shown in Figure 1 (Panel A and Panel B).

The alternative hypothesis H_A can be reported as:

$$-\Delta < \mu < \Delta \quad \text{and,}$$

$$\sigma < (\Delta - \mu) / z_p \wedge \sigma < (\Delta + \mu) / z_p \wedge \sigma > 0$$

ultimately as:

$$H_A = \left\{ (\mu; \sigma) : \left(-\Delta < \mu < \Delta \right) \wedge \left[0 < \sigma < (\Delta - |\mu|) / z_p \right] \right\} \quad \text{Formula 2}$$

Thus, the H_A space corresponds to the points $(\mu; \sigma)$ inside the ABC triangle, shown in Figure 1 (Panel B), determined by the lines: $\sigma = 0$; $\sigma = (\Delta - \mu) / z_p$; $\sigma = (\Delta + \mu) / z_p$

The null hypothesis H_0 is the complementary set of H_A and, therefore, is simply the set of points $(\mu; \sigma)$ of the positive half-plane of the ordinates ($\sigma > 0$) not inside the triangle ABC.

Formally, H_0 is the union of two subspaces: $H_0 = H_{0A} \cup H_{0B}$, where:

$$H_{0A} = \left\{ (\mu; \sigma) : (\mu - z_p \sigma \leq -\Delta) \right\} = \left\{ (\mu; \sigma) : (-\infty < \mu < \infty) \wedge \sigma \geq (\Delta + \mu) / z_p > 0 \right\}$$

is the region on the left of the half-line on which the side AC lies, and

$$H_{0B} = \left\{ (\mu; \sigma) : (\mu + z_p \sigma \geq \Delta) \right\} = \left\{ (\mu; \sigma) : (-\infty < \mu < \infty) \wedge \sigma \geq (\Delta - \mu) / z_p > 0 \right\}$$

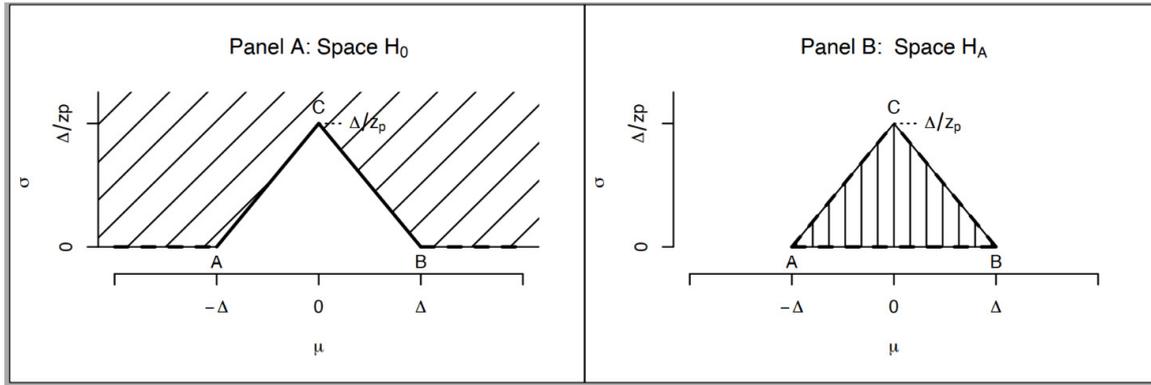
is the region on the right of the half-line on which the side BC lies.

Thereafter, H_0 can be briefly defined as:

$$H_0 = \left\{ (\mu; \sigma) : (-\infty < \mu < +\infty) \wedge \left[\sigma \geq (\Delta - |\mu|) / z_p > 0 \right] \right\} \quad \text{Formula 3}$$

Thus, the H_0 space is represented by the positive half-plane of the Y axis with the exclusion of the inner points of the ABC triangle, with vertices $A = (-\Delta; 0)$,

Figure 1. Space of the null hypothesis (H_0 , Panel A) and space of the alternative hypothesis (H_A , Panel B)



$B = (\Delta; 0)$, and $C = (0; \sigma = \Delta/z_p)$, shown in Figure 1 (Panel A).

It should be noted that the alternative hypothesis H_A specifies that there is at least $P = 2p - 1$ central portion of the distribution of the paired differences in the range $(-\Delta, \Delta)$, while the null hypothesis H_0 specifies that the central portion P included in the range $(-\Delta, \Delta)$, is less than $2p - 1$.

Outlines of the proposed procedure

Following Shieh [16], “a natural rejection region to assess”, at a significance level α , a central P portion $(\theta_{1-p}; \theta_p)$ of a Gaussian population of the differences $D \sim G(\mu; \sigma)$ with the p -centile “ p ” equal to: $p = (P + 1)/2$ is given by:

$$E = \{-\Delta < \hat{\theta}_{1-p} \wedge \hat{\theta}_p < \Delta\} \quad \text{Formula 4}$$

with $\hat{\theta}_{1-p} = \bar{D} - k_{1-\alpha} S / \sqrt{n}$ and $\hat{\theta}_p = \bar{D} + k_{1-\alpha} S / \sqrt{n}$, being $k_{1-\alpha}$ an opportune constant to be determined to control the Type I error probability; furthermore, “ \bar{D} ” and “ S ” are the sample mean and the sample standard deviation of the distribution of the differences, respectively.

Denoting with $\Pr\{E\}$ the probability of rejecting H_0 , we obtain: $\Pr\{E\} = \Pr\{(-\Delta < \hat{\theta}_{1-p}) \wedge (\hat{\theta}_p < \Delta)\}$

Our sample size calculation procedure is based on the two following main steps.

a. In order to keep the Type I error always less than or equal to the prefixed α for any pair of values $(\mu; \sigma)$ within the H_0 space, the $k_{1-\alpha}$ coefficient has to be determined so that the superior of the probability of the event E is equal to α : $\sup_{H_0} \Pr\{E\} = \alpha$. The superior is calculated in order that the Type I error is always $\leq \alpha$.

We will demonstrate that, under H_0 , the Type I error reaches its supremum value (α) in the two points $(\mu; \sigma) = [\mu; (\Delta - |\mu|)/z_p]$ for $\mu \rightarrow -\Delta^+$ and for

$\mu \rightarrow \Delta^-$ corresponding to the points A and B of the ABC triangle (Figure 1, Panel A), instead of the point $(\mu; \sigma) = (0; \Delta / z_p)$, claimed by Shieh [16], corresponding to the vertex C of the ABC triangle (Figure 1, Panel B).

b. The sample size “ n ” is calculated so that the $\inf_{H_A} \Pr\{E\} = 1 - \beta$. Thus, the power of the test is always not inferior to the prefixed threshold of $1 - \beta$ for any pair of values $(\mu; \sigma)$ in the H_A space. We will demonstrate that, under H_A , the power attains its lower extremum in the point $(\mu; \sigma) = (0; \Delta / z_p)$ corresponding to the vertex C of the ABC triangle (Figure 1, Panel B), in agreement with Shieh [16].

The sample size calculation is based on the non-central bivariate t distribution with correlation equal to 1 and Owen’s Q formulation, together with some related theorems [21].

The probability of rejecting H_0 : $\Pr\{E\}$

The event $E = \{-\Delta < \hat{\theta}_{1-p} \wedge \hat{\theta}_p < \Delta\}$ after some algebraic steps, can be written as:

$$E = \left\{ \begin{aligned} & \left(\frac{\bar{D} + \Delta}{S / \sqrt{n}} > k_{1-\alpha} \wedge \right. \\ & \left. \left(\frac{\bar{D} - \Delta}{S / \sqrt{n}} < -k_{1-\alpha} \right) \right\} \quad \text{Formula 5} \end{aligned} \right.$$

Owen [21] (page 437) wrote that the statistics:

$T_L = (\bar{D} + \Delta) / (S / \sqrt{n})$ and $T_U = (\bar{D} - \Delta) / (S / \sqrt{n})$ are distributed as non-central Student’s t distributions, with $v = n - 1$ degrees of freedom and non-centrality parameter τ given by: $\tau_L = (\mu + \Delta) / (\sigma / \sqrt{n})$ and $\tau_U = (\mu - \Delta) / (\sigma / \sqrt{n})$, respectively. Our demonstration of Owen’s affirmation [21] is given in the Appendix A.

Furthermore, if these statistics are jointly considered, they are distributed as a non-central bivariate t with correlation equal to 1, according to Owen [21] who refined the definition of a multivariate t -distribution given by Dunnett and Sobel [22].

Thus, the probability of the event E is:

$$\Pr\{E\} = \Pr\{(T_L > k_{1-\alpha}) \wedge (T_U < -k_{1-\alpha})\} = \Pr\{[t_{v,\tau_L} > k_{1-\alpha}] \wedge [t_{v,\tau_U} < -k_{1-\alpha}]\} \quad \text{Formula 6}$$

where $[t_{v,\tau_L}; t_{v,\tau_U}]$ is a non-central bivariate t-distribution with degrees of freedom v , non-centrality parameters τ_L and τ_U , and correlation equal to 1.

Particularly, $\Pr\{E\}$ can be calculated using statistical packages that implement the non-central bivariate t distribution such as Owen's package "OwenQ" [23] that implements also the well-known Owen's Q functions [21] or "PowerTOST" package [24] or using the package "mvtnorm" from Genz et al. [25] with a non-deterministic procedure.

From Owen's formulas [21] for calculating $\Pr\{E\}$, it is possible to note the following two useful properties of $\Pr\{E\}$:

- a. $\Pr\{E\}$ depends on the non-centrality parameters τ_L and τ_U . Particularly, $\Pr\{E\}$ increases when τ_L increases and when τ_U decreases, as it is possible to verify by calculating the area of the cumulative non-central bivariate t density function. For example, with $n = 134$ and $k_{1-\alpha} = 17$, when $\tau_L = 15$ and $\tau_U = -15$, $\Pr\{E\} = 0.00752$; moreover, when $\tau_L = 21$ and $\tau_U = -15$, $\Pr\{E\} = 0.0857$. Finally, when $\tau_L = 21$ and $\tau_U = -21$, $\Pr\{E\} = 0.99437$.
- b. $\Pr\{E | \tau_L, \tau_U\} = \Pr\{E | -\tau_U, -\tau_L\}$. That is, $\Pr\{E\}$ value does not change even when $\tau_L = -\tau_U$ and $\tau_U = -\tau_L$.

Calculation of $\sup_{H_0} \Pr\{E\}$

Theorem: $\sup_{H_0} \Pr\{E\}$, under H_0 , is attained as a limit when the generic point $(\mu; \sigma)$ tends to A $(-\Delta; 0)$ on the AC side or tends to B $(\Delta; 0)$ on the BC side of the ABC triangle shown in Figure 1 (Panel A).

Under H_0 , the conditions are: $(-\infty < \mu < +\infty)$ and $\sigma \geq (\Delta - |\mu|) / z_p > 0$.

It is well evident that the critical points lie on the boundary of the H_0 space; particularly, on the sides AC $(\mu_0 - z_p \sigma_0 = -\Delta)$ or CB $(\mu_0 + z_p \sigma_0 = \Delta)$ of the ABC triangle excluded its A and B points since, obviously, $\sigma > 0$.

