A Re-examination of Japanese Sample Size Calculation for Multi-regional Clinical Trial Evaluating Survival Endpoint

Nobuya Hayashi and Yohji Itoh
Biometrics Department, AstraZeneca K.K.
3-1, Ofuka-cho, Kita-ku, Osaka 530-0011, Japan
e-mail: nobuya.hayashi@astrazeneca.com

In planning a multi-regional clinical trial including Japan, the sample size for Japanese patients is often considered based on the probability of obtaining a consistent result between Japanese subpopulation and the overall study population, as recommended in the Japanese guideline “Basic Principles on Global Clinical Trials.” We review the commonly used existing method for Japanese sample size calculation based on obtaining a consistent result for survival endpoint. Through simulations, we note that Japanese sample size based on the existing method tended to have less actual power than the nominal power for consistency, especially when there is a large treatment effect. We propose alternative methods based on the delta method and numerical integrations. Our proposed methods give similar Japanese sample sizes, and our simulation studies show that our proposed methods provide the actual power close to the nominal power for consistency.

Key words: consistency, hazard ratio, delta method, numerical integration.

1. Introduction

When planning for a multi-regional clinical trial including Japan, the sample size for Japanese patients is often considered based on two examples provided in the guideline ‘Basic Principles on Global Clinical Trials’ issued by the Japanese Ministry of Health, Labour and Welfare (2007) that requires us to show “consistency” between results from the entire population and the Japanese population. In particular, Method 1 recommends setting sample size for Japanese patients in a multi-regional clinical trial to satisfy

\[ P \left( \frac{\hat{d}_{JP}}{\hat{d}_{All}} > \pi \right) \geq 1 - \beta' \]  

(1)

where \( \hat{d} \) is an estimated difference in an efficacy endpoint between test treatment and placebo groups, \( \pi \) is a criterion for consistency typically set at 0.5, \( 1 - \beta' \) is desired probability to satisfy the criterion typically set to at least 0.8 and subscripts ‘JP’ and ‘All’ indicate Japanese subpopu-
lation and whole study population, respectively. Hereafter, we refer the probability expressed by the left-hand side of (1) or its kindred probability as "consistency probability" for convenience.

For trials evaluating survival endpoint, hazard ratio is often used as an indicator of treatment effect. In such a study, a log transformed hazard ratio may be used as \( \hat{d} \), and formulations for normal endpoint case described in Quan et al (2010) can be applied to calculate the probability to satisfy the criterion. Alternatively, risk reduction (i.e. one minus hazard ratio) can be used as \( \hat{d} \), which is more clinically interpretable indicator:

\[
P\left( \frac{1 - \exp(\hat{\gamma}_{\text{JP}})}{1 - \exp(\hat{\gamma}_{\text{All}})} > \pi \right) = 1 - \beta'
\]

where \( \hat{\gamma}_{\text{All}} \) is the estimate of log hazard ratio in the whole study population and \( \hat{\gamma}_{\text{JP}} \) is the estimate of log hazard ratio in Japanese subpopulation. The consistency probability expressed by the left-hand side of (2) will be the main object of discussion in this paper. When planning for a multi-regional study, the joint probability of statistically significant overall study result and consistency is often also of interest. We will address this point as well in this paper.

Quan et al (2010) provided useful formulas for Japanese sample size calculation in survival endpoint case based on (2). However, their recommended formula gives Japanese sample size with the actual power slightly less than the nominal power, as noted in simulation studies by Quan et al (2010). In our simulations studies (Section 5), we also observed that the estimated consistency probability based on (6) are approximately 2 to 3% smaller than the nominal value. In Section 2, we review the existing method by Quan et al (2010). Then, in the subsequent sections, we propose two ways to obtain more accurate consistency probabilities along with methods to compute the joint probability of statistically significant overall study result and consistency. In Section 3, we propose a new formula based on alternative approximation of distribution for the statistics used in consistency criteria. We propose another approach based on numerical integration in Section 4. In Section 5, we compare our proposed methods with method by Quan et al (2010) through simulations in terms of actual power versus nominal power for consistency. We expand on the discussion of two possible approaches to apply Method 1 for survival endpoint case and their implications in practice in Section 6.

