PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. **9**: 207–216 (2010)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/pst.455

Sample size and proportion of Japanese patients in multi-regional trials‡

Kimitoshi Ikeda[1,*,†] and Frank Bretz[2,3]
[1]*Novartis Pharma KK, Tokyo, Japan*
[2]*Novartis Pharma AG, Basel, Switzerland*
[3]*Hannover Medical School, Hannover, Germany*

In recent years, multi-regional trials have received increasing attention by pharmaceutical companies carrying out global drug development programs. In Japan, new drugs are often approved several years after market release in other countries. The recently published guidance on 'Basic Principles on Global Clinical Trials' addresses specifically this time lag. A multi-regional trial has at least two main objectives. First, it is necessary to show a significant benefit in effect of a new drug in the entire population. Second, one needs to demonstrate that the results for a particular region are consistent with those from the entire population. In this paper, we discuss the methods proposed in the Japanese regulatory guidance document and derive closed form expressions for the resulting probabilities, which require the evaluation of multivariate normal or* t *probabilities. In addition, we propose an alternative method with better operating characteristics than the current approaches. Moreover, we examine the performance of our suggested method by simulating the probability of achieving the objectives and calculating the false-positive error rate. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: ***consistency; global trial; multivariate normal probability***

1. INTRODUCTION

A multi-regional trial is a single study conducted simultaneously across several regions under a common protocol. It usually aims at conclusions about the drug effect across all regions included in the study. A multi-regional trial conducted for bridging purposes in the context of a global development program would facilitate registration in all regions [1]. In recent years, multi-regional trials have received increasing attention, as they offer the opportunity to reduce resources by avoiding the conduct of duplicated trials.

The 11th Q&A for the ICH E5 guideline [1] gives two main objectives of multi-regional trials. One is to show that the drug is effective in the

*Correspondence to: Kimitoshi Ikeda, Novartis Pharma KK, Tokyo, Japan.
†E-mail: kimitoshi.ikeda@novartis.com
‡This article is published in *Pharmaceutical Statistics* as a special issue on *Multi-regional Clinical Trials — What are the challenges?*, edited by *Sue Jane Wang, Office of Biostatistics, Office of Translational Sciences, CDER, U.S. Food and Drug Administration, MD, USA*

Pharmaceutical STATISTICS

individual regions and the other is to compare the results of the study between regions with the goal of establishing that the drug is not sensitive to ethnic factors. Moreover, the guideline [1] states that a multi-regional study should be designed with sufficient numbers of subjects so that there is adequate power to have a reasonable likelihood of showing an effect in each region of interest. However, effectiveness criteria and details on the sample size determination for the individual regions are not given in [1]. These depend on scientific aspects and regulatory requirements, which may be different from region to region.

One may argue that the design and analysis of multi-regional trials are similar to those of multi-center trials. In both cases data are collected from multiple units, such as centers or regions, and intended to be analyzed as a whole, where heterogeneity of the treatment effect across units may exist [2]. In order to explore the presence of heterogeneity across units, one usually examines possible treatment-by-center interactions. In addition, individual center results should be presented and any extreme or opposite results among centers should be discussed [3–5]. The Q&A for the ICH E9 guideline states that the number of patients in each center should be at least 10. Shao and Chow [6] argued that this number should not be less than the number of centers. These proposals focus on investigating treatment-by-center interactions and the resulting sample sizes may not be adequate to show effectiveness of a new drug in an individual region, as required in a multi-regional trial.

In Japan, new drugs are often released several years after market release in other countries. The recent guidance on 'Basic Principles on Global Clinical Trials' published by the Japanese Ministry of Health, Labour, and Welfare addresses specifically this time lag [7]. This guidance introduces two methods for determining the size of the Japanese population in a global trial to achieve consistent results between the entire population and the Japanese population. In what follows, let D_{ALL} denote the observed effect difference in the entire study population across all regions between placebo and an active treatment under investigation. Similarly, let D_{J} denote the observed effect difference for the Japanese population. The first method in [7], 'Method 1' for short, requires that the number of Japanese patients is sufficiently large to ensure $D_{\mathrm{J}}/D_{\mathrm{ALL}} > \omega$ with a probability of at least 80%, where ω is a pre-specified threshold and $\omega \geqslant 0.5$ is generally recommended; see [8,9] for further details. The second method in [7], 'Method 2' for short, requires that the sample size is sufficiently large to demonstrate a consistent trend for all individual regions. Assume that, for example, three regions are included in the global trial and let D_1, D_2, and D_3 denote the observed treatment differences. Then the number of subjects is determined such that each individual difference D_1, D_2, and D_3 is larger than 0 with a probability of at least 80%; see [10] for further details.

In this paper, we focus on the following two objectives of a multi-regional trial and discuss the Japanese sample size requirement using Method 1. First, it is necessary to show a significant benefit in effect of a new drug over placebo in the entire study population. Second, one needs to demonstrate that the results for the Japanese subpopulation are consistent with those from the entire population. There are several possibilities to satisfy these general requirements, which all relate to interesting multiplicity issues. We discuss Method 1 from [7] and derive closed form expressions for the resulting probabilities, which require the evaluation of multivariate normal or t probabilities. In addition, we propose an alternative method, which has better operating characteristics than current approaches.

2. CHARACTERIZATION OF METHOD 1

In this section, we focus on a more detailed description of Method 1. We give closed form expressions for the probability of observing a consistent result, determine the necessary proportion of Japanese patients, and give a notable property of Method 1.