Considering the points on the side AC, the conditions are:

$$-\Delta < \mu \leq 0 \text{ and } \sigma = (\Delta - |\mu|) / z_p = (\Delta + \mu) / z_p.$$

Consequently, the two non-centrality parameters become:

$$\begin{aligned} \tau_L &= \frac{\mu + \Delta}{\sigma / \sqrt{n}} = \frac{\mu + \Delta}{\Delta - |\mu|} \cdot z_p \sqrt{n} \\ &= \frac{\mu + \Delta}{\Delta + \mu} \cdot z_p \sqrt{n} = z_p \sqrt{n} \text{ and} \\ \tau_U &= \frac{\mu - \Delta}{\sigma / \sqrt{n}} = \frac{\mu - \Delta}{\Delta - |\mu|} \cdot z_p \sqrt{n} = \frac{\mu - \Delta}{\Delta + \mu} \cdot z_p \sqrt{n} \end{aligned}$$

Therefore, on the side AC of the ABC triangle, τ_L is a constant, while τ_U is a branch of a hyperbola; particularly, we obtain the following limits:

$$\lim_{\mu \rightarrow -\Delta^+} \tau_L = z_p \sqrt{n} \text{ and } \lim_{\mu \rightarrow -\Delta^+} \tau_U = -\infty.$$

Furthermore, for the side BC of the ABC triangle the conditions are:

$$0 \leq \mu < \Delta \text{ and } \sigma = (\Delta - |\mu|) / z_p = (\Delta - \mu) / z_p$$

Consequently, the two non-centrality parameters become:

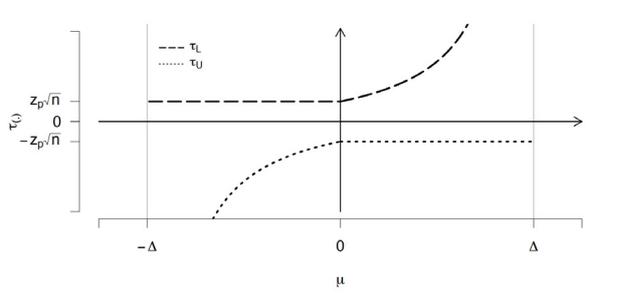
$$\begin{aligned} \tau_L &= \frac{\mu + \Delta}{\sigma / \sqrt{n}} = \frac{\mu + \Delta}{\Delta - \mu} \cdot z_p \sqrt{n} \text{ and} \\ \tau_U &= \frac{\mu - \Delta}{\sigma / \sqrt{n}} = \frac{\mu - \Delta}{\Delta - \mu} \cdot z_p \sqrt{n} = -z_p \sqrt{n} \end{aligned}$$

Therefore, on the side BC of the ABC triangle, τ_L is a branch of a hyperbola, while τ_U is a constant; particularly, we obtain the following limits:

$$\lim_{\mu \rightarrow \Delta^-} \tau_L = +\infty \text{ and } \lim_{\mu \rightarrow \Delta^-} \tau_U = -z_p \sqrt{n}.$$

Figure 2 shows the graphs of τ_L (long-dashed line) and τ_U (dashed line) functions against μ . It is possible to see that the two functions reach their vertical asymptotes $(+\infty; -\infty)$ when μ tends to $+\Delta$ or to $-\Delta$, respectively.

Figure 2. Graphs of the τ_L (long-dashed line) and τ_U (dashed line) functions against μ



By recalling the properties of $\Pr\{E\}$, $\sup_{H_0} \Pr\{E\}$ is obtained when $\mu \rightarrow +\Delta^-$ and $\sigma = \lim_{\mu \rightarrow \Delta^-} (\Delta - |\mu|) / z_p = 0^+$, since τ_L increases to $+\infty$ and τ_U is constant or when $\mu \rightarrow -\Delta^+$ and $\sigma = \lim_{\mu \rightarrow -\Delta^+} (\Delta - |\mu|) / z_p = 0^+$, since τ_L is constant and

τ_U decreases to $-\infty$. This occurs on the boundary of the H_0 space at the points B and A of the ABC triangle shown in Figure 1 (Panel A).

Thus, the superior under H_0 of the $\Pr\{E\}$ is given by:

$$\begin{aligned} \sup_{H_0} \Pr\{E\} &= \sup_{H_0} \Pr\left\{ \begin{array}{l} (T_L > k_{1-\alpha}) \wedge (T_U < -k_{1-\alpha}) \\ |\tau_L = +\infty; \tau_U = -z_p \sqrt{n} \end{array} \right\} = \text{Formula 7} \\ &= \Pr\left\{ \begin{array}{l} (t_{n-1, \tau_L} > k_{1-\alpha}) \wedge (t_{n-1, \tau_U} < -k_{1-\alpha}) \\ |\tau_L = +\infty; \tau_U = -z_p \sqrt{n} \end{array} \right\} \end{aligned}$$

at the point B. It has to be noted that $[t_{v, \tau_L}; t_{v, \tau_U}]$ is a non-central bivariate t with correlation equal to 1.

In the case of the point A of the ABC triangle, the conditioning is given by: $\tau_L = z_p \sqrt{n}$; $\tau_U = -\infty$.

It is worthwhile to stress that the superior of the $\Pr\{E\}$ under H_0 is equal in the points A and B of the ABC triangle, according to the previously reported property b of $\Pr\{E\}$.

This conclusion disagrees with Shieh's affirmation [16] that the upper extremum is obtained when $(\theta_{1-p} = -\Delta) \wedge (\theta_p = \Delta)$, or, equivalently, when $(\mu = 0) \wedge (\sigma = \Delta / z_p)$ at the coordinates of the vertex C of the ABC triangle.

We will demonstrate in the following that Shieh's affirmation [16] is valid only under a condition more restrictive than that foreseen by the canonical H_0 .

According to a geometrical approach, it has to be noted that starting from the vertex C ($\mu = 0$ and $\sigma = \Delta / z_p$) of the ABC triangle, and going down on the side AC the probability $\Pr\{E\}$ increases owing to the fact that τ_L is constant and τ_U decreases tending to $-\infty$. Furthermore, going down on the side BC, $\Pr\{E\}$ increases since τ_L increases tending to $+\infty$ and τ_U is constant. So, the point C is not the place of the supremum of $\Pr\{E\}$ as Shieh affirmed [16].

In fact, the point C of the ABC triangle has coordinates: $\tau_L = z_p \sqrt{n}$ and $\tau_U = -z_p \sqrt{n}$ very far from the points A or B where $\Pr\{E\}$ reaches its supremum value (Figure 1, Panel A).

Consequently, we have appropriately modified Shieh's [16] sample size calculation and we will give the correct formulation for the general canonical case.

Calculation of $\inf_{H_A} \Pr\{E\}$

Theorem: $\inf_{H_A} \Pr\{E\}$, under $H_{A'}$, corresponds to the point C ($\mu = 0$, $\sigma = \Delta / z_p$).

Under $H_{A'}$, the conditions are: $-\Delta < \mu < \Delta$ and $0 < \sigma < (\Delta - |\mu|) / z_p$ corresponding to the inner points of the ABC triangle.

Also in this case, the points at which $\Pr\{E\}$ attains its infimum are located on the edge of the H_A space,

particularly on the sides AC and BC of the ABC triangle.

Considering the points on the side AC and BC, we have: $-\Delta < \mu < \Delta$ and $\sigma = (\Delta - |\mu|) / z_p$; thus, the two non-centrality parameters are:

$$\begin{aligned} \tau_L &= (\mu + \Delta) / (\sigma / \sqrt{n}) \\ &= (\mu + \Delta) / (\Delta - |\mu|) / z_p \sqrt{n} \end{aligned}$$

and

$$\begin{aligned} \tau_U &= (\mu - \Delta) / (\sigma / \sqrt{n}) \\ &= (\mu - \Delta) / (\Delta - |\mu|) / z_p \sqrt{n}. \end{aligned}$$

From the graphs of the non-centrality parameters (τ_L and τ_U) functions shown in Figure 2, it is possible to see that, under H_A when $\mu = 0$ and, consequently, $\sigma = (\Delta - |\mu|) / z_p$, τ_L reaches its minimum value equal to $z_p \sqrt{n}$ and τ_U attains its maximum value of $-z_p \sqrt{n}$.

Focussing our interest on the minimum value, we remember that $\Pr\{E\}$ decreases when τ_L decreases and τ_U increases; therefore, $\inf_{H_A} \Pr\{E\}$ occurs just in the border point $C = (0; z_p \sqrt{n})$, vertex of the ABC triangle of the H_A space, shown in Figure 1 (Panel B).

Therefore, the inferior of the power under H_A is given by:

$$\begin{aligned} \inf_{H_A} \Pr\{E\} &= \inf_{H_A} \Pr\left\{ \begin{array}{l} (T_L > k_{1-\alpha}) \wedge (T_U < -k_{1-\alpha}) \\ |\tau_L = z_p \sqrt{n}; \tau_U = -z_p \sqrt{n} \end{array} \right\} = \text{Formula 8} \\ &= \Pr\left\{ \begin{array}{l} (t_{n-1, \tau_L} > k_{1-\alpha}) \wedge (t_{n-1, \tau_U} < -k_{1-\alpha}) \\ |\tau_L = z_p \sqrt{n}; \tau_U = -z_p \sqrt{n} \end{array} \right\} \end{aligned}$$

A more immediate alternative proof is based on the fact that $\inf_{H_A} \Pr\{E\}$ has to be calculated on the boundary of $H_{A'}$; therefore, since both conditions must be satisfied, it must be:

$$\theta_{1-p} = -\Delta \text{ and } \theta_p = \Delta.$$

Remembering the definitions of θ_{1-p} and θ_p , this formulation corresponds to a linear system of two equations in the unknowns μ and σ , whose solution is: $\mu = 0$ and $\sigma = \Delta / z_p$.

This conclusion is in agreement with Shieh's affirmation [16] that the power of the test in the point with coordinates $(\mu = 0; \sigma = \Delta / z_p)$ is not inferior to a prefixed value $1 - \beta$, considering fixed the remaining parameters of the H_A space.

Finally, it has to be noted that $\sup_{H_0} \Pr\{E\}$ corresponds to the maximum significance level (usually 0.05) and that $\inf_{H_A} \Pr\{E\}$ corresponds to the minimum required power value (usually 0.8).

Sample size calculation

It has to be remembered that the null hypothesis (H_0) is a hypothesis of non-equivalence (non-agreement) being the thresholds ($\pm\Delta$) delimiting the interval considered of equivalence or "of practical equality"; therefore, values greater than $-\Delta_0$ or lower than Δ_0 allow to reject the non-equivalence H_0 hypothesis.

Otherwise, the non-equivalence H_0 hypothesis can be formulated in the terms of a population proportion; accordingly, a population proportion p_0 (or, equivalently, of a central population portion $P_0 = 2 p_0 - 1$) corresponds to the threshold of the non-equivalence and values greater than p_0 allow to reject the non-equivalence H_0 hypothesis.

It has to be stressed that p_0 or Δ_0 have to be appropriately chosen according to clinical/laboratory considerations based on literature findings.

Furthermore, it is well-known that, in the sample size calculation, a specific alternative hypothesis (a sub-space of H_A) has to be settled in the H_A space delimited by the ABC triangle which characterizes our case of interest (Figure 1, Panel B).