2. Review of Existing Method

Unless otherwise noted, we will assume the following in this paper. We will assume \( \hat{\gamma}_{\text{All}} = w\hat{\gamma}_{\text{JP}} + (1 - w)\hat{\gamma}_{\text{NJ}} \) where \( w \) is the fraction of events in Japanese subjects among events in the whole study population, and \( \hat{\gamma}_{\text{NJ}} \) is the estimator of the hazard ratio for non-Japanese. We will further assume that the estimators of log hazard ratios are normally distributed:

\[
\hat{\gamma}_{\text{JP}} \sim N\left( \gamma_{\text{JP}}, \frac{1}{w\theta(1 - \theta)E} \right),
\]

\[
\hat{\gamma}_{\text{NJ}} \sim N\left( \gamma_{\text{NJ}}, \frac{1}{(1 - w)\theta(1 - \theta)E} \right),
\]
\[ \hat{\gamma}_{\text{All}} \sim N \left( \gamma_{\text{All}}, \frac{1}{\theta(1-\theta)E} \right) , \]  

for \( \gamma_{\text{JP}} \) and \( \gamma_{\text{NJ}} \) are true log hazard ratio for Japanese and non-Japanese population, respectively, \( \gamma_{\text{All}} = w\gamma_{\text{JP}} + (1-w)\gamma_{\text{NJ}}, \) \( \theta \) is the probability that a patient is allocated to test treatment, e.g. \( \theta = 0.5 \) for 1 : 1 allocation, and \( E \) is the expected total number of events.

Based on the asymptotic normal distribution of \( \exp(\hat{\gamma}_{\text{JP}}) - \pi \exp(\hat{\gamma}_{\text{All}}) \) in hazard ratio scale derived by applying the delta method, Quan et al (2010) derived the following approximation to the consistency probability (2):

\[ P \left( \frac{1 - \exp(\hat{\gamma}_{\text{JP}})}{1 - \exp(\hat{\gamma}_{\text{All}})} > \pi \right) \approx \Phi(z_{\text{Quan}}) \]  

where

\[ z_{\text{Quan}} = \frac{1 - \pi + \pi \exp(\hat{\gamma}_{\text{All}}) - \exp(\hat{\gamma}_{\text{JP}})}{\exp(\hat{\gamma}_{\text{JP}}) \sqrt{\left[ \pi^2 \exp(2(\gamma_{\text{All}} - \gamma_{\text{JP}})) \right] - 2\pi \exp(\gamma_{\text{All}} - \gamma_{\text{JP}}) + \frac{1}{w} \frac{1}{\theta(1-\theta)E} \left( 1 - \exp(\gamma) \right)^2}} \]

Given the desired probability of consistency \( 1 - \beta' \), if we assume \( \gamma_{\text{All}} = \gamma_{\text{JP}} = \gamma_{\text{NJ}} = \gamma \), we can solve (7) for \( w \) with \( z_{\text{Quan}} = z_{1-\beta'} \) to obtain the required fraction of events from Japanese patients, \( w_{\text{Quan}} \) (6) in Quan et al, 2010):

\[ w_{\text{Quan}} = \frac{\exp(2\gamma)z_{1-\beta'}^2}{\theta(1-\theta)E(1-\pi)^2(1-\exp(\gamma))^2 + \exp(2\gamma)(2\pi - \pi^2)z_{1-\beta'}^2} . \]

In their simulation studies, Quan et al (2010) noted that the estimates of consistency probability using the required number of events for Japanese population calculated based on \( w_{\text{Quan}} \) are 'slightly smaller than the nominal value.' In another word, (6) overestimates the consistency probability. In our simulations studies (Section 5), we also observed that the estimated consistency probability using the required number of events in Japanese population calculated based on \( w_{\text{Quan}} \) is approximately 2 to 3% smaller than the nominal value.

Note that when \( \pi = 0 \), (6) provides an approximation to the probability of Japanese population result having the same direction as the overall population result, however, it can be shown analytically that (6) overestimates the true probability under the assumption (3). (See Appendix A.)

These observations suggest that it is desirable to have an approximation that provides closer to true value, and when \( \pi = 0 \), reduces to the correct expression for the probability of directional consistency (A.1). To this end, an alternative approximation method is presented in the next section.

3. Alternative approximation formula

The \( z \)-value (7) was derived based on asymptotic normal distribution of \( \exp(\hat{\gamma}_{\text{JP}}) - \pi \exp(\hat{\gamma}_{\text{All}}) \) in hazard ratio scale using the delta method (Quan et al, 2010). Since log hazard ratios are assumed to be normally distributed, it is natural to consider asymptotic normal distribution in log
Here, we propose an alternative delta method approach based on asymptotic normal distribution in log(hazard ratio) scale.

