Conditional Type I Error Rate for Superiority Test Conditioned on Establishment of Noninferiority in Clinical Trials

Jiacheng Yuan, PhD
Research Data Science, Astellas Pharma Global Development, Inc, Deerfield, Illinois

Tiejun Tong, PhD
Department of Applied Mathematics, University of Colorado, Boulder, Colorado

Tie-Hua Ng, PhD
Center for Biologics Evaluation and Research, Food and Drug Administration, Rockville, Maryland

In clinical trials, it is often desirable to test for superiority conditioned on establishment of noninferiority based on the same primary endpoint. According to a guidance document issued by the European regulatory agency Committee for Proprietary Medicinal Products in 2001, no type I error rate adjustment is necessary for switching between superiority and noninferiority because the family-wise type I error rate is controlled at the same nominal level. However, Ng raised the issues of switching between superiority and noninferiority even though there is no inflation of the family-wise type I error rate and showed that the false discovery rate could be increased.

To alleviate these concerns, we propose to control the conditional type I error rate of the second-step superiority test at the nominal significance level, which leads to a lower (unconditional) significance level of the second-step superiority test. The suggested adjustment posts a more rigorous condition to claim superiority, which is an effort to decrease the number of erroneous claims of superiority.

Key Words
Clinical trial; Conditional test; Noninferiority; Superiority; Type I error rate

Correspondence Address
Jason Yuan, Research Data Science, Astellas Pharma Global Development, Inc, 3 Parkway North, Deerfield, IL 60015 (email: jason.yuan@us.astellas.com).

The views expressed in this article are those of the author and not an official position of the US Food and Drug Administration.

INTRODUCTION

In active controlled clinical trials, it is common to see designs with the primary objective to show that the test treatment is not too much inferior to the active control and continue to test for superiority if noninferiority is shown. For this design, Morikawa and Yoshida (1) argued that there is no need to adjust the type I error rate (significance level) based on the closed testing principle proposed by Marcus et al. (2). The European Agency for the Evaluation of Medicinal Products, Committee for Proprietary Medicinal Products (3) argued that multiplicity is not a concern when the study objective is switched between superiority and noninferiority because the test procedure is closed. More specifically, (a) superiority can be tested if the null hypothesis is rejected in a noninferiority trial, and (b) noninferiority may be tested if failing to reject the null hypothesis in a superiority trial with a prespecified margin; and no multiplicity adjustment is necessary. However, Ng (4,5) raised the issues of switching between superiority and noninferiority even though there is no inflation of the family-wise type I error rate and showed that the false discovery rate could be increased.

We suggest a new testing procedure that controls the conditional type I error rate, instead of the unconditional type I error rate, for the second-step superiority test. Specifically, we perform the usual single superiority test with a lower significance level so that the conditional type I error rate equals the nominal significance level. With our suggested adjustment, the significance level of the second-step test depends on the first-step test. The proposed adjustment posts a more rigorous condition to claim superiority, which is an effort to decrease the incidence of erroneous claims of superiority.

The rest of the article is organized as follows. In the next section, we formulate the statistical problem of switching from noninferiority to superiority. Following that, we elaborate the concerns raised by Ng (4). Then we derive the conditional type I error rate of the superiority test conditional on establishment of noninferiority and suggest an adjustment to the second-step test, followed by a hypothetical example to illustrate the proposed adjustment. Finally, we conclude with some discussion.