2.1. Probability of observing a consistent result

Let the random variables $Y_{\mathrm{J}a} \sim \mathcal{N}(\mu_{\mathrm{J}a}, \sigma^2)$ and $Y_{\mathrm{J}b} \sim \mathcal{N}(\mu_{\mathrm{J}b}, \sigma^2)$ denote the normal distributed

DOI: 10.1002/pst

outcomes of a primary variable for Japanese patients receiving either an active drug a or placebo b, where σ^2 denotes the common variance. For simplicity, we assume that σ is known. Extensions to unknown σ are straightforward and briefly discussed further below. Similarly, let $Y_{\mathrm{NJ}a} \sim \mathcal{N}(\mu_{\mathrm{NJ}a}, \sigma^2)$ and $Y_{\mathrm{NJ}b} \sim \mathcal{N}(\mu_{\mathrm{NJ}b}, \sigma^2)$ for non-Japanese patients. Let n_a and n_b denote the total numbers of patients in the study groups a and b, respectively. In addition, let p denote the proportion of Japanese patients in the study. The differences between the placebo group and the study drug group in the Japanese subpopulation and the non-Japanese subpopulation are distributed as

$$D_{\mathrm{J}} \sim \mathcal{N}\left(\delta_{\mathrm{J}}, \frac{n_a+n_b}{n_a n_b p}\sigma^2\right) \quad \text{and}$$

$$D_{\mathrm{NJ}} \sim \mathcal{N}\left(\delta_{\mathrm{NJ}}, \frac{n_a+n_b}{n_a n_b (1-p)}\sigma^2\right)$$

where $\delta_{\mathrm{J}} = \mu_{\mathrm{J}a} - \mu_{\mathrm{J}b}$ and $\delta_{\mathrm{NJ}} = \mu_{\mathrm{NJ}a} - \mu_{\mathrm{NJ}b}$. Let $D_{\mathrm{ALL}} = pD_{\mathrm{J}} + (1-p)D_{\mathrm{NJ}}$ denote the difference between the two treatment groups in all patients. Then the vector $\boldsymbol{d} = (D_{\mathrm{ALL}}, D_J)^t$ follows the bivariate normal distribution

$$\boldsymbol{d} = \begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix}\begin{pmatrix} D_{\mathrm{NJ}} \\ D_{\mathrm{J}} \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\delta}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\delta} = (\delta_{\mathrm{ALL}}, \delta_{\mathrm{L}})^t$, $\delta_{\mathrm{ALL}} = p\delta_{\mathrm{J}} + (1-p)\delta_{\mathrm{NJ}}$ and

$$\boldsymbol{\Sigma} = \frac{n_a+n_b}{n_a n_b}\sigma^2\begin{pmatrix} 1 & 1 \\ 1 & \frac{1}{p} \end{pmatrix}$$

To understand better the characteristics of Method 1, we visualize the joint distribution of D_{ALL} and D_{J} in Figure 1. The shaded area (a) in Figure 1 is the area of obtaining a consistent result with Method 1 and at the same time detecting a statistically significant difference between groups a and b for the global population, when the variance σ^2 is known. More formally, area (a) is defined through the linear constraints $(D_{\mathrm{ALL}}/\sigma)\sqrt{n_a n_b/(n_a+n_b)} > z_{1-\alpha}$ and $D_{\mathrm{J}}/D_{\mathrm{ALL}} > \omega$, where $z_{1-\alpha}$ denotes the $1-\alpha$ quantile of the standard normal distribution and α denotes the significance level for the one-sided test conducted in all patients.

On the basis of these preliminaries, we now investigate the calculation of four relevant probabilities. For simplicity, we consider here only the

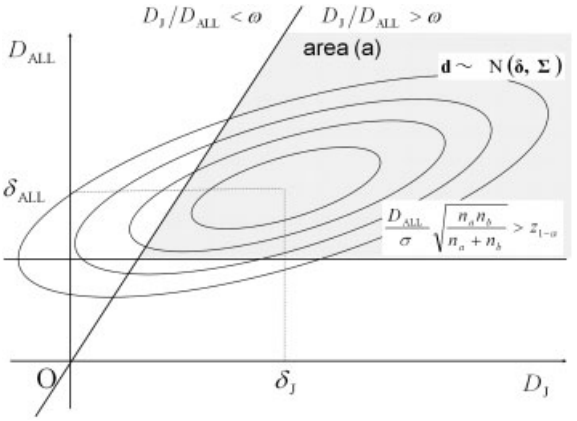

Figure 1. Joint distribution of D_{ALL} and D_{J}.

case that the treatment effect is the same in all regions, i.e. $\delta_{\mathrm{ALL}} = \delta_{\mathrm{J}} = \delta_{\mathrm{NJ}} = \delta$. First, the probability of detecting a statistically significant difference between groups a and b in all patients is

$$\eta_A = \Pr\left(\frac{D_{\mathrm{ALL}}}{\sigma}\sqrt{\frac{n_a n_b}{n_a+n_b}} > z_{1-\alpha}\right)$$

Typically, η_{A} is required to achieve a power of $1-\beta$.