Since \( P(\hat{\gamma}_{\text{All}} > 0) \) is very close to zero (Quan et al, 2010), we have

\[ P \left( \frac{1 - \exp(\hat{\gamma}_{\text{JP}})}{1 - \exp(\hat{\gamma}_{\text{All}})} > \pi \right) \approx P(\exp(\hat{\gamma}_{\text{JP}}) - \pi \exp(\hat{\gamma}_{\text{All}}) < 1 - \pi) \]

\[ = P(g(\hat{\gamma}_{\text{JP}}, \hat{\gamma}_{\text{NJ}}) < 0) \tag{8} \]

where

\[ g(\hat{\gamma}_{\text{JP}}, \hat{\gamma}_{\text{NJ}}) = \hat{\gamma}_{\text{JP}} - \log\{1 - \pi + \pi \exp(\hat{\gamma}_{\text{All}})\} \]

\[ = \hat{\gamma}_{\text{JP}} - \log[1 - \pi + \pi \exp\{w\hat{\gamma}_{\text{JP}} + (1 - w)\hat{\gamma}_{\text{NJ}}\}] \]

Applying the delta method, the asymptotic distribution of \( g(\hat{\gamma}_{\text{JP}}, \hat{\gamma}_{\text{NJ}}) \) is given by

\[ g(\hat{\gamma}_{\text{JP}}, \hat{\gamma}_{\text{NJ}}) \sim N(\mu_g, \sigma_g^2) \tag{9} \]

where \( \mu_g = g(\gamma_{\text{JP}}, \gamma_{\text{NJ}}) = \gamma_{\text{JP}} - \log\{1 - \pi + \pi \exp(\gamma_{\text{All}})\} \) and

\[ \sigma_g^2 = \frac{1}{w\theta(1 - \theta)E\left\{1 - \pi + (1 - w)\exp(\gamma_{\text{All}})\right\}^2} + \frac{1}{(1 - w)\theta(1 - \theta)E\left\{1 - \pi + \pi \exp(\gamma_{\text{All}})\right\}^2} . \]

Based on (8) and the distribution (9), we have

\[ P \left( \frac{1 - \exp(\hat{\gamma}_{\text{JP}})}{1 - \exp(\hat{\gamma}_{\text{All}})} > \pi \right) \approx P(\hat{\gamma}_{\text{JP}} - \log\{1 - \pi + \pi \exp(\hat{\gamma}_{\text{All}})\} < 0) \]

\[ = \Phi(z_{\text{All}}) \tag{10} \]

where

\[ z_{\text{All}} = -\frac{\mu_g}{\sigma_g} = -\frac{\gamma_{\text{JP}} - \log\{1 - \pi + \pi \exp(\gamma_{\text{All}})\}\{1 - \pi + \pi \exp(\gamma_{\text{All}})\}}{\sqrt{\frac{1}{\theta(1 - \theta)E\left\{1 - \pi + \pi \exp(\gamma_{\text{All}})\right\}}^2}} . \tag{11} \]

It is easy to check that when \( \pi = 0 \), (11) reduces to the expression, (A.1) in Appendix A, that gives the correct probability of \( \hat{\gamma}_{\text{JP}} < 0 \).

Given the desired probability of consistency \( 1 - \beta' \), if we assume that \( \gamma_{\text{All}} = \gamma_{\text{JP}} = \gamma_{\text{NJ}} = \gamma \), we can derive the required fraction, \( w_{\text{All}} \), of events from Japanese patients by solving (11) for \( w \) with \( z_{\text{All}} \equiv z_{1 - \beta'} \), that is,

\[ w_{\text{All}} = \left[ \theta(1 - \theta)E\cdot \frac{g(\gamma, \gamma)^2}{2_{1 - \beta'}^2} + \frac{2(1 - \pi)\pi \exp(\gamma) + \pi^2 \exp(2\gamma)}{\{1 - \pi + \pi \exp(\gamma)\}^2} \right]^{-1} . \tag{12} \]

Once we have the required total numbers of events, \( E \), and the required fraction of events from Japanese patients, the sample sizes can be calculated by dividing the number of events by

---

Quan et al (2010) provided the required fraction of Japanese events based on the asymptotic normal distribution of \( \frac{1 - \exp(\gamma_{\text{JP}})}{1 - \exp(\gamma_{\text{All}})} \). However, this method requires taking log-transformations on both sides of inequality in the consistency probability expression and when \( \pi = 0 \), the directional consistency probability calculated based on this method is always 1 which is clearly not correct. In addition, according to their simulation studies, Japanese sample size based on this method provided actual power that is substantially smaller than the nominal power.