THE STATISTICAL PROBLEM

Consider an active controlled clinical trial, where the test and control treatments are com-

Drug Information Journal, *Vol. 45, pp. 331–336, 2011 • 0092-8615/2011*
Submitted for publication: April 1, 2010
Accepted for publication: November 15, 2010

pared based on a continuous primary endpoint. Suppose that the primary endpoint is normally distributed, with means μ_t and μ_c and known variances σ_t^2 and σ_c^2, for the test and control arms respectively. Without loss of generality, we assume that a larger value of the primary endpoint indicates a better outcome. The primary objective is to show noninferiority, and if noninferiority is shown, then superiority will be tested. Specifically, in the first step, we perform the following noninferiority test at a significance level of α_1:

$$(T_1)\ H_{10}: \mu_t - \mu_c \leq -\delta \text{ vs } H_{1a}: \mu_t - \mu_c > -\delta,$$

where δ (>0) denotes the noninferiority margin. If the null hypothesis H_{10} is rejected, then, in the second step, the superiority test will be conducted at a significance level of α_2 as follows:

$$(T_2)\ H_{20}: \mu_t - \mu_c \leq 0 \text{ vs } H_{2a}: \mu_t - \mu_c > 0.$$

For a given nominal significance level of α, we set $\alpha_1 = \alpha$. Furthermore, without an adjustment, we set $\alpha_2 = \alpha$, and the overall type I error rate will be controlled at α, as noted earlier. In a later section, we propose that the second-step superiority be tested at a significance level of α_2, which is less than α.

CONCERNS OF NO ADJUSTMENT

Ng (4) pointed out that switching between superiority and noninferiority without any adjustment reduces to simultaneous testing of both hypotheses with a one-sided confidence interval, and raised the issues of simultaneous testing of both hypotheses. In fact, simultaneous testing of both hypotheses allows a test treatment that is expected to have the same effect as an active control to claim superiority by chance alone without losing the chance of showing noninferiority. This would lead to a higher number for erroneous claims of superiority compared with the situation where only one null hypothesis is to be tested because of the following. If only one null hypothesis is to be tested, expecting the test treatment to have the same effect as an active control, researchers will likely choose to test noninferiority rather than superiority. However, with simultaneous testing, superiority will be tested regardless of the expectation. Therefore, more test treatments that are expected to have the same effect as an active control would be tested for superiority with simultaneous testing than would be if only one null hypothesis is to be tested, resulting in more erroneous claims of superiority.

CONDITIONAL TYPE I ERROR RATE

To alleviate the concerns raised by Ng (4,5), we propose to control the conditional type I error rate of the second-step superiority test at the nominal significance level of α. To do so, let $\Phi(\cdot)$ and z_α be the cumulative distribution function (CDF) and the upper α-quantile of a standard normal distribution, respectively. Let $\sigma^2 = \sigma_t^2/n_t + \sigma_c^2/n_c$ where n_t and n_c denote the sample size for the test and control arms, respectively, and σ^2 is known because σ_t^2 and σ_c^2 are known by assumption. Let $W = \sum X_i/n_t - \sum Y_j/n_c$, where, for $i = 1, \ldots, n_t$ and $j = 1, \ldots, n_c$, X_i and Y_j are the individual values of the primary endpoint for the test treatment and control, respectively. Let the first-step noninferiority hypothesis be tested at a significance level of α_1; so that the null hypothesis H_{10} is rejected when $W > -\delta + z_{\alpha 1}\sigma$. In addition, let the second-step superiority hypothesis be tested at a significance level of α_2; so that the null hypothesis H_{20} is rejected when $W > z_{\alpha 2}\sigma$. If $\alpha_2 \leq \alpha_1$, then the conditional type I error rate, Ψ, for the second-step superiority test T_2 is given by

$$\begin{aligned}\psi &= \sup_{\mu_t - \mu_c \leq 0} P(W > z_{\alpha_2}\sigma \mid W > -\delta + z_{\alpha_1}\sigma)\\ &= \alpha_2 / \Phi(-z_{\alpha_1} + \delta/\sigma).\end{aligned}$$

The derivation of Ψ is given in the appendix.

We suggest control of the conditional type I error rate of T_2 at the nominal significance level of α. To do so, we set $\Psi = \alpha$. We then solve for α_2, and get $\alpha_2 = \alpha\Phi(-z_{\alpha 1} + \delta/\sigma)$.