Second, the probability of showing a consistent result between all and Japanese patients is considered. Although the guidance [7] recommends $D_{\mathrm{J}}/D_{\mathrm{ALL}} > \omega$ as a criterion of consistency, we can consider adding $D_{\mathrm{ALL}} > 0$ to the criterion to focus on the case that the effect of the active drug is better than the effect of placebo. Then, the probability of showing a consistent result is

$$\begin{aligned}\eta_B &= \Pr(D_{\mathrm{J}}/D_{\mathrm{ALL}} > \omega \quad \text{and} \quad D_{\mathrm{ALL}} > 0) \\ &= \frac{1}{2\pi|\boldsymbol{R}|^{\frac{1}{2}}}\int_{l_1}^{\infty}\int_{l_2}^{\infty}\exp\left(-\frac{1}{2}\boldsymbol{x}^t\boldsymbol{R}^{-1}\boldsymbol{x}\right)\mathrm{d}x_2\mathrm{d}x_1\end{aligned} \tag{1}$$

where

$$l_1 = -\frac{\delta}{\sigma}\sqrt{\frac{n_a n_b}{n_a+n_b}}$$

$$l_2 = -\frac{\delta(1-\omega)\sqrt{n_a n_b p}}{\sigma\sqrt{(n_a+n_b)(p\omega^2 - 2p\omega + 1)}}$$

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{R})$$

DOI: 10.1002/pst

Pharmaceutical STATISTICS

and

$$\boldsymbol{R} = \begin{pmatrix} 1 & \frac{\sqrt{p}(1-\omega)}{\sqrt{p\omega^2-2p\omega+1}} \\ \frac{\sqrt{p}(1-\omega)}{\sqrt{p\omega^2-2p\omega+1}} & 1 \end{pmatrix}$$

Alternatively, we can approximate $\Pr(D_J/D_{ALL} > \omega)$ by $\Pr(D_J > \omega D_{ALL})$ because the probability of $D_{ALL} < 0$ is very low when $\alpha = 0.025$ and $1-\beta \geqslant 0.8$. In this case, (1) can be approximated by

$$\begin{aligned} \eta_B &= \Pr(D_J/D_{ALL} > \omega) \\ &\approx \Pr(D_J > \omega D_{ALL}) \\ &= \Phi\left(\frac{\delta(1-\omega)\sqrt{n_a n_b p}}{\sigma\sqrt{(n_a+n_b)(p\omega^2-2p\omega+1)}}\right) \end{aligned} \tag{2}$$

where $\Phi(\cdot)$ denotes the cumulative density function of $\mathcal{N}(0,1)$. In the Appendix A, we provide the derivations of the expressions (1) and (2), see also [8].

Third, the joint probability of detecting a significant difference between the study drug and placebo in all patients and showing a consistent result between all and Japanese patients is

$$\begin{aligned} \eta_C &= \Pr\left(\frac{D_J}{D_{ALL}} > \omega \text{ and } \frac{D_{ALL}}{\sigma}\sqrt{\frac{n_a n_b}{n_a+n_b}} > z_{1-\alpha}\right) \\ &= \frac{1}{2\pi|\boldsymbol{R}|^{\frac{1}{2}}}\int_{m_1}^{\infty}\int_{m_2}^{\infty}\exp\left(-\frac{1}{2}\boldsymbol{x}^t\boldsymbol{R}^{-1}\boldsymbol{x}\right)\mathrm{d}x_2\mathrm{d}x_1 \end{aligned} \tag{3}$$

where

$$m_1 = z_{1-\alpha} - \frac{\delta}{\sigma}\sqrt{\frac{n_a n_b}{n_a+n_b}}$$

and

$$m_2 = -\frac{\delta(1-\omega)\sqrt{n_a n_b p}}{\sigma\sqrt{(n_a+n_b)(p\omega^2-2p\omega+1)}}$$

A similar expression using single integration was derived by [8,9]. The probability η_C is particularly relevant because it defines the probability of achieving two objectives according to Method 1 in [7]. Note that the expressions for η_B and η_C can be simplified when using the common sample size formula $(\delta/\sigma)\sqrt{n_a n_b/(n_a+n_b)} = z_{1-\alpha}+z_{1-\beta}$. The expressions (1) and (3) can be calculated using efficient numerical integrations methods [11]. Note that η_B can be calculated using (1) or (2) even if the variance σ^2 is unknown. In addition, the probabilities η_A and η_C can be calculated for unknown variances by calculating univariate and multivariate t probabilities [11].

To illustrate the previous expressions, we calculated η_B and η_C as a function of p. In Figure 2, we display the results for $1-\beta = 0.9$ (left plot) and $1-\beta = 0.8$ (right plot), respectively, where $\alpha = 0.025$ and $\omega = 0.5$. For $1-\beta = 0.9$, we conclude that the proportion of Japanese patients in the global study needs to be at least 22.5% (42.6%) to assure that the probability η_B of a consistency result is at least 80% (90%). For $1-\beta = 0.8$, the proportion of Japanese patient needs to be at least 28.6% (51.7%) to assure at least $\eta_B = 0.8$ ($\eta_B = 0.9$).

Fourth, we calculate the probability of obtaining a consistent result between Japanese and all patients conditional on detecting a significant difference between the study drug and placebo in all patients. This conditional probability is

$$\begin{aligned} \eta_D &= \Pr\left(\frac{D_J}{D_{ALL}} > \omega \middle| \frac{D_{ALL}}{\sigma}\sqrt{\frac{n_a n_b}{n_a+n_b}} > z_{1-\alpha}\right) \\ &= \eta_C/\eta_A \end{aligned}$$

For $1-\beta = 0.9$, $\alpha = 0.025$ and $\omega = 0.5$, the proportion of Japanese patients needs to be at least 20% to assure that the probability η_D is at least 80%. For $1-\beta = 0.8$, the proportion of Japanese patients needs to be at least 22.9%. These proportions are lower when compared with those calculated using η_B.