A) Setting the alternative hypothesis according to the two proposed different approaches

Case 1: population proportion p or quantile z_p fixed (different Δ thresholds)

The population proportions are kept fixed ($p_A = p_0 = p$) and lower Δ_A thresholds are selected ($\Delta_A < \Delta_0$), as it is shown in Figure 3 (Panel A).

Accordingly, the alternative hypothesis becomes:

$$H_A : (-\Delta_A < \theta_{1-p}) \wedge (\theta_p < \Delta_A).$$

Case 2: thresholds Δ fixed (different population proportions $p_A > p_0$)

The thresholds are kept fixed $\Delta_0 = \Delta_A = \Delta$ and a greater p_A population proportion (or quantile z_{pA}) is selected: namely, $p_A > p_0$ (or $z_{pA} > z_{p0}$).

The alternative hypothesis becomes:

$H_A : (-\Delta < \theta_{1-p_A}) \wedge (\theta_{p_A} < \Delta)$; its space is the hatched area with vertical lines delimited by the ABC* triangle shown in Figure 3 (Panel A and Panel B).

It has to be noted that Case 2 corresponds to the Shieh's approach [16, page 5], based on two population proportion values under H_0 and H_A , implicitly considering Δ as fixed.

For both the above considered Cases (Case 1 and Case 2), the sample size is calculated accordingly to the fulfilment of the conditions for controlling the Type I and Type II errors.

B)-Determination of the non-centrality parameters

A fundamental point is the fact that the sample size calculation depends on the calculation of $\Pr\{E\}$ that, in its turn, depends on the non-centrality parameters of the bivariate t distribution.

B.1)-The non-centrality parameters under H_0

The non-centrality parameters have to be calculated as a limit along a determined direction at the point of the supremum of $\Pr\{E\}$, that is at the point $B = (\Delta; 0)$, or, indifferently, at the point $A = (-\Delta; 0)$ of the ABC triangle, as we have shown in the previous paragraph "Calculation of $\sup_{H_0} \Pr\{E\}$ ".

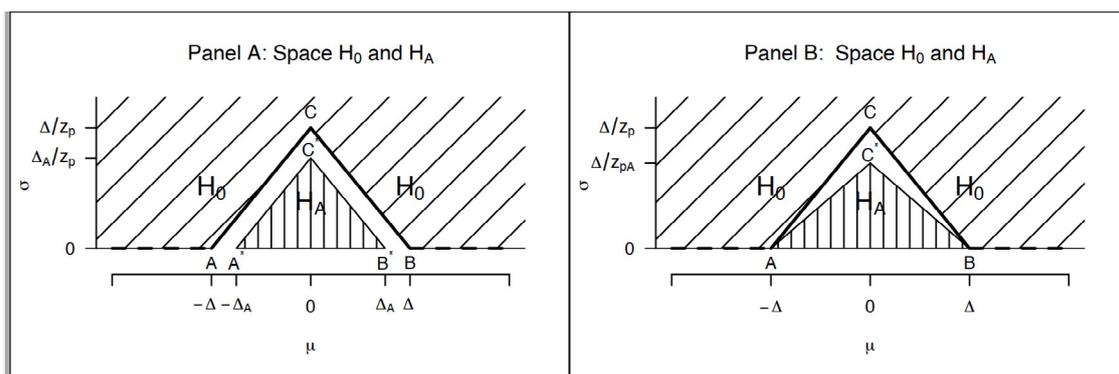
Thus, $\tau_{L_{H_0}} = +\infty$ and $\tau_{U_{H_0}} = -z_p \sqrt{n}$ at the point B or $\tau_{L_{H_0}} = z_p \sqrt{n}$ and $\tau_{U_{H_0}} = -\infty$ at the point A of the ABC triangle.

B.2)-The non-centrality parameters under H_A

The non-centrality parameters have to be calculated at the point of the infimum of $\Pr\{E\}$, that is, as already reported, at the vertex C^* of the space H_A (inner points of the ABC triangle).

Since this vertex is different in the two above

Figure 3. Panel A: Case 1: population proportion p or quantile z_p fixed (different Δ thresholds); Panel B: Case 2: thresholds Δ fixed (different population proportions $p_A > p_0$)



outlined cases, the non-centrality parameters will consequently be different.

Case 1: model with fixed population proportion p_0

In this case, $p_0 = p_A = p$ and $C^* = (\mu_{C^*}; \sigma_{C^*}) = (0; \Delta_A / z_p)$.

The non-centrality parameters become:

$$\tau_{L_{H_A}}(C^*) = (\mu_{C^*} + \Delta) / (\sigma_{C^*} / \sqrt{n}) = \Delta / (\Delta_A / z_p) \sqrt{n} = \Delta / \Delta_A \cdot z_p \sqrt{n}$$

$$\tau_{U_{H_A}}(C^*) = (\mu_{C^*} - \Delta) / (\sigma_{C^*} / \sqrt{n}) = -\Delta / (\Delta_A / z_p) \sqrt{n} = -\Delta / \Delta_A \cdot z_p \sqrt{n}$$

It should be noted that the above non-centrality parameters depend only on p , Δ/Δ_A , and n .

Case 2: model with fixed Δ

In this case, $\Delta_0 = \Delta_A = \Delta$, and $C^* = (\mu_{C^*}; \sigma_{C^*}) = (0; \Delta / z_{p_A})$

The non-centrality parameters become:

$$\tau_{L_{H_A}}(C^*) = (\mu_{C^*} + \Delta) / (\sigma_{C^*} / \sqrt{n}) = \Delta / (\Delta / z_{p_A}) \sqrt{n} = \Delta / \Delta_A \cdot z_{p_A} \sqrt{n}$$

$$\tau_{U_{H_A}}(C^*) = (\mu_{C^*} - \Delta) / (\sigma_{C^*} / \sqrt{n}) = -\Delta / (\Delta / z_{p_A}) \sqrt{n} = -z_{p_A} \sqrt{n}$$

It should be noted that the above non-centrality parameters depend only on p_A and n .

C)-Sample size calculation procedure

The sample size (n) is calculated according to an iterative procedure by using the non-central bivariate t distribution with correlation equal to 1. We have to obtain a “ n ” value that satisfies the two conditions of $\sup_{H_0} \Pr \{E\} = \alpha$ and $\inf_{H_A} \Pr \{E\} = 1 - \beta$.

Particularly:

- a. A starting very low n value ($n=2$) is iteratively increased until the calculated $k_{1-\alpha}$ gives:

$$\sup_{H_0} \Pr \{E\} = \Pr \left\{ \begin{array}{l} \left(t_{v, \tau_{H_0}} > k_{1-\alpha} \right) \wedge \\ \Pr \left(t_{v, \tau_{U_{H_0}}} < -k_{1-\alpha} \right) \end{array} \right\} = \alpha$$

- b. Then, we calculate the power:

$$\inf_{H_A} \Pr \{E\} = \Pr \left\{ \begin{array}{l} \left(t_{v, \tau_{H_A}} > k_{1-\alpha} \right) \wedge \\ \Pr \left(t_{v, \tau_{U_{H_A}}} < -k_{1-\alpha} \right) \end{array} \right\}$$

- c. Furthermore, we increase (or decrease) a low(high) “ n ” value and we repeat the steps a) and b) until the power reaches the $1 - \beta$ required threshold with $v = n - 1$, $\tau_{L_{H_0}}$, $\tau_{U_{H_0}}$, $\tau_{L_{H_A}}$, and $\tau_{U_{H_A}}$ appropriately calculated for the Case 1 or for the Case 2.

Furthermore, it is possible to make more efficient the procedure at the level a) since, as demonstrated in the Appendix B, the $\sup_{H_0} \Pr \{E\}$ can be calculated by using a non-central univariate Student’s distribution with v degrees of freedom and non-centrality parameter $\tau = z_p \sqrt{n}$. Then, our $k_{1-\alpha}$ calculated from the bivariate t distribution corresponds to the $100(1-\alpha)$ th percentile of the non-central univariate Student’s t distribution.

RESULTS

Tables of the sample size

As was demonstrated in the previous paragraph, the sample size calculation, fixed the Type I and Type II errors (α , β), does not depend on μ_0 , σ_0 , μ_A , and σ_A , but only on p_0 and the ratio Δ_A/Δ_0 in the Case 1 and on p_0 and p_A in the Case 2.

Therefore, the sample size tables have been built by considering two values of the significance level ($\alpha = 0.05$ or $\alpha = 0.01$) for two power values ($1-\beta = 0.80$ or 0.90). Then, for the Case 1 (Tables 1.1 and 1.2) some selected values of the parameters described above are: $p_0 = \{0.800, 0.900, 0.950, 0.975, \text{ and } 0.990$ as an extreme value for the non-equivalence} and $\Delta_A/\Delta_0 = \{0.80, 0.70, 0.65, 0.60, \text{ and } 0.55\}$. Furthermore, for the Case 2 (Tables 2.1 and 2.2), the selected values are: $p_0 = \{0.800, 0.900, 0.950, 0.975, \text{ and } 0.990\}$ and $p_A = \{0.80, 0.90, 0.95, 0.975, \text{ and } 0.99\}$.

For example, at a fixed $p_0 = 0.9$, a null hypothesis of $\Delta_0 = 1$ against $\Delta_A = 0.8$ ($\Delta_A/\Delta_0 = 0.8$) is rejected at a significance level of 0.05 and at a power of 0.8 with the specimens obtained from 169 units (Table 1.1, left).

The sample size becomes 214 if the power is increased to 0.90 (Table 1.2, right). Of course, at fixed values of p_0 , the sample size decreases at decreasing values of the ratio Δ_A/Δ_0 . Finally, the value of $\Delta_0 = 1$ has been considered for simplicity, but any value is possible as long as the value of the ratio is maintained.

For example, $H_0: p_0 = 0.9$ vs. $H_A: p_A = 0.95$, is rejected at a significance level of 0.05 and at a power of 0.8 with the specimens obtained from 134 units (Table 2.1, left).

The sample size becomes 169 if the power is increased to 0.90 (Table 2.2, right). Of course, for this approach the values of p_0 and p_A have to be specified.