In summary, our suggested testing procedure is as follows: (a) perform the first-step noninferiority test as usual with type I error rate $\alpha_1 = \alpha$, and (b) perform the conditional superiority test as a single superiority test with type I error rate $\alpha_2 = \alpha\Phi(-z_\alpha + \delta/\sigma)$ instead of the nominal level of α. Noting that $\alpha_2 < \alpha$, our suggestion posts a more rigorous condition to establish superiority in the second-step conditional test.

A HYPOTHETICAL EXAMPLE

In this section, we consider a hypothetical example where the primary endpoint is normally distributed with a larger value indicating a better outcome. Let's assume equal means ($\mu_0 = 5.0$) and equal standard deviations ($\sigma_0 = 5.0$) between the two arms, 1:1 randomization, a noninferiority margin of $\delta = 1.6$, a power of 80%, and a significance level of 0.025 for the noninferiority test. The design is to first test for noninferiority:

$$(T_1)\ H_{10}: \mu_t - \mu_c \le -1.6 \text{ vs } H_{1a}: \mu_t - \mu_c > -1.6.$$

If noninferiority can be established, then test for superiority:

$$(T_2)\ H_{20}: \mu_t - \mu_c \le 0 \text{ vs } H_{2a}: \mu_t - \mu_c > 0.$$

According to our suggestion, the conditional superiority test, if performed, should be conducted as a usual single superiority test with a significance level of $\alpha_2 = \alpha\Phi(-z_\alpha + \delta/\sigma)$. With a power of $(1 - \beta)$ to conclude noninferiority assuming no treatment difference, we have $\delta/\sigma = z_\alpha + z_\beta$. Then $\alpha_2 = \alpha\Phi(z_\beta) = \alpha(1 - \beta) = 0.020$.

DISCUSSION

UNKNOWN VARIANCES

The derivation in the section Conditional Type I Error Rate is based on the assumption that the variances are known. When σ_t^2 and σ_c^2 are unknown, for ease of exposition, in what follows we assume equal variance so that the statistical test based on the t distribution may be performed. Note that with unequal variance, Welch's t test statistic should be employed with the degrees of freedom approximated by the Welch-Satterthwaite equation (6).

Let s_p be the pooled estimate of the common standard deviation, and denote $s^* = s_p\sqrt{1/n_t + 1/n_c}$. Similarly, as previously, the usual tests reject H_{10} when $(W+\delta)/s^* > t_{\nu,\alpha_1}$ and reject H_{20} when $W/s^* > t_{\nu,\alpha_2}$, where $t_{\nu,\alpha}$ is the upper α-quantile of the Student's t distribution with $\nu = n_t + n_c - 2$ degrees of freedom. Then for any given $\alpha_2 \le \alpha_1$, the conditional type I error rate for the second-step superiority test is

$$\psi = \sup_{\mu_t - \mu_c \le 0} P\left(\frac{W}{s^*} > t_{\nu,\alpha_2} \,\middle|\, \frac{W+\delta}{s^*} > t_{\nu,\alpha_1}\right).$$

Let $F_{\nu,\theta}(\cdot)$ denote the CDF of the noncentral t distribution with ν degrees of freedom and noncentrality parameter θ. Then by the fact that W/s^* follows a noncentral t distribution with ν degrees of freedom and noncentrality parameter $(\mu_t - \mu_c)/\sigma$, we have

$$\psi = \sup_{\mu_t - \mu_c \le 0} \frac{P\left(\frac{W}{s^*} > t_{\nu,\alpha_2}\right)}{P\left(\frac{W+\delta}{s^*} > t_{\nu,\alpha_1}\right)} = \sup_{\mu_t - \mu_c \le 0} \frac{\bar{F}_{\nu,(\mu_t-\mu_c)/\sigma}(t_{\nu,\alpha_2})}{\bar{F}_{\nu,(\mu_t-\mu_c+\delta)/\sigma}(t_{\nu,\alpha_1})},$$

where $\bar{F}_{\nu,\theta}(\cdot) = 1 - F_{\nu,\theta}(\cdot)$ is defined as the right tail probability. With different values of $\alpha = \delta/\sigma$, ν, α_1, and α_2, Figure 1 shows that $\bar{F}_{\nu,(\mu_t-\mu_c)/\sigma}(t_{\nu,\alpha_2})/\bar{F}_{\nu,(\mu_t-\mu_c+\delta)/\sigma}(t_{\nu,\alpha_1})$ is a monotonically increasing function of the parameter $\mu_t - \mu_c$. (In the figure, the ratio increases with $x = (\mu_t - \mu_c)/\sigma$ and σ is a fixed number.)