2.2. Required proportion of Japanese patients

Recall that according to Method 1 in [7], the number of Japanese patients needs to be sufficiently large to ensure that $D_J/D_{ALL} > \omega$ with probability of at least 80%. Re-arranging expression (2) for η_B, the required proportion of Japanese patients can be approximated by solving

DOI: 10.1002/pst

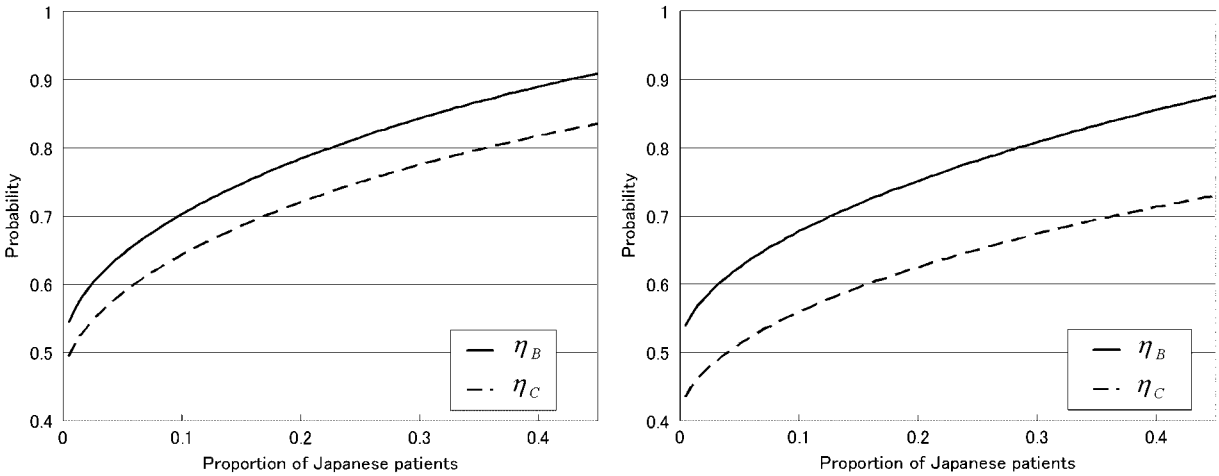

Figure 2. Probability of observing a consistent result (η_B and η_C).

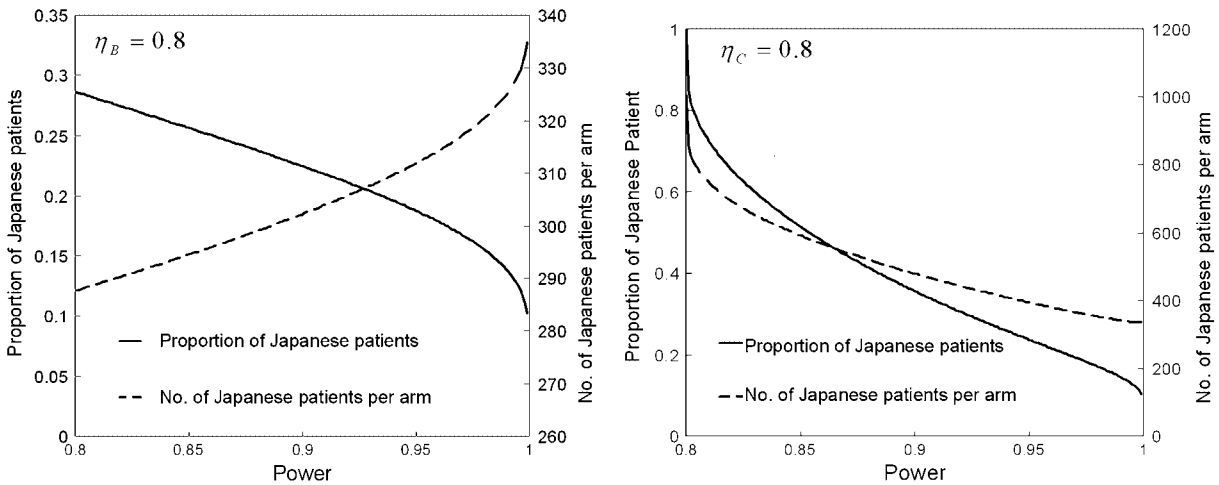

Figure 3. Proportion and number of Japanese patients (η_B and η_C).

$\Pr(D_J > \omega D_{ALL}) = 1-\gamma$, $0<\gamma<1$, for p. That is,

$$p \approx \frac{z_{1-\gamma}^2\sigma^2(n_a+n_b)}{\delta^2(1-\omega)^2 n_a n_b - z_{1-\gamma}^2\sigma^2\omega(n_a+n_b)(\omega-2)} = \frac{z_{1-\gamma}^2}{(1-\omega)^2(z_{1-\alpha}+z_{1-\beta})^2 - z_{1-\gamma}^2\omega(\omega-2)} \quad (4)$$

The previous expression again simplifies by using the sample size formula, assuming that the variance σ^2 is known, see also [8]. To illustrate the last result, we calculate p as a function of $1-\beta$. The left plot in Figure 3 displays the proportion and number of Japanese patients for $\alpha = 0.025$ and $\omega = 0.5$ so that $\eta_B = 0.8$. The number of Japanese patients is calculated when the effect size δ/σ is 0.125. With increasing power, the required proportion p to show a consistent result decreases. The absolute number of Japanese patients increases with increasing power, because of the increasing total number of patients.