Table 1.1. Sample size with fixed p (Case 1): $\alpha = 0.05$

Power	0.80					0.90				
Δ_A/Δ_0	0.80	0.70	0.65	0.60	0.55	0.80	0.70	0.65	0.60	0.55
$p_0=0.800$	282	102	67	46	32	355	127	83	57	40
0.900	169	63	42	30	21	214	79	53	37	27
0.950	134	51	35	25	18	171	65	44	31	22
0.975	118	45	31	22	17	151	58	39	28	21
0.990	106	41	29	21	15	136	53	36	26	19

Table 1.2. Sample size with fixed p (Case 1): $\alpha = 0.01$

Power	0.8					0.9				
Δ_A/Δ_0	0.80	0.70	0.65	0.60	0.55	0.80	0.70	0.65	0.60	0.55
$p_0=0.800$	215	77	51	35	24	279	100	65	44	31
0.900	128	48	32	22	16	168	62	41	29	21
0.950	101	38	26	19	14	134	50	34	24	17
0.975	88	34	23	17	12	117	45	30	22	16
0.990	79	31	21	16	12	106	41	28	20	15

Table 2.1. Sample size with fixed Δ (Case 2): $\alpha = 0.05$

Power	0.80					0.90				
p_A	0.800	0.900	0.950	0.975	0.990	0.800	0.900	0.950	0.975	0.990
$p_0=0.800$	*	71	25	15	10	*	88	31	18	12
0.900		*	134	44	22		*	169	55	27
0.950			*	220	54			*	282	69
0.975				*	201				*	259
0.990					*					*

*indicates that the sample size $\rightarrow \infty$

Table 2.2. Sample size with fixed Δ (Case 2): $\alpha = 0.01$

Power	0.80					0.90				
p_A	0.800	0.900	0.950	0.975	0.990	0.800	0.900	0.950	0.975	0.990
$p_0=0.800$	*	54	19	11	8	*	69	24	14	9
0.900		*	101	33	16		*	132	43	21
0.950			*	166	41			*	220	53
0.975				*	152				*	202
0.990					*					*

*indicates that the sample size $\rightarrow \infty$

Comparisons between the sample sizes calculated with p fixed (Case 1) and Δ fixed (Case 2)

The power $(\inf_{H_A} \Pr\{E\})$ increases when τ_L increases and τ_U decreases, keeping all other parameters (α , Δ , and n) fixed. Therefore, to compare the sample sizes calculated for the two considered cases, it is sufficient to compare the corresponding non-centrality parameters, under H_A , indicated shortly as $\tau_{(j)}(1)$ and $\tau_{(j)}(2)$, for the Case 1 and the Case 2, respectively.

The power of the Case 1 is greater or equal to the power of Case 2 when: $\tau_L(1) \geq \tau_L(2)$ and $\tau_U(1) \leq \tau_U(2)$, that is $(\Delta/\Delta_A) \cdot z_p \sqrt{n} \geq z_{p_A} \sqrt{n}$, or, equivalently, $\Delta/z_{p_A} \geq \Delta_A/z_p$. Consequently, the sample size of Case 1 is less than or at most equal to Case 2.

In geometrical terms, these quantities correspond to the ordinates of the respective points C^* under H_A in the Figure 3 (Panel A and Panel B, respectively) equal to $Y_{C^*}(1) = \Delta_A / z_p$; $Y_{C^*}(2) = \Delta / z_{p_A}$.

Therefore, the sample size of Case 1 is less than or at most equal to the sample size of the Case 2 iff $Y_{C^*}(1) \leq Y_{C^*}(2)$. In other terms, it is sufficient to compare the two Y_{C^*} values: the lower Y_{C^*} corresponds to the greater power and the smaller sample size. Finally, the two sample sizes will be equal when $\Delta_A/z_p = \Delta/z_{p_A}$.

We show a numerical example with $\alpha = 0.05$ and power = 0.80.

Case 2 with: $p_0 = 0.80$ ($z_{p_0} = 0.8416$), $\Delta_0 = 1$, and $p_A = 0.90$ ($z_{p_A} = 1.2816$), we have: $\Delta_0 / z_{p_A} = 0.7803$ and $n = 71$;

Case 1 with: $p_0 = 0.80$ ($z_{p_0} = 0.8416$), $\Delta_0 = 1$, and $\Delta_A = 0.70$, we have: $z_{p_A} = 1.2023 \rightarrow p_A = 0.8854$, $\Delta_0 / z_{p_0} = 1.1882$ and $n = 102$;

Case 1 with: $p_0 = 0.80$ ($z_{p_0} = 0.8416$), $\Delta_0 = 1$, and $\Delta_A = 0.65$, we have:

$$z_{p_A} = 1.2023 \rightarrow p_A = 0.8854, \\ \Delta_0 / z_{p_0} = 1.1882, \text{ and } n = 67;$$

Then, for obtaining under the Case 1 the same sample size calculated under the Case 2 it is sufficient to determine the value of Δ_A from the equation: $\Delta_A/z_{p_A} = \Delta_0/z_{p_0}$.

Consequently, we have for the Case 2, $p_0 = 0.80$ ($z_{p_0} = 0.8416$), $\Delta_0 = 1$, and $p_A = 0.90$ ($z_{p_A} = 1.2816$)

giving $\Delta_A = (\Delta_0 / z_{p_A}) \cdot z_{p_0} = (1/1.2816) \cdot 0.8416 = 0.6567$. Therefore, $\Delta_A / z_{p_0} = 0.7803$ and $n = 71$.

A particular approach: the sample size calculation under two simple hypotheses

It has to be pointed out that our new AC procedure

considers complex hypotheses on the full H_0 and H_A spaces. However, taking into account that $\Pr\{E\}$, being fixed the other parameters (α , $1-\beta$, Δ , and n), is function of the pair $(\mu; \sigma)$, it is possible to fix a pair $(\mu; \sigma)$ under H_0 and a pair under H_A and calculate $\Pr\{E\}$ under the two simple hypotheses, together with their corresponding sample size. This sample size calculation, particularly useful for a further verification of the theoretical results previously outlined, differs from the one outlined in the previous paragraph "Sample size calculation procedure" because $\Pr\{E|H_0\}$ and $\Pr\{E|H_A\}$ are now calculated at the fixed pair $(\mu_0; \sigma_0)$ and $(\mu_A; \sigma_A)$ without searching for the supremum and the infimum probability values, respectively.

Consequently, the non-centrality parameters under H_0 are:

$$\tau_{L_{H_0}} = (\mu_0 + \Delta / z_{p_0}) / (\sigma_0 / \sqrt{n}) \text{ and} \\ \tau_{U_{H_0}} = (\mu_0 - \Delta / z_{p_0}) / (\sigma_0 / \sqrt{n})$$

and, under H_A , are:

$$\tau_{L_{H_A}} = (\mu_A + \Delta / z_{p_A}) / (\sigma_A / \sqrt{n}) \text{ and} \\ \tau_{U_{H_A}} = (\mu_A - \Delta / z_{p_A}) / (\sigma_A / \sqrt{n})$$

It is possible to calculate these non-centrality parameters directly when are known the points $(\mu_0; \sigma_0)$ and $(\mu_A; \sigma_A)$ and a starting n value. Then, the pertinent sample sizes for comparing two simple hypotheses can be iteratively calculated.

Choosing the points $(\mu_0; \sigma_0)$ and $(\mu_A; \sigma_A)$ appropriately, it is possible to obtain the sample sizes calculated with our new AC procedure and those calculated by Shieh [16] by suitably choosing the two simple hypotheses.

Let's make an example according to our procedure with the following parameters values: $\alpha = 0.05$, $1 - \beta = 0.80$, $\Delta = 1$, $p_0 = 0.90$, and $p_A = 0.95$. If we use our procedure, we obtain a sample size of 134 (Table 2.1). The same sample size of 134 is obtained if we formulate as simple hypothesis H_0 the coordinates of the point A, namely: $H_0: (\mu_0; \sigma_0) = (-\Delta^+; 0^+) = \lim_{\mu \rightarrow -\Delta^+} (\mu; (\Delta - |\mu|) / z_{p_0})$ and as simple hypothesis H_A the coordinates of the point C^* , namely: $H_A: (\mu_A; \sigma_A) = (0; \Delta / z_{p_A})$.

The same result is obtained if we consider the coordinates of the point B, under $H_0: (\mu_0; \sigma_0) = (\Delta; 0^+) = \lim_{\mu \rightarrow \Delta^-} (\mu; (\Delta - |\mu|) / z_{p_0})$.

In addition, if as simple hypotheses H_0 and H_A , we consider the coordinates of the points $C = (0; \Delta / z_{p_0})$ and $C = (0; \Delta / z_{p_A})$ respectively, according to Shieh [16], the sample size becomes 62, equal to the value calculated by Shieh's procedure [16].

Simulation studies

These simulation studies have been carried out for having an experimental confirmation of our theoretical conclusions.

A) Check of the fulfilment of the nominal significance level α

It must verify that $\sup_{H_0} \Pr \{E\} = \alpha$.

The fixed parameters of this simulation study are: $\alpha = 0.05$, $\Delta = 1$, $p = 0.90$, with $p = (P+1)/2$, where P is the population central proportion under H_0 , that Shieh [16] calls "Null proportion"; otherwise, the parameter μ_0 has been made to vary within the interval $-\Delta, \Delta$ (precisely, for selected values from -0.99 to 0 and from 0 to 0.99) and, consequently, $\sigma_0 = (\Delta - |\mu|) / z_p$.

For each of the seven pair of $(\mu_0; \sigma_0)$ under H_0 , 10,000 samples of size $n = 50$ have been simulated and the proportion of H_0 rejection (corresponding to the Type I error and defined by Shieh [16] "simulated proportion") has been calculated according to Shieh's procedure [16] and to our procedure. The results are shown in Table 3.

It is possible to see that for $\mu_0 = 0$ the results are in agreement with Shieh's affirmation [16] that his procedure controls adequately the Type I error. Indeed, this proportion, calculated according to Shieh's procedure [16], is very near to the nominal significance level of $\alpha = 0.05$, while the Type I error proportion calculated according to our new AC procedure is much lower. However, when μ_0 differs slightly from 0 , the Type I error proportion of Shieh's procedure [16] increases to values much greater than the nominal significance level of 0.05 until a maximum of 0.2222 or 0.2176 (Table 3) for μ_0 values of -0.99 or 0.99 , very near to the value of $\Delta = -1$ or $\Delta = +1$ corresponding to the points A or B of the ABC triangle.

Otherwise, the significance level α from our procedure reaches the nominal value and it remains in the required validity intervals $[0.04, 0.06]$ and

$[0.0450, 0.0550]$ proposed by Cochran [26] and Bradley [27], respectively.

So, the simulation results confirm, as we expected, the theoretically obtained results. In conclusion, only the sample size calculated according to our procedure allows to respect the nominal significance level α .

B) Check of the fulfilment of the nominal power $(1 - \beta)$

It needs to be verified that $\inf_{H_A} \Pr \{E\} = 1 - \beta$.

We limited ourselves in checking the actual test power under H_A only in Case 2 with fixed Δ .

We considered the same scenarios shown by Shieh [16] (Table 5, page 5): that is, nominal power = 0.80 or 0.90 , population central proportion (P_0) under H_0 equal to $0.80, 0.90$, and 0.95 and, consequently, $p_0 = (1 + P_0)/2$, central population proportion (P_A) under H_A equal to $0.90, 0.95$, and 0.99 and, consequently, $p_A = (1 + P_A)/2$, $\Delta = 1$, and significance level $\alpha = 0.05$.

For each scenario, we calculated the pertinent sample size (n) according to our new procedure and we generated, under H_A , 10,000 samples of n units with $\mu_A = 0$ and $\sigma_A = \Delta / z_{p_A}$ (see Case 2).

The proportion of H_0 rejection corresponds to the "simulated power" (Tables 4.1 and 4.2), according to Shieh's terminology [16], and it has to be compared with the "estimated power" calculated from the sample size according to Shieh's procedure [16] and our procedure. Of course, the "estimated power" has to be very near to the required power under which the sample size has been calculated.

For easiness of comparison with the sample sizes, the simulated and estimated power shown in the Shieh's paper [16] have been reported in italic between brackets in the pertinent columns of the Tables 4.1 and 4.2.

It is possible to see that only the sample sizes obtained with our procedure fulfil the expected nominal power.