However, due to the complexity of the density function of noncentral t distribution, it is not straightforward to analytically verify that $\bar{F}_{\nu,(\mu_t-\mu_c)/\sigma}(t_{\nu,\alpha_2})/\bar{F}_{\nu,(\mu_t-\mu_c+\delta)/\sigma}(t_{\nu,\alpha_1})$ is a monotonically increasing function of the parameter $\mu_t - \mu_c$ like that in Lemma 1 in the appendix.

If the monotonicity holds, then

$$\sup_{\mu_t - \mu_c \le 0} \frac{\bar{F}_{\nu,(\mu_t-\mu_c)/\sigma}(t_{\nu,\alpha_2})}{\bar{F}_{\nu,(\mu_t-\mu_c+\delta)/\sigma}(t_{\nu,\alpha_1})} = \frac{\bar{F}_{\nu}(t_{\nu,\alpha_2})}{\bar{F}_{\nu,\delta/\sigma}(t_{\nu,\alpha_1})} = \frac{\alpha_2}{\bar{F}_{\nu,\delta/\sigma}(t_{\nu,\alpha_1})}.$$

Similar to the situation with known variances in the section on conditional type I error rate, to control the conditional type I error rate of T_2 at the nominal significance level of α, with $\alpha_1 = \alpha$, we get $\alpha_2 = \alpha\bar{F}_{\nu,\delta/\sigma}(t_{\nu,\alpha})$. Then we have to estimate σ for the adjustment.