DOI: 10.1002/pst

Pharmaceutical STATISTICS

Similarly, we can determine the number of Japanese patients so that $\eta_C = 0.8$. The right plot in Figure 3 displays proportion and number of Japanese patients for $\alpha = 0.025$ and $\omega = 0.5$. With increasing power, the required proportion p decreases.

2.3. Notable property

In this section we consider a notable property of Method 1. Consider the shaded area (b) in Figure 4. This area satisfies the conditions $(D_{\mathrm{ALL}}/\sigma)\sqrt{n_a n_b/(n_a+n_b)} > z_{1-\alpha}$ and $D_{\mathrm{J}}/D_{\mathrm{ALL}} > \omega$. That is, observed differences D_{ALL} and D_{J} falling into this area would indicate a consistent result when using Method 1 and at the same time lead to a statistically significant difference between groups a and b. Consider further the shaded area (c) in Figure 4, which satisfies the condition $(D_{\mathrm{ALL}}/\sigma)\sqrt{n_a n_b/(n_a+n_b)} > z_{1-\alpha}$, but not $D_{\mathrm{J}}/D_{\mathrm{ALL}} > \omega$. Observed differences D_{ALL} and D_{J} falling into this area would fail to show a consistent result according to [7], although statistical significance has been achieved for the entire population. With larger values of η_B (at least 80% or even larger), though, the probability of observing area (c) may be small. Note, however, that any result falling into area (c) is uniformly better than any result falling into area (b), as by construction both D_{ALL} and D_{J} are larger in area (c) than in area (b). This is an undesirable property of Method 1, as despite observing better results in both the entire population and the Japanese subpopulation, consistency cannot be claimed with Method 1. This observation motivates us to suggest an alternative to Method 1 in the next section.

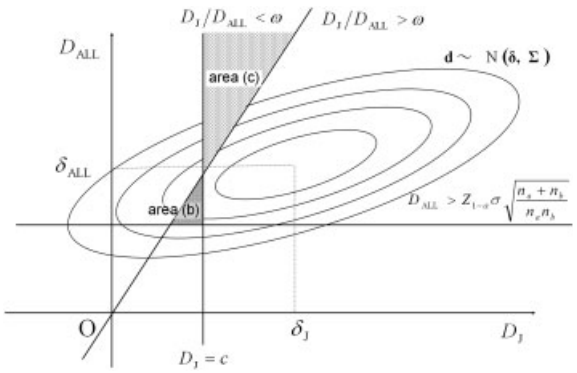

Figure 4. The property of Method 1.

3. NEW APPROACH

Motivated by the property described in Section 2.3, we now investigate a modification of Method 1. We describe the suggested new method and derive closed form expressions for the probability of observing a consistent result. Numerical comparisons with Method 1 are considered in Section 4.

3.1. Description

As seen from Figure 4, the geometric shape associated with Method 1 is the key reason for the property described in Section 2.3. As long as the region determining a consistency result is not of rectangular shape and parallel to the coordinate axes, this property will remain. We therefore investigate the alternative requirement $D_{\mathrm{J}} > c$ for a suitably chosen threshold c. There are many ways of choosing the threshold c. One possibility is to use a hypothesis test for comparing groups a and b within Japanese patients. The null hypothesis H_{J}: $\delta_{\mathrm{J}} \leqslant 0$ of no beneficial effect of the active drug over placebo is tested against the alternative K_{J}: $\delta_{\mathrm{J}} > 0$. If the associated p-value is less than or equal to a threshold on the probability scale, it is claimed that the study drug leads to consistent results between Japanese patients and the entire study population. The number of Japanese patients to be included in a multi-regional trial is then determined by the requirement to ensure a consistent result (according to the new approach) with a probability of at least 80% (say) under the assumption that the study drug has a same positive effect both in Japanese patients and entire study population, as required by the Methods 1 and 2 in [7].

Under the above normality assumptions and known variance σ^2, the test statistic for comparing groups a and b within Japanese patients is $z = (D_{\mathrm{J}}/\sigma)\sqrt{n_a n_b p/(n_a+n_b)}$. If z is larger than a

DOI: 10.1002/pst

threshold $z_{1-\phi}$ for a given value ϕ, consistency is claimed. Accordingly, this approach uses the criterion $D_J > c$, where $c = z_{1-\phi}\sigma\sqrt{(n_a+n_b)/n_a n_b p}$. If the variance σ^2 is unknown, a t-test can be used.

Moreover, this approach can be applied to various types of data, such as a normal or non-normal continuous, categorical, or censored data, because it does not depend on the study endpoints, type of tests, etc.

3.2. Probability of observing a consistent result

We now investigate the calculation of relevant success probabilities for the new approach in a similar way as in Section 2.1. We consider here only the case that the treatment effect is the same in all regions, i.e. $\delta_{\mathrm{ALL}} = \delta_{\mathrm{J}} = \delta_{\mathrm{NJ}} = \delta$. The probability η_A remains unaffected by the new approach. However, the probability of showing a consistent result between all patients and Japanese patients becomes

$$\begin{aligned}\eta'_B &= \Pr\left(\frac{D_{\mathrm{J}}}{\sigma}\sqrt{\frac{n_a n_b p}{n_a+n_b}} > z_{1-\phi}\right)\\ &= \Phi((z_{1-\alpha}+z_{1-\beta})\sqrt{p} - z_{1-\phi})\end{aligned}$$

Consequently, the necessary proportion of Japanese patients to show a consistent result with probability of at least $100(1-\gamma)\%$ for some $0<\gamma<1$ is given by

$$p \geqslant \frac{(z_{1-\phi}+z_{1-\gamma})^2}{(z_{1-\alpha}+z_{1-\beta})^2}$$

Note that for a fixed value ϕ, the required number of Japanese patients remains unchanged regardless of the power.