Table 3. Type I error rates of the agreement test: simulation study

μ_0	σ_0	Simulated proportion Our new procedure - AC	Simulated proportion Shieh procedure
0.00	0.780304146	0.0042	0.0487
0.10 (-0.10)	0.702273731	0.0245 (0.0235)	0.1625 (0.1662)
0.30 (-0.30)	0.546212902	0.0507 (0.0489)	0.2176 (0.2222)
0.50 (-0.50)	0.390152073	0.0507 (0.0489)	0.2176 (0.2222)
0.80 (-0.80)	0.156060829	0.0507 (0.0489)	0.2176 (0.2222)
0.90 (-0.90)	0.078030415	0.0507 (0.0489)	0.2176 (0.2222)
0.99 (-0.99)	0.007803041	0.0507 (0.0489)	0.2176 (0.2222)

Probability of the population centile $p = 0.90$ obtained from a population central proportion value of $P = 0.80$, $\Delta = 1$, nominal significance level $\alpha = 0.05$ and μ varying from 0 to -0.99 (a value near to $-\Delta$) or from 0 to 0.99 (a value near to Δ).

Comparisons among the different considered procedures: our new AC procedure, Shieh [16], Liu and Chow [17], B&A [2], Lin et al. [18] and Lu et al. [10].

All proposals can be formulated in a unifying way by considering the two sample distributions of:

$\bar{D} - k_{1-\alpha} S / \sqrt{n}$ and $\bar{D} + k_{1-\alpha} S / \sqrt{n}$ considered by Shieh [16].

However, it is worthwhile to underline that the values of T_{bl} and T_{bu} , shown in the formulas 22 (B&A's procedure [2]), 24 (Lin et al.'s procedure [18]), and 26 (Lu et al.'s procedure [10]) of Shieh's paper [16], have to be calculated with the quantiles $\hat{\theta}_L$ and $\hat{\theta}_U$, shown after equations 18 and 19 of Shieh's paper [16], instead of the quantiles $\hat{\theta}_{bl}$ and $\hat{\theta}_{bu}$ reported in

Shieh's paper [16].

The probability $\Pr\{E\} = \Pr\{\text{"reject } H_0\}$, on which the sample sizes substantially depend, can be formulated as:

$$\Pr\{E\} = \Pr\left\{ \begin{aligned} &(-\Delta < \bar{D} - k_{1-\alpha} S / \sqrt{n}) \wedge \\ &(\bar{D} + k_{1-\alpha} S / \sqrt{n} < \Delta) \end{aligned} \right\}$$

By using the non-central bivariate t distribution:

$$\Pr\{E\} = \Pr\{(k_{1-\alpha} < T_L) \wedge (T_U < -k_{1-\alpha})\} \text{ where}$$

$$T_L = (\bar{D} + \Delta) / (S / \sqrt{n}) \text{ and}$$

$$T_U = (\bar{D} - \Delta) / (S / \sqrt{n})$$

Table 4.1. Calculated sample size, "simulated power", and "estimated power" of the exact agreement test for $\Delta=1$, significance level α 0.05, and power = 0.80

Nominal Power	Null proportion	Alternative proportion	Sample size our new procedure (SS from Shieh)	Simulated power (Shieh)	Estimated power (Shieh)	Difference (Shieh)
0.8	0.8	0.90	134 (62)	0.8022 (0.3809)	0.8028 (0.3832)	-0.0006 (-0.0023)
0.8	0.8	0.95	44 (21)	0.8081 (0.3998)	0.8096 (0.3963)	-0.0015 (0.0035)
0.8	0.8	0.99	16 (9)	0.8337 (0.4791)	0.8288 (0.4747)	0.0049 (0.0044)
0.8	0.9	0.95	220 (118)	0.8005 (0.4745)	0.8008 (0.4753)	-0.0003 (-0.0008)
0.8	0.9	0.99	32 (18)	0.8086 (0.4857)	0.8064 (0.4858)	0.0022 (-0.0001)
0.8	0.95	0.99	78 (47)	0.8029 (0.5456)	0.8037 (0.5413)	-0.0008 (0.0043)

Table 4.2. Calculated sample size, "simulated power", and "estimated power" of the exact agreement test for $\Delta=1$, significance level α = 0.05, and power = 0.90

Nominal Power	Null proportion	Alternative proportion	Sample size our new procedure (SS from Shieh)	Simulated power (Shieh)	Estimated power (Shieh)	Difference (Shieh)
0.9	0.8	0.90	169 (86)	0.9021 (0.5542)	0.9006 (0.5552)	0.0015 (-0.0010)
0.9	0.8	0.95	55 (28)	0.9057 (0.5544)	0.9050 (0.5509)	0.0007 (0.0035)
0.9	0.8	0.99	19 (11)	0.9089 (0.6030)	0.9062 (0.5999)	0.0027 (0.0031)
0.9	0.9	0.95	282 (163)	0.9022 (0.6467)	0.9010 (0.6457)	0.0012 (0.0010)
0.9	0.9	0.99	40 (24)	0.9019 (0.6568)	0.9019 (0.6480)	0.0000 (0.0088)
0.9	0.95	0.99	100 (64)	0.8998 (0.6990)	0.9031 (0.7048)	-0.0033 (-0.0058)

So, the comparison among the sample sizes for the different procedures becomes the comparison among the different values of the coefficient $k_{1-\alpha}$.

Table 5 shows the values of "k" of the formulas from the different considered procedures.

Table 6 shows the $k_{1-\alpha}$ values for $\alpha = 0.05$ and $p=0.95$ at some sample sizes values. It has to be noted that, within each approach, the values of the sample size and the values of $k_{1-\alpha}$ increase accordingly.

Furthermore, within each sample size value, Shieh's $k_{1-\alpha}$ [16] is the lowest, but, as we have previously theoretically demonstrated and verified by simulation studies, the Type I and the Type II are not fulfilled. Then, in increasing order there are: the $k_{1-\alpha}$ from Lin et al. [18], the $k_{1-\alpha}$ from B&A [2], and the $k_{1-\alpha}$ from our new AC procedure, surprisingly, at first glance, equal the $k_{1-\alpha}$ coefficient obtained from Liu and Chow [17]. Furthermore, the coefficient $k_{1-\alpha}$ from Lu et al. [10] is less than our coefficient at $n = 30$ and more at $n = 200$; this pattern also applies to the sample size functions shown in Figure 4 which reports the sample size functions of all the considered approaches.

It is well evident that, among the considered procedures, the sample size function of Shieh's procedure [16] has the lowest values, but it has been calculated under a restrictive H_0 , as we have already reported. In addition, our sample size function is between those of Lu et al. [10] more liberal and B&A [5] more conservative. Lin et al. [18] function is practically superimposed on that of B&A [2]. The sample size function of Liu and Chow's procedure [17] does not appear since it is completely superimposed by the sample size function of our procedure owing to the fact that they are equivalent, as we will demonstrate. Finally, it has to be noted that for fixed values of p_0 , when the p_A values increase ($p_A > 0.98$, in Figure 4) the sample sizes decrease, becoming progressively very similar.

Check of the fulfilment of the nominal significance level α of the various methods

Considering Shieh's affirmation [16] about the conservatism of the TOST methods, we have carried out a simulation study to empirically estimate their significance level.

Table 5. Values of "k" of the formulas from the different considered procedures

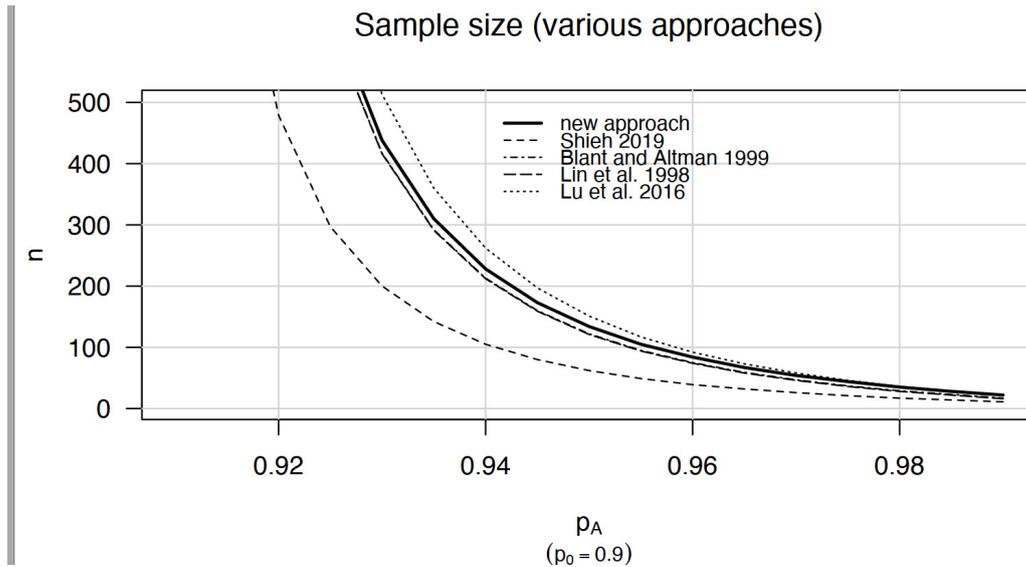
Methods	Coefficient $k_{1-\alpha}$
Our new procedure - AC	Solution $k_{1-\alpha}$ of: $\sup_{H_0} \Pr \{E\} = \alpha$
Shieh [16]	Solution $\gamma_{1-\alpha}$ of: $\Pr \{E \mid \mu = 0; \sigma = \Delta / z_p\} = \alpha$
Liu and Chow [17]	$t_{1-\alpha, n-1} z_p \sqrt{n}$
B&A [5]	$w_{1-\alpha} = z_p \sqrt{n} + t_{1-\alpha, n-1} \sqrt{b}$
Lin et al. [18]	$z_{1-\alpha} \sqrt{b} + z_p \sqrt{n}$
Lu et al. [10]	$t_{1-\alpha/2, n-1} \sqrt{b} + z_p \sqrt{n}$

Where $b = 1 + z_p^2 / 2$

Table 6. $k_{1-\alpha}$ values for $\alpha = 0.05$ and $p=0.95$ at some sample sizes values

Procedures \ sample size	30	50	100	200
Our new procedure - AC	12.15855	14.60171	19.26539	25.98244
Shieh [16]	10.74401	13.24633	17.96149	24.71098
Liu and Chow [17]	12.15855	14.60171	19.26539	25.98244
B&A [5]	11.61548	14.20249	18.99537	25.79654
Lin et al. [18]	11.53223	14.15387	18.97154	25.78474
Lu et al. [10]	12.14636	14.71331	19.49208	26.28648

Figure 4. Sample Size functions for $p_0 = 0.90$ with $p_A > 0.90$ of the four considered procedures



We used the same parameter values used by Shieh for obtaining the results shown in the Tables 2, 3 and 4 of his paper [16].

Particularly, we used four sample size values ($n = 30, 50, 100, 200$), three values of the population proportion ($p = 0.900, 0.950, 0.975$), and, finally $\Delta = 1$. We changed Shieh's [16] $\mu = 0$ and $\sigma = \Delta / z_{p_0}$ with $\mu = (0.9999 \text{ or } -0.9999)$ and $\sigma = (\Delta - \mu) / z_{p_0}$, and, finally, we simulated 100,000 samples. The results are shown in Tables 7.1, 7.2, and 7.3.