---

Jpn J Biomet Vol. 38, No. 2, 2017
the event probability during study as described in Quan et al (2010).

When planning for a multi-regional study, the joint probability of statistically significant overall study result and consistency is often also of interest. Ikeda and Bretz (2010) proposed a method for calculating this joint probability for the case where the endpoint is normally distributed. Here, we show an approximation of the joint probability for survival endpoint with consistency criteria (2) using the delta method. The joint probability can be written as

\[
P \left( \frac{1 - \exp(\hat{\gamma}_{\text{JP}})}{1 - \exp(\hat{\gamma}_{\text{All}})} > \pi \quad \text{and} \quad \frac{\hat{\gamma}_{\text{All}}}{\sqrt{1/\{\theta(1-\theta)E\}}} < -z_{1-\alpha/2} \right) 
\]

\[
\approx P \left( \hat{\gamma}_{\text{JP}} - \log(1 - \pi + \pi \exp(\hat{\gamma}_{\text{All}})) < 0 \quad \text{and} \quad \frac{\hat{\gamma}_{\text{All}}}{\sqrt{1/\{\theta(1-\theta)E\}}} < -z_{1-\alpha/2} \right) .
\]

Applying the delta method, the asymptotic distribution of

\[
g(\hat{\gamma}_{\text{JP}}, \hat{\gamma}_{\text{NJ}}) = \frac{\partial g(\hat{\gamma}_{\text{JP}}, \hat{\gamma}_{\text{NJ}})}{\partial \hat{\gamma}_{\text{JP}}}
\]

is given by

\[
g(\hat{\gamma}_{\text{JP}}, \hat{\gamma}_{\text{NJ}}) \sim N(\mathbf{v}_g, \Sigma_g)
\]

where

\[
\mathbf{v}_g = \begin{pmatrix} \hat{\gamma}_{\text{JP}} - \log(1 - \pi + \pi \exp(\hat{\gamma}_{\text{All}})) \\ \hat{\gamma}_{\text{All}} \end{pmatrix},
\]

\[
\Sigma_g = \frac{1}{\theta(1-\theta)E} \begin{pmatrix} g_{11} & g_{12} \\ g_{12} & 1 \end{pmatrix},
\]

\[
g_{11} = \frac{1}{w} \left\{ \frac{1 - \pi + \pi(1 - w) \exp(\hat{\gamma}_{\text{All}})}{1 - \pi + \pi \exp(\hat{\gamma}_{\text{All}})} \right\}^2 + \frac{1}{1 - w} \left( \frac{\pi(1 - w) \exp(\hat{\gamma}_{\text{All}})}{1 - \pi + \pi \exp(\hat{\gamma}_{\text{All}})} \right)^2,
\]

\[
g_{12} = \frac{1 - \pi}{1 - \pi + \pi \exp(\hat{\gamma}_{\text{All}})}.
\]

Then, the joint probability is given by

\[
P \left( \frac{1 - \exp(\hat{\gamma}_{\text{JP}})}{1 - \exp(\hat{\gamma}_{\text{All}})} > \pi \quad \text{and} \quad \frac{\hat{\gamma}_{\text{All}}}{\sqrt{1/\{\theta(1-\theta)E\}}} < -z_{1-\alpha/2} \right) 
\]

\[
\approx \Phi_2 \left( z_{\text{Alt}}, -\frac{\hat{\gamma}_{\text{All}}}{\sqrt{1/\{\theta(1-\theta)E\}}} - z_{1-\alpha/2} \mid g_{12} \right)
\]

where \(\Phi_2(z_1, z_2 \mid \rho)\) represents the bi-variate standard normal distribution with correlation \(\rho\).

This joint probability can be calculated using statistical softwares such as PROBBNRM function of SAS or pnorm2d function in ‘fMultivar’ package of R. The joint probability based on \(z_{\text{Quan}}\) in place of \(z_{\text{Alt}}\) can be derived similarly using delta method, and can be expressed in similar form as (13) with \(z_{\text{Alt}}\) replaced by \(z_{\text{Quan}}\).

\[
g_{11} = \left[ \pi \exp(\hat{\gamma}_{\text{All}} - \hat{\gamma}_{\text{JP}}) \right] \left\{ \pi \exp(\hat{\gamma}_{\text{All}} - \hat{\gamma}_{\text{JP}}) - 2 \right\} + \frac{1}{w} \exp(2\hat{\gamma}_{\text{JP}})
\]

and

\[
g_{12} = 1 - \pi + \pi \exp(\hat{\gamma}_{\text{All}}) - \exp(\hat{\gamma}_{\text{JP}}).
\]

Jpn J Biomet Vol. 38, No. 2, 2017
4. Numerical integration

Using numerical integration, the exact probability for consistency can be calculated. In this section, we describe numerical integration techniques for calculating the consistency probability using R and SAS to make them have a practical use.