The joint probability of detecting a significant difference between groups a and b in all patients and showing a consistent result between all patients and Japanese patients for the new approach becomes

$$\begin{aligned}\eta'_C &= \Pr\left(\frac{D_{\mathrm{ALL}}}{\sigma}\sqrt{\frac{n_a n_b}{n_a+n_b}} > z_{1-\alpha} \quad \text{and} \quad \frac{D_{\mathrm{J}}}{\sigma}\sqrt{\frac{n_a n_b p}{n_a+n_b}} > z_{1-\phi}\right)\\ &= \frac{1}{2\pi|\boldsymbol{R}_1|^{\frac{1}{2}}}\int_{-z_{1-\beta}}^{\infty}\int_{l}^{\infty}\exp\left(-\frac{1}{2}\boldsymbol{v}'\boldsymbol{R}_1^{-1}\boldsymbol{v}\right)dv_2 dv_1\end{aligned}$$

where $l = z_{1-\beta} - (z_{1-\alpha}+z_{1-\beta})\sqrt{p}$, $\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \sim \mathcal{N}(0, \boldsymbol{R}_1)$ and

$$R_1 = \begin{pmatrix} 1 & \sqrt{p} \\ \sqrt{p} & 1 \end{pmatrix}$$

The probability of obtaining a consistent result between Japanese and all patients for the new approach under the condition of detecting a significant difference between the study drug and placebo in all patients becomes

$$\eta'_D = \Pr\left(\frac{D_{\mathrm{J}}}{\sigma}\sqrt{\frac{n_a n_b p}{n_a+n_b}} > z_{1-\phi} \middle| \frac{D_{\mathrm{ALL}}}{\sigma}\sqrt{\frac{n_a n_b}{n_a+n_b}} > z_{1-\alpha}\right) = \eta'_C/\eta_A$$

3.3. Choice of ϕ

Selecting ϕ is of critical importance for the new approach. The choice of ϕ has a similar impact on the operating characteristics of the new approach as the choice of ω has it for Method 1. In the following we give some suggested guidance on how to link the choice of ϕ to a specific value of ω in order to ensure the same level of strictness of Method 1. The guidance document [7] recommends using $\omega \geqslant 0.5$. To find the corresponding value for ϕ, we consider Method 1 for $\alpha = 0.025$ and $\omega = 0.5$. Using (1), we need 22.5–28.6% Japanese patients in the entire study population to demonstrate a consistent result with probability of at least 80% for Method 1 to achieve a power of $0.8 \leqslant 1-\beta \leqslant 0.9$. Setting $\phi = 0.25$ in the new approach requires 21.9–29.3% Japanese patients to ensure a power of 80–90%. Therefore, setting $\phi = 0.25$ in the new approach ensures a similar level of strictness as Method 1 with $\omega = 0.5$.

4. NUMERICAL COMPARISON STUDY

In this section, we report the results of a numerical study to compare the new approach from Section 3 with Method 1 from [7]. We investigate the operating characteristics for both methods, such as different probabilities of success and the false-positive error rate, under different scenarios.

DOI: 10.1002/pst

Pharmaceutical STATISTICS

4.1. Results for probabilities of showing a consistent result

In this study, we compare the performance of both the methods with respect to the four probabilities introduced in Sections 2.1 and 3.2. For the different scenarios used in the simulations, the total sample size n per arm is calculated such that the study has a power of $1-\beta$ for selected values of β to detect a significant difference at a level of $\alpha = 0.025$ in all patients. The number of Japanese patients n_J per arm is calculated such that $\Pr(D_J/D_{ALL} > \omega$ and $D_{ALL} > 0)$ is more than 80%. We set $\omega = 0.5$ for Method 1 and $\phi = 0.25$ for the new approach. In addition, we consider the scenarios that the study drug has an effect size of 0.125 or 0.25. Thus, without loss of generality, we set $\delta_{ALL} = \delta_J = \delta_{NJ} = 1$ and consider $\sigma = 4,8$. Larger probabilities of showing a consistent result indicate better performance.

Table I presents the results of 1,000,000 simulated trials per scenario. We conclude from Table I that η_B and η'_B are very similar for all power values under consideration. Consequently, the level of strictness for the new approach with $\phi = 0.25$ is about the same as for Method 1 with $\omega = 0.5$. However, η'_C and η'_D are consistently larger than η_C and η_D, respectively. That is, the probability to achieve the objectives for the entire study population and the Japanese subpopulation is larger for the new approach than for Method 1.