It is possible to see that Shieh's procedure [16] is too much liberal with α values ranging from 0.14504 to 0.23524 with an increasing trend, in agreement with the sample size increase. Therefore, Shieh's procedure [16] does not protect from the Type 1 error, as we have already shown (Table 3).

Our procedure and the equivalent Liu and Chow' procedure [17] show the best performance with α values ranging from 0.05012 to 0.05078; interestingly, the higher values (>0.05) are at the two extremes of the sample sizes and the lower values (<0.05) are at the intermediate sample sizes almost giving the impression of a curvilinear trend.

Furthermore, B&A's procedure [2] gives somewhat liberal α values ranging from 0.06131 to 0.08546 with a decreasing trend at the sample size increase.

Finally, Lu et al.'s procedure [10] tends to be somewhat/moderately conservative with α values ranging from 0.03401 to 0.05328 with only two values greater than 0.05 at the lowest sample size ($n = 20$).

Considerations about the canonical null hypothesis H_0 and Shieh's approach [16]

As reported in the Methods paragraph, H_0 is the union of two hypotheses: $H_{0A} \cup H_{0B}$.

Let's consider now a null hypothesis (H_0^*) more

restrictive, corresponding to the intersection of the two hypotheses H_{0A} and H_{0B} ($H_{0A} \cap H_{0B}$)

$H_0^* : \theta_{1-p} \leq -\Delta \text{ and } \theta_p \geq \Delta$ in which "and" has replaced the "or" of the canonical formulation of H_0 .

It has to be noted that H_0^* is contained in the null canonical hypothesis (H_0) as it is possible to see in Figure 5. The space of H_0^* corresponds to the region delimited by the extensions of the sides AC and BC of the triangle ABC in the half plane positive of the vertical Y axis ($Y > 0$) and it is well evident that is a subset of the space outside the ABC triangle of the canonical H_0 formulation.

According to this formulation, the individual agreement test should be formulated as:

$H_0^* : \theta_{1-p} \leq -\Delta \text{ and } \theta_p \geq \Delta \text{ vs. } H_A : -\Delta < \theta_{1-p} \text{ and } \theta_p < \Delta$

To determine the sample size according to H_0^* , it is necessary to carry out the same steps outlined for our new procedure in the canonical case, as previously reported.

a) -Determination of the supremum of $Pr\{E\}$ under H_0 .

Similarly to the canonical case, the $\sup_{H_0} Pr\{E\}$ has to be calculated on the boundary of H_0 that is, on the extensions of the sides AC and BC in the direction of the positive Y, but, unlike the canonical case, both conditions must be satisfied, i.e. it must be both: $\theta_{1-p} = -\Delta$ and $\theta_p = \Delta$.

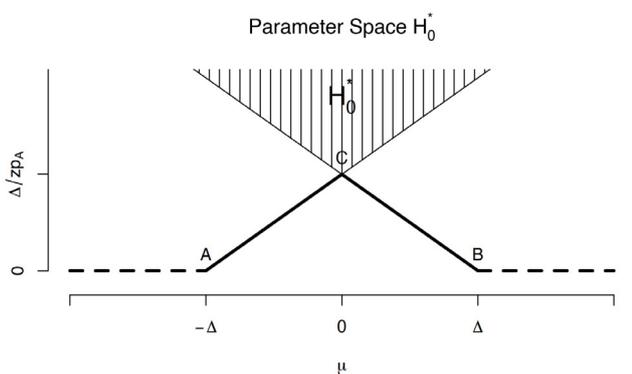
Remembering the definitions of θ_{1-p} and θ_p , this corresponds to a linear system of two equations in the unknowns μ and σ , whose solutions are $\mu = 0$ and $\sigma = \Delta / z_p$ the same ones reported by Shieh [16], corresponding to the vertex C of the triangle ABC.

Figure 5. Parameter space of the H_0^* hypothesis

Table 7.1	n=30	n=50	n=100	n=200
Our new procedure - AC	0.05031	0.04954	0.04938	0.05012
Shieh [16]	0.21658	0.22286	0.22895	0.23524
Liu and Chow [17]	0.05031	0.04954	0.04938	0.05012
B&A [5]	0.07958	0.07118	0.06526	0.06131
Lin et al. [18]	0.08596	0.07476	0.06708	0.06206
Lu et al. [10]	0.04791	0.04210	0.03647	0.03401

Table 7.2	n=30	n=50	n=100	n=200
Our new procedure - AC	0.05068	0.04978	0.04927	0.05031
Shieh [16]	0.17106	0.17548	0.18220	0.18723
Liu and Chow [17]	0.05068	0.04978	0.04927	0.05031
B&A [5]	0.08312	0.07379	0.06676	0.06255
Lin et al. [18]	0.08949	0.07755	0.06852	0.06356
Lu et al. [10]	0.05123	0.04436	0.03787	0.03447

Table 7.3	n=30	n=50	n=100	n=200
Our new procedure - AC	0.05078	0.04980	0.04941	0.05057
Shieh [16]	0.14505	0.14879	0.15511	0.15926
Liu and Chow [17]	0.05078	0.04980	0.04941	0.05057
B&A [5]	0.08546	0.07559	0.06794	0.06390
Lin et al. [18]	0.09171	0.07929	0.06972	0.06484
Lu et al. [10]	0.05328	0.04562	0.03860	0.03538



b)-Determination of the infimum of $Pr\{E\}$ under H_A

For this calculation, the H_A hypothesis is the same as previously reported and, consequently, the $\inf_{H_A} Pr\{E\}$ is on the point with coordinates $\mu = 0$; $\sigma = \Delta/z_p$ corresponding to the vertex C of the triangle ABC, similarly as it has been shown by Shieh [16].

Then, if we modify our procedure by using these values in the calculation of $\sup_{H_0} Pr\{E\}$ and $\inf_{H_A} Pr\{E\}$ or we use the procedure for the sample size calculation under two simple hypotheses (see paragraph titled "A particular approach: the sample size calculation under two simple hypotheses."), we obtain the same sample sizes as Shieh [16].

In conclusion Shieh's procedure [16] is valid only in

the case of the more restrictive hypothesis H_0^* , which is a part of the canonical null hypothesis H_0 . We have to restate that our procedure allows to obtain the sample size pertinent to the general case.

Considerations about Liu and Chow's TOST procedure [17] and our AC procedure

The explanation of the fact that these two procedures give the same sample size is based on the calculation of the coefficient $k_{1-\alpha}$ and of the test power. Indeed, our procedure calculates the coefficient $k_{1-\alpha}$ by referring to the non-central bivariate t distribution, while Liu and Chow [17] use a non-central univariate Student's t distribution. However, we demonstrate in the Appendix B that the calculation of the coefficient $k_{1-\alpha}$ using the non-central bivariate t distribution is equivalent to using the non-central univariate Student's t distribution.

Otherwise, regarding the power calculation, both procedures make use of the non-central bivariate t distribution. Thus, being equal the two quantities on which the sample size calculation is based, it is expected that the sample size values will be the same, fixed the parameters of the sample size calculation.

A relevant difference that has to be noted, is the fact that our procedure is based on a test of size α i.e. with a Type I error equal to α . Liu and Chow [17] claimed that "the proposed two one-sided tests procedure for

individual bioequivalence is size α test, but using the notation for “only a level α test (i.e. $\leq \alpha$)” according to Casella and Berger’s definition [20]. However, for completeness, we will demonstrate in the Appendix C that Liu and Chow’s test [17] is just a size α test (i.e. $= \alpha$), according to Casella and Berger’s definition [20].

DISCUSSION

The sample size calculation for agreement studies has to be appropriately done starting from sensible assumptions in addition to the usual power and significance level (two tailed).

Indeed, we think that all the studies need to be supported by a sample size calculation justified by a sound statistical methodology in order to have an adequate probability of obtaining their aim. So, the sentence: “All studies need a sample size justification. Not all studies need a sample size calculation.” from the Han et al.’s review [1] is, in our opinion, difficult to be justified and shareable. Indeed, we think that researchers have to be well aware of the effect sizes, differences, etc. that even a sample size of convenience allows to demonstrate.

Furthermore, we think that an approach based on the individual equivalence owing to its sensible rationale for assessing the agreement between measurement methods has to be absolutely supported even if the estimates of their systematic and proportional biases have to be absolutely supplied. As a further relevant point, sample sizes have to be such as to not compromise the real feasibility of an agreement study taking into account also the ethic aspects of the biomedical research [28].

A relevant point to be noted is that the sample sizes calculated according to Shieh [16] or our procedure does not depend on the values of μ and σ under the H_0 and H_A hypotheses.

Indeed, the sample size calculated with our procedure, fixed α , $1 - \beta$ and p_0 , depends in the Case 1 only on the ratio Δ^*/Δ and, in the Case 2, only on p_A , whatever the value assigned to Δ .

This rather surprising result occurs also in Shieh’s sample size calculation [16], fixed α , $1 - \beta$, p_0 and p_A ; indeed, the sample size is always the same independently from the value given to Δ , as it is possible to calculate from the IML® or R programs attached to Shieh’s paper [16].

Moreover, the sample size from the proposed our procedure is obtained with an exact statistical method similarly to the sample size obtained by Shieh’s procedure [16], instead of the approximate methods followed by B&A [2], Lin et al. [18], and Lu et al. [10].

In addition, our procedure allows to assess the individual bioequivalence and to calculate its pertinent sample size according to two different, but complementary approaches that are easy to switch between. Indeed, the first uses different Δ thresholds

(Δ_0, Δ_A) and the second uses different quantiles (z_{p_0}, z_{p_A}) or their corresponding probabilities p (p_0, p_A), under H_0 and under H_A . As a further relevant point, the sample sizes calculated for sensible agreement scenarios are adequate for the actual feasibility of the study.

As another relevant point, our approach considers the whole parameter space of H_0 and H_A hypotheses and allows to obtain a test of size α according to Casella and Berger [20].

Our procedure gives the same sample sizes of the Liu and Chow’s procedure [17], leading to conclude that the two procedures are equivalent and also that Liu and Chow’s procedure [17] is of size α as we have also directly demonstrated. Finally, it has to be stressed that these procedures have a better performance in comparison to the other considered procedures owing to their better control of the Type I and Type II errors, as we have theoretically demonstrated and as our simulation study has empirically confirmed (Tables 7.1, 7.2. and 7.3).

So, our procedure or Liu and Chow’s procedure [17] has to be warmly recommended.

We have demonstrated that Shieh’s proposal [16] is based on a particular case of the parameters space under H_0 and, consequently, it has some very important limitations. Particularly, under the canonical formulation of the H_0 and H_A hypotheses, it does not give the claimed statistical significance test of size α and do not fulfil the required power, being its sample sizes too much lower than the necessary ones.

However, it is debatable whether it is possible to shrink the H_0 space to Shieh’s formulation [16] without fulfilling the axiom of the complementarity between the null and the alternative hypotheses.