Earlier, we suggested performing the condi-

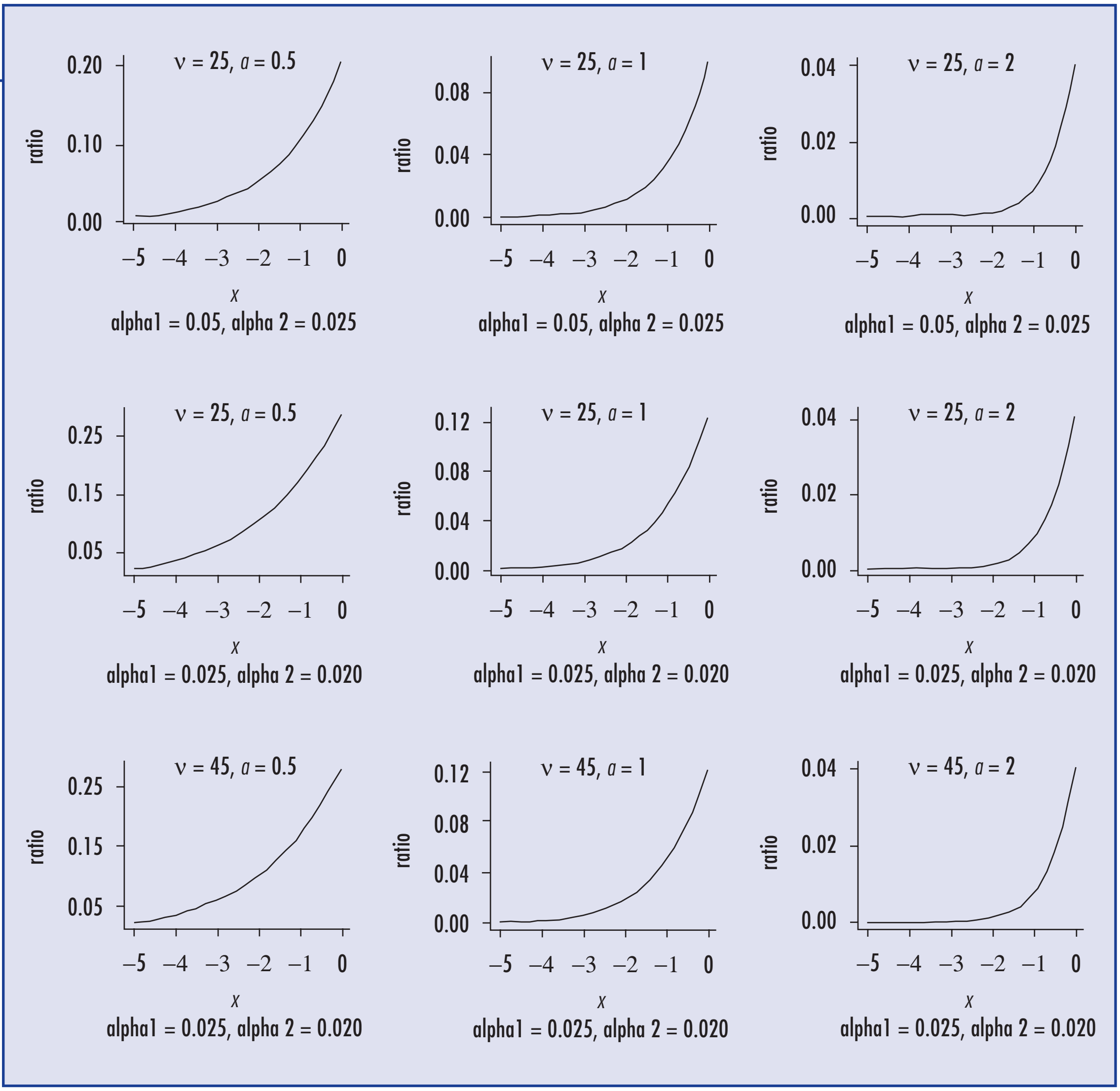

FIGURE 1

Monotonicity of the ratio. In the figure, x represents $(\mu_t - \mu_c)/\sigma$, *a represents* δ/σ, *and ratio represents* $\bar{F}_{\nu,(\mu_t-\mu_c)/\sigma}(t_{\nu,\alpha_2})/\bar{F}_{\nu,(\mu_t-\mu_c+\delta)/\sigma}(t_{\nu,\alpha_1})$.

tional superiority test as a single superiority test with type I error rate $\alpha_2 = \alpha\Phi(-z_\alpha + \delta/\sigma)$. Although it is "natural" to use this formula with σ being replaced with its estimate when it is not known, it is not appropriate to do so because this formula is based on known variances. When the variances are not known, the formula $\alpha_2 = \alpha\bar{F}_{\nu,\delta/\sigma}(t_{\nu,\alpha})$ should be used instead, where σ has to be estimated independent of the current trial. To be conservative, $\alpha_2 = \alpha^2$ may be used, where σ approaches infinity. Table 1 shows the comparison between the conservative $\alpha_{2,\text{cons}} = \alpha^2$ and the conditional error rate–based $\alpha_{2,cerb} = \alpha\bar{F}_{\nu,\delta/\sigma}(t_{\nu,\alpha})$ for $\alpha = 0.025$, where the ratio of $\alpha_{2,\text{cons}}/\alpha_{2,\text{cerb}}$ in percentages is shown with different combinations of ν and δ/σ. From this table, it is obvious that the conservative adjustment allows a much lower type I error rate. Figure 2 shows the ratio of $\alpha_{2,\text{cons}}/\alpha_{2,\text{cerb}}$ with δ/σ in the range of (0, 2) and ν taking values of 10, 20, 60, and 100. The ratio does not change much as the degree of freedom changes.