The consistency probability can be expressed as a double integral:

\[
P\left(\frac{1 - \exp(\hat{\gamma}_{JP})}{1 - \exp(\hat{\gamma}_{All})} > \pi\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_{JP} | \gamma_{JP}, \sigma_{JP}^2) f(x_{NJ} | \gamma_{NJ}, \sigma_{NJ}^2) I\left\{\frac{1 - \exp(x_{JP})}{1 - \exp(wx_{JP} + (1-w)x_{NJ})} > \pi\right\} dx_{JP} dx_{NJ}
\]

(14)

where \(\sigma_{JP}^2 = \frac{1}{w\theta(1-\theta)E}\), \(\sigma_{NJ}^2 = \frac{1}{(1-w)\theta(1-\theta)E}\), \(f(x | \mu, \sigma^2)\) denotes the univariate normal density function with mean \(\mu\) and variance \(\sigma^2\), and \(I\{A\}\) denotes an indicator function that is one if condition \(A\) is true and zero if \(A\) is not true. This double integration provides us the exact probability and can be executed using numerical integration technique. This calculation can be done in R by calling function ‘integrate’ twice for inner and outer integrals together with function ‘ifelse’ for defining indicator function.

While this calculation can be done successfully in R, we were not able to successfully calculate the double integral using CALL QUAD in SAS probably due to the existence of discontinuous points in the integrand of (14). This issue can be solved by transforming the expression to explicitly defining the limits of integration in (14) rather than relying on the indicator function.

The inequality \(\{1 - \exp(\hat{\gamma}_{JP})\}/\{1 - \exp(\hat{\gamma}_{All})\} > \pi\) can be transformed to be

\[
\exp\{(1-w)\hat{\gamma}_{NJ}\} > \frac{\exp(\hat{\gamma}_{JP}) + \pi - 1}{\pi \exp(w\hat{\gamma}_{JP})}.
\]

(15)

When \(\exp(\hat{\gamma}_{JP}) + \pi - 1 > 0\), that is, \(\hat{\gamma}_{JP} > \log(1-\pi)\), (15) can be transformed to be

\[
\hat{\gamma}_{NJ} > \frac{1}{1-w} \log\left\{\frac{\exp(\hat{\gamma}_{JP}) + \pi - 1}{\pi \exp(w\hat{\gamma}_{JP})}\right\}.
\]

When \(\exp(\hat{\gamma}_{JP}) + \pi - 1 \leq 0\), inequality (15) always holds irrespective of the value of \(\hat{\gamma}_{NJ}\), and thus the range of the value that \(\hat{\gamma}_{NJ}\) can take is \(-\infty < \hat{\gamma}_{NJ} < \infty\). Therefore, the region satisfying the condition \(\{1 - \exp(\hat{\gamma}_{JP})\}/\{1 - \exp(\hat{\gamma}_{All})\} > \pi\) can be depicted as shown in Figure 1.

The double integration for the consistency probability can be reduced to the single integration by dividing the integration range of the outer integration into two separate ranges as shown below:

\[
P\left(\frac{1 - \exp(\hat{\gamma}_{JP})}{1 - \exp(\hat{\gamma}_{All})} > \pi\right)
= \int_{-\infty}^{\log(1-\pi)} f(x_{JP} | \gamma_{JP}, \sigma_{JP}^2) \left\{\int_{-\infty}^{\infty} f(x_{NJ} | \gamma_{NJ}, \sigma_{NJ}^2) dx_{NJ}\right\} dx_{JP}
+ \int_{\log(1-\pi)}^{\infty} f(x_{JP} | \gamma_{JP}, \sigma_{JP}^2) \left\{\int_{\max(x_{JP})}^{\infty} f(x_{NJ} | \gamma_{NJ}, \sigma_{NJ}^2) dx_{NJ}\right\} dx_{JP}
\]
Fig. 1. Region to be integrated

Parameters assumed: \( \pi = 0.5, \theta = 0.5, E = 161, E_{JP} = 30 \) and \( w = 30/161 = 0.186 \)

\[
\hat{\gamma}_{JP} = \frac{1}{1-w} \log \left( \frac{\exp(\hat{\gamma}_{JP}) + \pi - 1}{\pi \exp(w \hat{\gamma}_{JP})} \right)
\]

An R code for the calculation based on (16) is given in Appendix B. The formula (16) is more complicated than (14), but the CPU time for this calculation in R is much shorter than that for (14). Although we were not able to code a successful SAS program for the double integration based on (14), the numerical integration based on (16) is possible using CALL QUAD in SAS/IML and an example SAS code is provided in Appendix C. We have confirmed that this SAS code gives the same result as the R code for the numerical integration approach in all parameter settings that we have tried.