REFERENCES

1. Bland M, Altman A (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1(8476): 307-310.
2. Bland MJ, Altman D (1999) Measuring Agreement in Method Comparison Studies. *Stat Methods Med Res* 8(2):135-160.
3. Han O, Tan HW, Julious S, Sutton L, Jacques R, et al. (2022) Review of samples sizes used in agreement studies published in the PubMed repository *BMC Med Res Methodol* 22(1): 242.
4. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, et al. (2011) Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 64(1): 96-106.
5. Gerke O, Pedersen AP, Debrabant B, Halekoh U, Möller S (2022) Sample size determination in method comparison and observer variability studies. *J Clin Monit Comput* 36(5): 1241-1243.
6. Bland JM, Altman D (1990) A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement

- Comput Biol Med 20(5): 337-340.
7. Taffè P When can the Bland & Altman limits of agreement method be used and when it should not be used. *J Clin Epidemiol* 2021;137:176-181.
 8. Bland JM Website <https://www-users.york.ac.uk/~mb55/meas/sizemeth.htm>
 9. Cesana BM, Antonelli P (2012) Agreement analysis: further statistical insights. *Ophthal Physiol Opt* 32(5): 436-440.
 10. Lu MJ, Zhong WH, Liu YX, Miao HZ, Li YC, et al. (2016) Sample Size for Assessing Agreement Between Two Methods of Measurement by Bland-Altman Method. *The International Journal of Biostatistics* 12(2): 1-8.
 11. Shieh G (2016) Exact Power and Sample Size Calculations for the Two One-Sided Tests of Equivalence *Plos One* 11(9): e0162093.
 12. Shieh G (2018) The appropriateness of Bland-Altman's approximate confidence intervals for limits of agreement. *BMC Med Res Methodol* 18(1): 45.
 13. Carkeet A (2015) Exact Parametric Confidence Intervals for Bland-Altman Limits of Agreement. *Optom Vis Sci* 92(3): e71-80.
 14. Carkeet A, Goh YT (2018) Confidence and coverage for Bland-Altman limits of agreement and their approximate confidence intervals. *Stat Methods Med Res* 27(5):1559-74.
 15. Jan SL, Shieh G (2018) The Bland-Altman range of agreement: Exact interval procedure and sample size determination. *Comput Biol Med* 100:247-252.
 16. Shieh G (2020) Assessing Agreement Between Two Methods of Quantitative Measurements: Exact Test Procedure and Sample Size Calculation. *Stat Biopharm Res* 12(3):352-359.
 17. Liu JP, Chow SC (1997) A Two One-Sided Tests Procedure for Assessment of Individual Bioequivalence. *J Biopharm Stat* 7(1): 49-61.
 18. Lin SC, Whipple DM, Ho CS (1998) Evaluation of statistical equivalence using limits of agreement and associated sample size calculation. *Commun Stat A-Theor* 1998;27(6): 1419-1432.
 19. Krishnamoorthy K, Mathew T *Statistical Tolerance Regions: Theory, Applications, and Computation*. 2009, New York: Wiley. (page 35, equation 2.3.11).
 20. Casella G, Berger RL *Statistical Inference*, 1990, Duxbury Press, Belmont 2nd edition, 2002 Duxbury Pacific Grove Ca USA
 21. Owen DB (1965) A special case of a bivariate non-central t-distribution. *Biometrika* 52(3-4): 437-446.
 22. Dunnett CW, Sobel MA (1954) Bivariate Generalization of Student's t-distribution, with Tables for certain special cases. *Biometrika* 41(1-2): 153-169.
 23. <https://CRAN.R-project.org/package=OwenQ>
 24. <https://CRAN.R-project.org/package=PowerTOST>
 25. Genz A, Bretz F, Miwa T, Mi X, Leisch F, et al. (2021) mvtnorm: Multivariate Normal and t Distributions. R package version 1.1-3.
 26. Cochran WG (1954) Some Methods for Strengthening the Common χ^2 Tests. *Biometrics* 10(4): 417-451.
 27. Bradley JV (1978) Robustness? *Brit J Math Stat Psy* 31(2): 144-152.
 28. Cesana BM, Antonelli P (2019) Sample size calculations need to be adequate and parsimonious. *J Clin Epidemiol* 108:140-141.

Appendix A: demonstration that T_L and T_U are non-central Student's t distributions.

First of all, let's remember that: $T_L = (\bar{D} + \Delta) / (S / \sqrt{n})$ and $T_U = (\bar{D} - \Delta) / (S / \sqrt{n})$.

Considering the value of T_L , it is possible to write:

$$T_L = \frac{(\bar{D} + \Delta)}{S / \sqrt{n}} = \frac{(\bar{D} + \mu - \mu + \Delta)}{S / \sqrt{n}} \cdot \frac{\sigma / \sqrt{n}}{\sigma / \sqrt{n}} = \left[\frac{(\bar{D} - \mu)}{\sigma / \sqrt{n}} + \frac{(\mu + \Delta)}{\sigma / \sqrt{n}} \right] \cdot \frac{\sigma / \sqrt{n}}{S / \sqrt{n}}$$

Where μ and σ refer to the distribution of the differences (D).

Putting $\tau_L = \frac{(\mu + \Delta)}{\sigma / \sqrt{n}}$ and noting that $\frac{(\bar{D} - \mu)}{\sigma / \sqrt{n}} = Z$ (the standardized Gaussian distribution) and

$\frac{(n-1)S^2}{\sigma^2} = \chi_{n-1}^2$ (a χ^2 distribution with n-1 degrees of freedom independent from Z), we obtain:

$$T_L = \frac{Z + \tau_L}{S / \sigma} = \frac{Z + \tau_L}{\sqrt{\frac{(n-1)S^2 / \sigma^2}{n-1}}} = \frac{Z + \tau_L}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}}$$

That is, by definition a non-central univariate Student's t distribution with $\nu = n - 1$ degrees of freedom and non-centrality parameter given by τ_L , formally: $T_L \sim t_{\nu, \tau_L}$.

Similarly, considering the value of T_U , we obtain:

$$T_U = \frac{Z + \tau_U}{S / \sigma} = \frac{Z + \tau_U}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}} \text{ with } \tau_U = \frac{\mu - \Delta}{\sigma / \sqrt{n}}$$

Formally, $T_U \sim t_{\nu, \tau_U}$.

Appendix B

Demonstration that the calculation of $\sup_{H_0} P\{E\}$ can be done by means of a non-central univariate Student's t distribution and determination of the $k_{1-\alpha}$ coefficient corresponding to its $100(1-\alpha)$ th percentile.

Owen's Q functions

Non-central univariate Student's t

Owen [21] introduced the following Q-functions:

$$Q_1(t, \nu, \delta, R) := \frac{1}{\Gamma(\nu/2) \cdot 2^{(\nu-2)/2}} \int_0^R \Phi\left(\frac{tx}{\sqrt{\nu}} - \delta\right) x^{\nu-1} e^{-x^2/2} dx$$

$$Q_2(t, \nu, \delta, R) := \frac{1}{\Gamma(\nu/2) \cdot 2^{(\nu-2)/2}} \int_R^{+\infty} \Phi\left(\frac{tx}{\sqrt{\nu}} - \delta\right) x^{\nu-1} e^{-x^2/2} dx$$

where R is every value >0 .

Furthermore, the cumulative probability function of a non-central univariate Student's t with ν degrees of freedom and non-centrality parameter δ , is defined as:

$$F(t, \nu, \delta) := \frac{1}{\Gamma(\nu/2) \cdot 2^{(\nu-2)/2}} \int_0^{+\infty} \Phi\left(\frac{tx}{\sqrt{\nu}} - \delta\right) x^{\nu-1} e^{-x^2/2} dx$$

This function can be split in the sum of two terms, corresponding to the above reported Owen's function, as:

$$F(t, \nu, \delta) := \frac{1}{\Gamma(\nu/2) \cdot 2^{(\nu-2)/2}} \int_0^R \Phi\left(\frac{tx}{\sqrt{\nu}} - \delta\right) x^{\nu-1} e^{-x^2/2} dx + \frac{1}{\Gamma(\nu/2) \cdot 2^{(\nu-2)/2}} \int_R^{+\infty} \Phi\left(\frac{tx}{\sqrt{\nu}} - \delta\right) x^{\nu-1} e^{-x^2/2} dx =$$
$$= Q_1(t, \nu, \delta, R) + Q_2(t, \nu, \delta, R) \text{ for each } R > 0$$

It follows immediately that, for $R \rightarrow +\infty$, $F(t, \nu, \delta) = Q_1(t, \nu, \delta, +\infty)$

Non-central bivariate t

The cumulative probability function of a non-central bivariate t with ν degrees of freedom for $t_1 > t_2$ and $\delta_1 > \delta_2$, with correlation equal to 1, according to Owen [21] is:

$$\Pr(T_1 \leq t_1 \wedge T_2 \leq t_2) = \frac{1}{\Gamma(\nu/2) \cdot 2^{(\nu-2)/2}} \int_0^R \Phi\left(\frac{t_1 x}{\sqrt{\nu}} - \delta_1\right) x^{\nu-1} e^{-x^2/2} dx +$$
$$+ \frac{1}{\Gamma(\nu/2) \cdot 2^{(\nu-2)/2}} \int_R^{+\infty} \Phi\left(\frac{t_2 x}{\sqrt{\nu}} - \delta_2\right) x^{\nu-1} e^{-x^2/2} dx$$

$$\text{with } R = \frac{\delta_1 - \delta_2}{t_1 - t_2} \sqrt{\nu}.$$

So, by using the Owen's Q functions [21]:

$$\Pr(T_1 \leq t_1 \wedge T_2 \leq t_2) = Q_1(t_1, \nu, \delta_1, R) + Q_2(t_2, \nu, \delta_2, R)$$

From this distribution it is possible to calculate the probabilities of the following events:

$$E_2 = \{T_1 > t_1 \wedge T_2 \leq t_2\}, E_3 = \{T_1 \leq t_1 \wedge T_2 \geq t_2\}, \text{ and } E_4 = \{T_1 \geq t_1 \wedge T_2 \geq t_2\}$$

Particularly, the probability of the event E_2 , corresponding to our event E, is:

$$\begin{aligned}
\Pr\{E_2\} &= \Pr\{T_2 \leq t_2\} - \Pr\{T_1 \leq t_1 \wedge T_2 \leq t_2\} = \\
&= Q_1(t_2, \nu, \delta_2, R) + Q_2(t_2, \nu, \delta_2, R) - [Q_1(t_1, \nu, \delta_1, R) + Q_2(t_2, \nu, \delta_2, R)] = \\
&= Q_1(t_2, \nu, \delta_2, R) - Q_1(t_1, \nu, \delta_1, R)
\end{aligned}$$

Then, the probability of the event E, crucial for the sample size calculation, is given by:

$$\Pr\{E\} = \Pr\{T_L > k_{1-\alpha} \wedge T_U < -k_{1-\alpha}\} = \Pr\{t_{\nu, \tau_L} > k_{1-\alpha} \wedge t_{\nu, \tau_U} < -k_{1-\alpha}\}$$

where $[t_{\nu, \tau_L}; t_{\nu, \tau_U}]$ is a non-central bivariate t-distribution with degrees of freedom ν , non-centrality parameters τ_L and τ_U , and with correlation equal to 1.