EXTREME SITUATION

The value of δ should be decided on a clinically meaningful basis, and not extremely small. In fact, if $\delta \to 0$, establishment of noninferiority should be equivalent to that of superiority. Although it makes no sense to test for noninferiority with a very large margin, the discussion here is an exercise to see if α_2 makes sense mathematically. In the situation where the noninferiority margin approaches $+\infty$, the adjusted significance level for the conditional superiori-

TABLE 1

Comparison Between the Conservative $\alpha_{2,cons} = \alpha^2$ and the Conditional Error Rate–Based $\alpha_{2,cerb} = \alpha \bar{F}_{\nu,\delta/\sigma}(t_{\nu,\alpha})$

	δ/σ = 0.2 (%)	0.5 (%)	1 (%)	2 (%)
ν = 15	65.5	36.8	16.3	5.4
20	65.1	36.2	15.9	5.2
25	64.8	35.9	15.7	5.2
30	64.6	35.7	15.6	5.1
35	64.5	35.6	15.5	5.1

The ratio of $\alpha_{2,cons}/\alpha_{2,cerb}$ in percentages is shown with different combination of ν and δ/σ.

ty test is $\alpha_2 = \alpha\Phi(-z_\alpha + \delta/\sigma) \to \alpha\Phi(+\infty) = \alpha$, which means no adjustment. Is this reasonable? When $\delta \to +\infty$, the first step noninferiority test becomes

$$(T_1)\ H_{10}: \mu_t - \mu_c \le -\infty \text{ vs } H_{1a}: \mu_t - \mu_c > -\infty,$$

which does not leave the chance for a type I error for itself because the alternative hypothesis $\mu_t - \mu_c > -\infty$ is always true. In this case, it is reasonable for no adjustment.

APPENDIX: DERIVATION OF ψ WHEN σ_t^2 AND σ_c^2 ARE KNOWN

Denote the first and second significance levels with α_1 and α_2, respectively, and assume $\alpha_2 \le \alpha_1$. Then the conditional type I error rate of the second-step test is

$$\begin{aligned}\psi &= \sup_{\mu_t-\mu_c\le 0} P(W > z_{\alpha_2}\sigma \mid W > -\delta + z_{\alpha_1}\sigma)\\ &= \sup_{\mu_t-\mu_c\le 0} \frac{P(W > z_{\alpha_2}\sigma)}{P(W > -\delta + z_{\alpha_1}\sigma)}\\ &= \sup_{\mu_t-\mu_c\le 0} \frac{\Phi\left(\frac{\mu_t-\mu_c}{\sigma} - z_{\alpha_2}\right)}{\Phi\left(\frac{\mu_t-\mu_c}{\sigma} - z_{\alpha_1} + \frac{\delta}{\sigma}\right)}\\ &= \sup_{x\le -z_{\alpha_2}} \frac{\Phi(x)}{\Phi\left(x + z_{\alpha_2} - z_{\alpha_1} + \frac{\delta}{\sigma}\right)},\end{aligned}$$

where $x = \frac{\mu_t - \mu_c}{\sigma} - z_{\alpha_2}$. Note that $z_{\alpha_2} - z_{\alpha_1} + \frac{\delta}{\sigma} > 0$. Thus by Lemma 1, we have

$$\begin{aligned}\psi &= \frac{\Phi(-z_{\alpha_2})}{\Phi\left(-z_{\alpha_2} + z_{\alpha_2} - z_{\alpha_1} + \frac{\delta}{\sigma}\right)}\\ &= \alpha_2/\Phi(-z_{\alpha_1} + \delta/\sigma).\end{aligned}$$

Lemma 1: For any constant $a > 0$, the ratio $\Phi(x)/\Phi(x + a)$ is a monotonically increasing function of $x \in (-\infty, \infty)$, where $\Phi(x)$ is the CDF of a standard normal random variable.