4.2. Results for false-positive error rates

We now investigate the false-positive error rates under different scenarios, i.e. we consider the probability for a method to declare a consistent result despite the fact that the study drug has no effect in either the Japanese patients or the entire study population. Then, η_B and η'_B are calculated as the false-positive error rates for the new approach and Method 1, respectively. For the calculation of η_B, $\Pr(D_J/D_{ALL} > \omega$ and $D_{ALL} > 0)$ is used. Larger false-positive error rates indicate worse performance. For the calculations, we use the same total sample sizes n as considered in Section 4.1. We assume the scenarios $\delta_{ALL} = \delta_J = \delta_{NJ} = 0$ and $\delta_{NJ} = 1$ for $\sigma = 4,8$. Under these assumptions, the study drug has no effect in at least one of the populations and consistency should not be declared.

Table II presents the results for the different scenarios. First, we consider the case that the study drug has no effect in the entire study population, in particular no effect for the Japanese patients, i.e. $\delta_{ALL} = \delta_J = \delta_{NJ} = 0$. We conclude from the upper half of Table II that the false-positive error rates of the new approach are lower than the ones of Method 1 for the given nominal power values of 0.8, 0.9 and 0.95. Note that the difference in error rate increases for smaller nominal power values. It can also be seen that the new approach keeps the error rate of 0.25. The same conclusions hold when considering the case that the study drug has an

Table I. Probability of obtaining a consistent result (in %).

	Power (n, n_J)	Method	η_A	η_B, η'_B	η_C, η'_C	η_D, η'_D
$\sigma = 8$	0.80 (1006, 288)	New approach	79.89	79.29	68.61	85.88
		Method 1	79.89	79.92	66.58	83.34
	0.90 (1346, 303)	New approach	89.93	80.57	75.40	83.84
		Method 1	89.93	80.02	73.47	81.70
	0.95 (1665, 312)	New approach	95.00	81.50	78.99	83.16
		Method 1	95.00	80.21	76.97	81.02
$\sigma = 4$	0.80 (253,73)	New approach	80.15	79.53	68.91	85.98
		Method 1	80.15	79.96	66.82	83.36
	0.90 (338,76)	New approach	90.05	80.70	75.57	83.92
		Method 1	90.05	80.13	73.63	81.77
	0.95 (417,78)	New approach	95.02	81.18	78.70	82.82
		Method 1	95.02	80.00	76.81	80.84

DOI: 10.1002/pst

Table II. Results for false-positive error rates (in %).

	$\sigma = 8$		$\sigma = 4$	
Nominal power	New approach	Method 1	New approach	Method 1
$\delta_J = \delta_{NJ} = 0$				
0.80	25.00	29.88	25.00	29.91
0.90	25.00	29.19	25.00	29.19
0.95	25.00	28.75	25.00	28.75
$\delta_J = 0$, $\delta_{NJ} = 1$				
0.80	25.00	27.10	25.00	27.01
0.90	25.00	25.62	25.00	25.58
0.95	25.00	24.68	25.00	24.67

effect in the total study population, but no effect for Japanese patients, i.e. $\delta_J = 0$, $\delta_{NJ} = 1$, see the bottom half Table II for details. In summary, the numerical comparison study suggests that the new approach has better operating characteristics than Method 1, as it leads to (*i*) smaller false-positive error rates, and (*ii*) larger probabilities of showing consistent results.

5. DISCUSSION

One of the main objectives of multi-regional trials is to confirm the effectiveness in a particular region. In this paper, we focused on Method 1 from [7] and derived closed form expressions for the resulting probabilities, which require the evaluation of multivariate normal or t probabilities. In addition, we proposed an alternative method, which has better operating characteristics than current approaches. The numerical results indicate that the new approach using a hypothesis test within Japanese patients has a better performance than Method 1. That is, it provides higher probabilities to achieve statistical significance in all patients and consistency between Japanese and all patients, when the study drug is effective in both Japanese and all patients. Moreover, the false-positive error rates of the new approach are comparable or even lower than those of Method 1 when a study drug has no effect in either Japanese or all patients.

In this paper, we set $\phi = 0.25$ so that Japanese sample sizes based on the new method are similar to those obtained with Method 1 using $\omega = 0.5$. Other choices of ϕ are possible and need to be investigated in future. One may argue that setting the significance level at $\phi = 0.25$ is too large compared with the significance level of 0.025 we generally use. Consequently, $\phi < 0.25$ should be preferred. However, setting a lower ϕ value would lead to large sample size requirements for the region, thereby reducing the benefit of conducting a multi-regional trial. Alternatively, selecting $\phi = 0.5$ means that we have to show $D_J > 0$, which may lack the persuasiveness of effectiveness in the region. These significance levels would relate to the 'hierarchy of persuasiveness' stated in the 11th Q&A for the ICH E5 guideline [1].

The problem of showing effectiveness in multi-regional trials remains a difficult problem. In this paper, we considered necessary sample sizes to demonstrate the effectiveness for Japanese patients. In addition, we would need to compare the results across the regions as well. In case that differences in treatment effect exist across regions such that heterogeneity of the treatment effect across regions or large treatment-by-region interaction effect may exist, we need to investigate the reason for the difference and show that they do not have an impact on the overall results. It would be important to consider effectiveness in the particular region and differences of the treatment effect across regions.

REFERENCES

1. International Conference on Harmonization. *Q&A for the ICH-E5 Guideline on 'Ethnic Factors in the Acceptability of Foreign Clinical Data'*, 2006. Available at www.ich.org/cache/compo/475-272-1.html.
2. Anello C, O'Neil RT, Dubey S. Multicentre trials: a US regulatory perspective. *Statistical Methods in Medical Research* 2005; **14**:303–318.
3. Lewis JA. Statistical issues in the regulation of medicine. *Statistics in Medicine* 1995; **14**:127–136.
4. US Food and Drug Administration. *Guideline on the 'Format and Content of the Clinical and Statistical Sections of New Drug Applications'*, 1988.
5. International Conference on Harmonization. *The ICH E9 Guideline on Statistical Principles for Clinical Trials*, 1998. Available at www.ich.org/cache/compo/475-272-1.html.