Since we cannot explicitly solve for the Japanese fraction \( w \) required for attaining a given consistency probability based on numerical integration approach, we need to obtain the solution by a grid search or by numerically solving the nonlinear equation (2) with respect to \( w \) using functions for numerical solutions for nonlinear equations such as function ‘uniroot’ in R.

We can also calculate the joint probability of overall study result being statistically significant.
and observing a consistent result using the numerical integration technique. The region to be integrated is indicated by the shaded area shown in Figure 2. The area is separated by two curves:

\[ \hat{\gamma}_{NJ} = \frac{1}{1-w} \log \left\{ \frac{\exp(\hat{\gamma}_{JP}) + \pi - 1}{\pi \exp(w_{\gamma_{JP}})} \right\} \] and \[ \hat{\gamma}_{NJ} = \frac{-z_{1-\alpha/2} \sigma_{NJ} - w\hat{\gamma}_{JP}}{1-w}, \]

where \[ \sigma_{NJ}^2 = \frac{1}{\theta(1-\theta)E}. \]

Let the value of \[ \hat{\gamma}_{JP} \] at the intersection points of the two curves be denoted by \( c \) which can be expressed as

\[ c = \log \{ \pi \exp(-z_{1-\alpha/2} \sigma_{NJ}) + 1 - \pi \}, \]

and let \( v \) be a function such that

\[ v(x) = \frac{-z_{1-\alpha/2} \sigma_{NJ} - wx}{1-w}, \]

then the joint probability can be expressed as follows:

\[
P \left( \frac{1 - \exp(\hat{\gamma}_{JP})}{1 - \exp(\hat{\gamma}_{NJ})} > \pi \text{ and } \frac{\hat{\gamma}_{NJ}}{\sqrt{1/\theta(1-\theta)E}} < -z_{1-\alpha/2} \right) \\
= \int_{-\infty}^{\log(1-\pi)} f(x_{JP} \mid \gamma_{JP}, \sigma_{JP}^2) \left[ \int_{-\infty}^{v(x_{JP})} f(x_{NJ} \mid \gamma_{NJ}, \sigma_{NJ}^2) dx_{NJ} \right] dx_{JP} \\
+ \int_{\log(1-\pi)}^{c} f(x_{JP} \mid \gamma_{JP}, \sigma_{JP}^2) \left[ \int_{u(x_{JP})}^{\infty} f(x_{NJ} \mid \gamma_{NJ}, \sigma_{NJ}^2) dx_{NJ} \right] dx_{JP} \\
= \int_{-\infty}^{\log(1-\pi)} f(x_{JP} \mid \gamma_{JP}, \sigma_{JP}^2) \int_{u(x_{JP})}^{\infty} F(v(x_{JP}) \mid \gamma_{NJ}, \sigma_{NJ}^2) dx_{NJ} dx_{JP} \]

\[ \text{P} \]
\[ + \int_{ \log(1 - \pi) }^{ \infty } f(x_{JP} \mid \gamma_{JP}, \sigma_{JP}^2) \{ F(v(x_{JP} \mid \gamma_{NJ}, \sigma_{NJ}^2) - F(u(x_{JP} \mid \gamma_{NJ}, \sigma_{NJ}^2)) \} dx_{JP} \]

where \( u \) is the function defined in (17).

5. Comparison of various computation methods

We compared 3 approaches (Quan’s method, alternative approximation and numerical integration) for calculating consistency probability stated in this paper based on Monte Carlo simulations. For each method, the required number of events in Japanese with \( \pi = 0.5 \), \( \theta = 0.5 \) and \( 1 - \beta' = 0.8 \) was calculated and the sample sizes for overall and Japanese were calculated using the same assumptions as in simulations by Quan et al (2010), i.e., survival is exponentially distributed with hazard rates for placebo and active groups set to \( \lambda_0 = 0.05 \) and \( \lambda_1 = \lambda_0 \exp(\gamma) \), respectively, with no drop-out, and fixed study duration \( L = 36 \). The required number of events for the overall population was calculated using Schoenfeld’s method (Schoenfeld, 1981). Table 1 shows the consistency probability based on the simulations.