Proof. Let $g(x) = \Phi(x)/\Phi(x + a)$. To show that $g(x)$ is a monotonically increasing function, it suffices to show that $(d/dx)\, g(x) > 0$ for any $x \in (-\infty, \infty)$. Note that

$$\begin{aligned}\frac{d}{dx} g(x) &= \frac{d}{dx}\left(\frac{\int_{-\infty}^{x} \phi(t)dt}{\int_{-\infty}^{x+a} \phi(t)dt}\right)\\ &= \frac{\phi(x)\int_{-\infty}^{x+a}\phi(t)dt - \phi(x+a)\int_{-\infty}^{x}\phi(t)dt}{\left(\int_{-\infty}^{x+a}\phi(t)dt\right)^2}.\end{aligned}$$

Thus, we need to show that $\phi(x)\int_{-\infty}^{x+a}\phi(t)dt - \phi(x+a)\int_{-\infty}^{x}\phi(t)dt > 0$, or equivalently

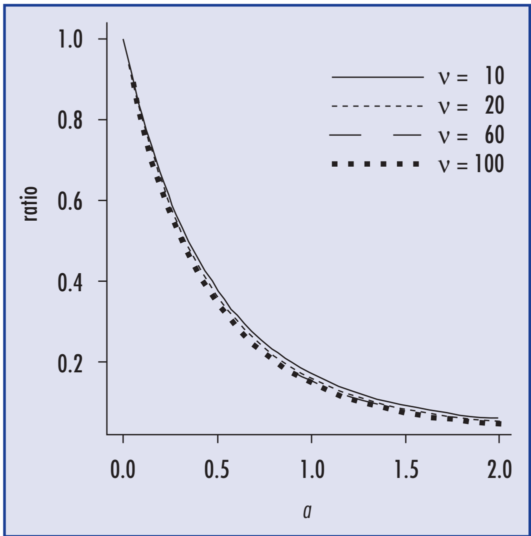

FIGURE 2

Conservatism with different values of δ/σ. In the figure, ratio represents $\alpha_{2,cons}/\alpha_{2,cerb}$, a represents δ/σ, and ν is degrees of freedom.

$$\int_{-\infty}^{x+a} \frac{\phi(t)}{\phi(x+a)} dt > \int_{-\infty}^{x} \frac{\phi(t)}{\phi(x)} dt, \quad \text{for any } x \in (-\infty, \infty). \tag{A1}$$

By letting $u = t - a$, we have

$$\int_{-\infty}^{x+a} \frac{\phi(t)}{\phi(x+a)} dt = \int_{-\infty}^{x} \frac{\phi(u+a)}{\phi(x+a)} du = \int_{-\infty}^{x} \frac{\phi(t+a)}{\phi(x+a)} dt.$$

Then by the fact that

$$\frac{\phi(t+a)}{\phi(x+a)} - \frac{\phi(t)}{\phi(x)} = \frac{e^{-(t+a)^2/2}}{e^{-(x+a)^2/2}} - \frac{e^{-t^2/2}}{e^{-x^2/2}}$$
$$= \frac{e^{-t^2/2}}{e^{-x^2/2}} \left(e^{(x-t)a} - 1\right) > 0,$$

for any $t < x$, (A1) holds and so is the lemma.

REFERENCES

1. Morikawa T, Yoshida M. A useful testing strategy in phase III trials: combined test of superiority and test of equivalence. *J Biopharm Stat*. 1995;5:297–306.
2. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 1976;63:655–660.
3. Committee for Proprietary Medicinal Products. Points to consider on switching between superiority and non-inferiority. *Br J Clin Pharmacol*. 2001;52:223–228.
4. Ng TH. Issues of simultaneous tests for non-inferiority and superiority (with comments by G. Pennello and W. Maurer, and rejoinder). *J Biopharm Stat*. 2003;13:629–662.
5. Ng TH. Simultaneous testing of non-inferiority and superiority increases the false discovery rate. *J Biopharm Stat*. 2007;17:259–264.
6. Welch BL. The generalization of "student's" problem when several different population variances are involved. *Biometrika*. 1947;34:28–35.

The authors report no relevant relationships to disclose.