Then, putting: $t_1 = k_{1-\alpha}, \delta_1 = \tau_L, t_2 = -k_{1-\alpha}, \delta_2 = \tau_U$, for keeping Owen's terminology [21], $\Pr\{E\}$ can be also expressed as: $\Pr\{E\} = Q_1(-k_{1-\alpha}, \nu, \tau_U, R) - Q_1(k_{1-\alpha}, \nu, \tau_L, R)$

Calculation of $\sup_{H_0} \Pr\{E\}$

Let's remember that the supremum of $\Pr\{E\}$ under H_0 is obtained in the point $(\mu, \sigma) = \left[-\Delta^+; (\Delta - |\mu|) / z_p = 0^+ \right]$ leading, consequently, to the following non-centrality parameters:

$$\tau_L = z_p \sqrt{n} \text{ and } \tau_U = -\infty, \text{ and, finally, } R = +\infty.$$

Otherwise, it can be also obtained in the point $(\mu, \sigma) = \left[\Delta^-; (\Delta - |\mu|) / z_p = 0^+ \right]$ with non-centrality parameters given by: $\tau_L = +\infty$ and $\tau_U = -z_p \sqrt{n}$.

$$\text{So: } \sup_{H_0} \Pr\{E\} = Q_1(-k_{1-\alpha}, \nu, -\infty, +\infty) - Q_1(k_{1-\alpha}, \nu, z_p \sqrt{n}, +\infty)$$

Let's analyse the two terms at the right of the above equality; we can see that:

$$\text{a) } Q_1(-k_{1-\alpha}, \nu, -\infty, +\infty) = 1.$$

In fact, in the formula of Q_1 the term $\Phi\left(\frac{tx}{\sqrt{v}} - \delta\right)$ becomes $\Phi(+\infty) = 1$; consequently, the integrand function of Q_1 becomes a χ density function that, integrating over all its domain, returns a value of 1.

b) From the relationships among Owen's Q functions [21] and the non-central univariate Student's t distribution, previously reported, we can write:

$$Q_1(k_{1-\alpha}, \nu, z_p \sqrt{n}, +\infty) = Q_1(k_{1-\alpha}, \nu, z_p \sqrt{n}, R) + Q_2(k_{1-\alpha}, \nu, z_p \sqrt{n}, R) = F_t(k_{1-\alpha}, \nu, z_p \sqrt{n})$$

Where $F_t(k_{1-\alpha}, \nu, \tau)$ indicates the cumulative distribution of the non-central t with ν degrees of freedom and non-centrality parameter τ .

$$\text{In conclusion: } \sup_{H_0} \Pr\{E\} = 1 - F_t(k_{1-\alpha}, \nu, z_p \sqrt{n})$$

Calculation of the coefficient $k_{1-\alpha}$

The coefficient $k_{1-\alpha}$ is calculated so that: $\sup_{H_0} \Pr\{E\} = \alpha$. Therefore, it is sufficient to solve the equation $1 - F_t(t, \nu, z_p \sqrt{n}) = \alpha$ in respect to t, that is $F_t(t, \nu, z_p \sqrt{n}) = 1 - \alpha$; the solution is exactly the 100(1- α)% centile of a non-central univariate Student's t with ν degrees of freedom and non-centrality parameter $\tau = z_p \sqrt{n}$ given by:

$$k_{1-\alpha} = t_{1-\alpha, \nu, z_p \sqrt{n}}$$

It can be seen that the coefficient $k_{1-\alpha}$ corresponds to the coefficient $\tau_{1-\alpha}$ of the Liu and Chow's procedure [17], described by Shieh [16] (formula 13 at page 3).

Numerical check of the previous theoretical results

In the open-source R language with $p_0 = 0.9$, $p_A = 0.95$, $\alpha = 0.05$, power = 0.80; we obtain, according to our procedure, a sample size of $n = 134$ with $k_{1-\alpha} = 17.25068$.

1) Verify that $Q_1(-k_{1-\alpha}, v, -\infty, +\infty) = 1$

```
> OwenQ::OwenQ1(nu = 133, t = -17.25068, delta = -Inf, R = .Machine$double.xmax)
[1] 1
```

It has to be noted that the parameters of OwenQ are: "nu" = degrees of freedom, t = our $k_{1-\alpha}$, delta = our non-centrality parameter, and R a constant in the Owen's formula. In addition, ".Machine\$double.xmax" represents the maximum numerical value obtainable in the open-source R language and approximates +Inf.

2) Verify that Owen's Q_1 function is equal to the cumulative probability of the non-central t distribution:

$$Q_1(k_{1-\alpha}, v, z_p \sqrt{n}, +\infty) = F_t(k_{1-\alpha}, v, z_p \sqrt{n})$$

2.1)- Calculation of $Q_1(k_{1-\alpha}, v, z_p \sqrt{n}, +\infty)$

```
>OwenQ::OwenQ1(nu= 133,t = 17.25068, delta = qnorm(p = 0.90)*sqrt(134), R =
.Machine$double.xmax)
[1] 0.9499998
```

2.2)- Calculation of $F_t(k_{1-\alpha}, v, z_p \sqrt{n})$ the cumulative probability of a non-central t distribution, using "pt" function of R.

```
> pt(q = 17.25068,df = 133, ncp = qnorm(p = .90)*sqrt(134))
[1] 0.9499998
```

3) Verify that $k_{1-\alpha}$ is equal to the $1-\alpha$ centile of the non-central t distribution: $k_{1-\alpha} = t_{1-\alpha, v, z_p \sqrt{n}}$

3.1)-Calculation of $k_{1-\alpha}$ by using an ad hoc written function for the new proposed our procedure

```
> k_coeff(n = 134,alpha = .05,p = .9)["k_coef"]
[1] 17.25068
```

3.2)-Calculation of $t_{1-\alpha, v, z_p \sqrt{n}}$ by using the "qt" function of R

```
> qt(p = .95, df = 133, ncp = qnorm(p = 0.90)*sqrt(134)) #
[1] 17.25068
```

Appendix C

Demonstration that the generalized procedure two-one sided test (TOST) for the assessment of the individual agreement is a size α test, in contrast to Shieh's affirmation [16].

In order for it to be a size α test, it is necessary to verify that the two conditions, indicated in theorem 8.3.24 of Casella and Berger's book [20, page 396], are satisfied.

The two tests that make up the TOST procedure are:

$$H_{01} : \theta_{1-p} \leq -\Delta \text{ vs. } H_{A1} : \theta_{1-p} > -\Delta$$

$$H_{02} : \theta_{1-p} \geq \Delta \text{ vs. } H_{A2} : \theta_{1-p} < \Delta$$

Let's us consider the first test that make up the TOST procedure.

Casella and Berger's first condition [20] is that a sequence of parameter points $(\mu_i; \sigma_i)$ in H_{01} exists such that:

$$\lim_{i \rightarrow +\infty} \Pr(\text{reject } H_{01} | (\mu_i; \sigma_i)) = \alpha.$$

The test function is: $\hat{\theta}_{1-p} = \bar{D} - k \cdot S / \sqrt{n}$, where k is an appropriate constant to be determined.

The rejection Region of H_{01} is $\hat{\theta}_{1-p} \leq -\Delta$ and it is equivalent, after some algebraic steps, to:

$$(\bar{D} + \Delta) / (S / \sqrt{n}) > k_{1-\alpha}.$$

Given, $T_L = (\bar{D} + \Delta) / (S / \sqrt{n})$, we have demonstrated, in Appendix A, that T_L is distributed as a non-central univariate Student's t with non-centrality parameter $\tau_L = (\mu + \Delta) / (\sigma / \sqrt{n})$.

In order that the partial test is of size α , k must satisfy the following condition:

$$\sup_{H_{01}} \Pr\{\hat{\theta}_{1-p} > -\Delta\} = \alpha \text{ or, equivalently : } \sup_{H_{01}} \Pr\{T_L > k\} = \alpha.$$

The superior has to be searched on the boundary of H_{01} and, in this particular case, is on the half straight line in the positive half plane of the Y axis on which lies the segment AC.

On the half line AC, $\sigma = ((\Delta + \mu) / z_p)$ and, consequently, $\tau_L = (\mu + \Delta) / ((\Delta + \mu) / (z_p \sqrt{n})) = z_p \sqrt{n}$

Thus, the condition $\sup_{H_{01}} \Pr\{T_L > k\} = \alpha$ becomes $\Pr\{T_L > k | \tau_L = z_p \sqrt{n}\} = \alpha$.

Since T_L is a non-central Student's t, the coefficient k which fulfils the above condition corresponds to the $(1-\alpha)$ th quantile of this non-central univariate Student's t distribution with non-centrality parameter $\tau_L = z_p \sqrt{n}$ and consequently it will be denoted as $k_{1-\alpha}$.

Thus, the test function becomes $\hat{\theta}_{1-p} = \bar{D} - k_{1-\alpha} S / \sqrt{n}$.

Let's consider a succession of points on the segment AC (see Figure 1, Panel A) belonging to the border of H_{01} (and also of H_0), which from C converge towards A.

We have:

$$\lim_{\mu \rightarrow -\Delta^+} \left(\mu; \sigma = \frac{\Delta + \mu}{z_p} \right) = (-\Delta^+; 0^+) \text{ converging to the point A}$$

$$\lim_{\mu \rightarrow -\Delta^+} \tau_L = \lim_{\mu \rightarrow -\Delta^+} \frac{\mu + \Delta}{\Delta + \mu} z_p \sqrt{n} = z_p \sqrt{n}$$

Thus, $\lim_{\mu \rightarrow -\Delta^+} \Pr\{T_L > k_{1-\alpha}\} = \Pr\{t_{(n-1); \tau_L} > k_{1-\alpha} \mid \tau_L = z_p \sqrt{n}\} = \alpha$.

The last equality is a consequent of what we have just reported about $k_{1-\alpha}$.
Therefore, Casella and Berger's first condition [20] is satisfied.

Casella and Berger's second condition [20] is that a sequence of parameter points $(\mu_i; \sigma_i)$ in H_{01} exists such that: $\lim_{i \rightarrow +\infty} \Pr(\text{reject } H_{02} \mid (\mu_i; \sigma_i)) = 1$.

Let us now consider the second test of the TOST procedure.

$H_{02} : \theta_p \geq \Delta$ vs. $H_{A2} : \theta_p < \Delta$

The test function is: $\hat{\theta}_{1-p} = \bar{D} + k_{1-\alpha} \cdot S / \sqrt{n}$ and the rejection Region of H_{02} is $\hat{\theta}_p < \Delta$ and, given

$T_U = (\bar{D} - \Delta) / (S / \sqrt{n})$ it is equivalent to $T_U < -k_{1-\alpha}$, where T_U is distributed as a non-central univariate

Student's t with non-centrality parameter $\tau_U = (\mu - \Delta) / (\sigma / \sqrt{n})$.

Let's consider the same succession of points on the segment AC (see Figure 1, Panel A) as we did previously.

We have:

$$\lim_{\mu \rightarrow -\Delta^+} \tau_U = \lim_{\mu \rightarrow -\Delta^+} \frac{\mu - \Delta}{\Delta + \mu} z_p \sqrt{n} = -\infty$$

Thus, $\lim_{\mu \rightarrow -\Delta^+} \Pr\{T_U < -k_{1-\alpha}\} = \Pr\{t_{n-1; \tau_U} < -k_{1-\alpha} \mid \tau_U = -\infty\} = 1$

The last equality is a consequence of the fact that the non-central univariate cumulative probability of the Student's t distribution is a decreasing function with respect to its non-centrality parameter.
Thus, Casella and Berger's second condition [20] is also satisfied.

We can therefore conclude that Liu and Chow's TOST procedure [17] is a size α test.