**Table 1.** Comparison of various methods for proportional hazards model based on simulations

| HR  |  | Method               | \( E_{JP} \) | Consistency Probability based on simulation |
|-----|-----|---------------------|-------------|-------------------------------------------|
| 0.8 | 844 | Alternative approximation | 175.0       | 0.795                                     |
|     |     | Numerical integration | 173.5       | 0.794                                     |
|     |     | Quan’s Method        | 156.2       | 0.780                                     |
| 0.7 | 330 | Alternative approximation | 65.4        | 0.797                                     |
|     |     | Numerical integration | 64.6        | 0.797                                     |
|     |     | Quan’s Method        | 54.1        | 0.776                                     |
| 0.6 | 161 | Alternative approximation | 30.3        | 0.807                                     |
|     |     | Numerical integration | 29.8        | 0.799                                     |
|     |     | Quan’s Method        | 22.8        | 0.767                                     |
| 0.5 | 87  | Alternative approximation | 15.5        | 0.815                                     |
|     |     | Numerical integration | 15.2        | 0.799                                     |
|     |     | Quan’s Method        | 10.3        | 0.765                                     |

Simulations based on 10000 runs. \( E \) is calculated based on Schoenfeld’s method (Schoenfeld, 1981) using 2-sided significance level \( \alpha = 0.05 \), power \( 1 - \beta = 0.9 \) and allocation ratio \( \theta = 0.5 \). The number of subjects, not shown in table, is then calculated assuming placebo hazard rate \( \lambda_0 = 0.05 \) and fixed follow-up period \( L = 36 \). \( E_{JP} \) is calculated based on target power for consistency \( 1 - \beta' = 0.8 \). The consistency criteria used is \( \pi = 0.5 \).

The required number of events in Japanese based on alternative approximation and numerical integration are similar, and the actual power based on simulation is close to the nominal power. On the other hand, the method of Quan et al (2010) gives smaller required number of events in Japanese than the other methods resulting in the actual power that is approximately 2 to 3% smaller than the nominal value for the consistency probability. As the hazard ratio decreases, Quan’s method tended have slightly lesser actual power of showing consistency.
6. Risk reduction versus log(hazard ratio) as \( \hat{d} \)

As mentioned in Section 1, when we use hazard ratio as a measure of efficacy, there are two ways of defining \( \hat{d} \) in the condition of consistency (1): log-transformed hazard ratio, \( \hat{\gamma} \), or risk reduction, \( 1 - \exp(\hat{\gamma}) \). Here we will compare these two definitions of consistency to understand the property of (2) in more detail.

Depending on the choice of \( \hat{d} \), the meaning of \( \pi \) is different, so we may need to use different value of \( \pi \). However, if we assume the same value of \( \pi \) for both \( \hat{\gamma} \) and \( 1 - \exp(\hat{\gamma}) \) as \( \hat{d} \), we can compare the required Japanese fraction between them for a given value of \( 1 - \beta' \). Figure 3 shows the required fraction of Japanese for \( \pi = 0.5, \theta = 0.5, \) and \( 1 - \beta' = 0.8 \) under the condition that the sample size of the whole study population is calculated to satisfy power 90% at 5% two-sided significance level. It is clear that the fraction of Japanese is constant (22.4%) irrespective of hazard ratio for the definition of consistency in terms of log-transformed hazard ratio. On the other hand, the fraction of Japanese decreases along with hazard ratio for the definition of consistency in terms of risk reduction and it approaches the same level of that for log(hazard ratio) case as hazard ratio approaches 1. There is a remarkable difference between the two criteria and we can fulfill the required condition with the smaller number of Japanese patients if we use the criteria defined in terms of risk reduction.

When hazard ratio is expected to be very low (that is, test treatment is very efficacious), the calculated sample size of Japanese sample size based on the criteria (2) will be small. In
such a case, the observed number of events in Japanese subpopulation may be very small (or, in extreme case, zero events in the new treatment arm), providing almost no information for evaluating ethnic differences in efficacy as well as in safety. Therefore, the determination of Japanese sample size should not only be based on consistency probability, but also take other factors into account such as probability of obtaining zero event in Japanese subpopulation and the minimum number of subjects required for safety evaluation in Japanese.