DOI: 10.1002/pst

Pharmaceutical STATISTICS

6. Shao J, Chow SC. Two-stage sampling with pharmaceutical applications. *Statistics in Medicine* 1993; **21**: 1727–1742.
7. Ministry of Health, Labour and Welfare. *Basic Principles on Global Clinical Trials*, 2007.
8. Quan H, Zhao P-L, Zhang J, Roessner M, Aizawa K. Sample size considerations for Japanese patients in a multi-regional trial based on MHLW Guidance. *Pharmaceutical Statistics* 2010; **9**(2):100–112.
9. Uesaka H. Sample size allocation to regions in a multiregional trial. *Journal of Biopharmaceutical Statistics* 2009; **19**:580–594.
10. Kawai N, Chuang-Stein C, Komiyama O, Ii Y. An approach to rationalize partitioning sample size into individual regions in a multiregional trial. *Drug Infomation Journal* 2008; **42**:139–147.
11. Genz A, Bretz F. *Computation of Multivariate Normal and t Probabilities*. Springer: Heidelberg, 2009.

APPENDIX A

Equation (1) is derived as follows. With the notation from Section 2, we have

$$\begin{aligned}\eta_B &= \Pr(D_J/D_{ALL}>\omega \text{ and } D_{ALL}>0)\\ &= \frac{1}{2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\int_0^\infty\int_{\omega D_{ALL}}^\infty \exp\left(-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{\delta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{d}-\boldsymbol{\delta})\right)\\ &\quad\times \mathrm{d}D_J\mathrm{d}D_{ALL}\\ &= \frac{1}{2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\int_0^\infty\int_{\omega D_{ALL}}^\infty \exp\left(-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{\delta})^t\boldsymbol{C}^t\boldsymbol{C}^{-t}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}^{-1}\boldsymbol{C}(\boldsymbol{d}-\boldsymbol{\delta})\right)\\ &\quad\times \mathrm{d}D_J\mathrm{d}D_{ALL}\\ &= \frac{1}{2\pi|\boldsymbol{R}|^{\frac{1}{2}}}\int_{l_1}^\infty\int_{l_2}^\infty \exp\left(-\frac{1}{2}\boldsymbol{x}^t\boldsymbol{R}^{-1}\boldsymbol{x}\right)\mathrm{d}x_2\mathrm{d}x_1\end{aligned}$$

where

$$\boldsymbol{C}=\boldsymbol{S}_2\boldsymbol{S}_1=\sqrt{\frac{n_an_b}{(n_a+n_b)\sigma^2}}\begin{pmatrix}1 & 0\\ -\frac{\omega\sqrt{p}}{\sqrt{p\omega^2-2p\omega+1}} & \frac{\sqrt{p}}{\sqrt{p\omega^2-2p\omega+1}}\end{pmatrix},$$

$$\boldsymbol{S}_1=\begin{pmatrix}1 & 0\\ -\omega & 1\end{pmatrix},$$

$$\boldsymbol{S}_2=\sqrt{\frac{n_an_b}{(n_a+n_b)\sigma^2}}\begin{pmatrix}1 & 0\\ 0 & \frac{\sqrt{p}}{\sqrt{p\omega^2-2p\omega+1}}\end{pmatrix}$$

$\boldsymbol{x}=\boldsymbol{C}(\boldsymbol{d}-\boldsymbol{\delta})$ and $\boldsymbol{R}=\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^t$. Note that S_1 is a linear transformation of the integration region to a rectangular shape and S_2 is a linear transformation to adjust the scale so that the variances of x are 1. Similarly, Equation (2) follows from

$$\begin{aligned}\eta_B &= \Pr(D_J/D_{ALL}>\omega) \approx \Pr(D_J>\omega D_{ALL})\\ &= \frac{1}{2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\int_{-\infty}^\infty\int_{\omega D_{ALL}}^\infty \exp\left(-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{\delta})^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{d}-\boldsymbol{\delta})\right)\\ &\quad\times \mathrm{d}D_J\mathrm{d}D_{ALL}\\ &= \frac{1}{2\pi|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\int_{-\infty}^\infty\int_{\omega D_{ALL}}^\infty \exp\left(-\frac{1}{2}(\boldsymbol{d}-\boldsymbol{\delta})^t\boldsymbol{C}^t\boldsymbol{C}^{-t}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}^{-1}\boldsymbol{C}(\boldsymbol{d}-\boldsymbol{\delta})\right)\\ &\quad\times \mathrm{d}D_J\mathrm{d}D_{ALL}\\ &= \frac{1}{2\pi|\boldsymbol{R}|^{\frac{1}{2}}}\int_{-\infty}^\infty\int_{l_2}^\infty \exp\left(-\frac{1}{2}\boldsymbol{x}^t\boldsymbol{R}^{-1}\boldsymbol{x}\right)\mathrm{d}x_2\mathrm{d}x_1\\ &= \Phi\left(\frac{\delta(1-\omega)\sqrt{n_an_bp}}{\sigma\sqrt{(n_a+n_b)(p\omega^2-2p\omega+1)}}\right).\end{aligned}$$

DOI: 10.1002/pst