7. Discussion/Summary

In this paper, we proposed two methods for calculating the consistency probability in (2) and compared them with Quan’s method. Through simulations, we noted that our proposed methods provided similar required number of events in Japanese and the actual power close to the nominal power, while Quan’s method tended to require less number of events and have lesser actual power than the nominal power as the assumed hazard ratio decreases. Whether the smaller required event based on Quan’s method at the cost of a lower actual power for consistency (2 to 3% lower than the nominal power in the settings considered in this paper) is worth it or not depends on how the sponsor or investigator planning for the clinical trial perceives the risk. However, when the expected hazard ratio is very small, Quan’s method should be used with a caution and our proposed methods can be used to obtain Japanese sample sizes that provide actual power for consistency close to the nominal power.

The methods for calculating consistency probability for proportional hazards model described in this paper can be applied to other models where the treatment effect is assumed to be log-normally distributed. For example, in negative binomial regression, which is commonly used to model recurrent events such as COPD (chronic obstructive pulmonary disease) exacerbations (Keene et al, 2007), the estimator of treatment effect in rate ratio (RR) is usually assumed to be log-normally distributed with the variance Var(log(RR)) (Zhu and Lakkis, 2014). By substituting Var(log(RR)) in place of Var(log(HR)) (i.e. $\frac{1}{\theta(1-\theta)E}$) in the formulas for proportional hazards models, we can obtain the formula for calculating the consistency probability for the negative binomial regression. The same argument is also applicable to the risk reduction of events that are modeled by the Poisson regression. The discussions in Section 6 holds analogously for the relation between log(RR) and $1 - RHR$ in consistency criteria as well.

Acknowledgements

We would like to give special thanks to Kenjiro Fukase for helpful discussion, review and comments throughout the development of this manuscript. We would like to thank the two anonymous referees for their helpful comments to improve the manuscript. We would also like to thank Hyosung Kim, Masatoshi Sugeno and Masahiro Nii for discussing the topics of this manuscript within Statistics Group of AstraZeneca Japan.
REFERENCES

Ikeda, K. and Bretz, F. (2010). Sample size and proportion of Japanese patients in multi-regional trials. *Pharmaceutical Statistics*. 9, 207–216.

Keene, O.N., Jones, M.R.K., Lane, P.W., and Anderson, J. (2007). Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN. *Pharmaceutical Statistics*. 6, 89–97.

Ministry of Health, Labour and Welfare (2007). Basic Principles on Global Clinical Trials, PFSB/ELD Notification No.0928010 dated September 28, 2007.

Quan, H., Zhao, P.-H., Zhang, J., Roessner, M., and Aizawa, L. (2010). Sample size considerations for Japanese patients in a multi-regional trial based on MHLW guidance. *Pharmaceutical Statistics*. 9, 100–112.

Schoenfeld, D. (1981). The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*. 68, 316–319.

Zhu, H. and Lakkis, H. (2014). Sample size calculation for comparing two negative binomial rates. *Statistics in Medicine*. 10, 376–387.

Appendix

A.

Here, we will show that (6) overestimates the true consistency probability under the assumption (3) for the case $\pi = 0$. When $\pi = 0$, (7) reduces to

$$z_{Quan,0} = \{\exp(-\gamma) - 1\} \sqrt{w\alpha(1 - \alpha)}$$

and $P(\hat{\gamma}_{JP} < 0)$ is approximated by $\Phi(z_{Quan,0})$ from (6). On the other hand, based on (3), the $z$-value providing the correct probability of $P(\hat{\gamma}_{JP} < 0)$ is given by

$$z_0 = -\gamma \sqrt{\text{var}(\hat{\gamma}_{JP})} = -\gamma \sqrt{w\alpha(1 - \alpha)}$$

(A.1)

The difference between these $z$-values is

$$z_{Quan,0} - z_0 = \{\exp(-\gamma) - 1 + \gamma\} \sqrt{w\alpha(1 - \alpha)}.$$

The first derivative of $z_{Quan,0} - z_0$ with respect to $\gamma$ is

$$\frac{d}{d\gamma}(z_{Quan,0} - z_0) = \{-\exp(-\gamma) + 1\} \sqrt{w\alpha(1 - \alpha)}$$

which is negative when $\gamma < 0$. This means that $z_{Quan,0} - z_0$ is a monotone decreasing function of $\gamma$. In addition, when $\gamma = 0$, $z_{Quan,0} - z_0 = 0$. Therefore, we conclude that $z_{Quan,0} - z_0 > 0$ when $\gamma < 0$ and thus $\Phi(z_{Quan,0})$ overestimates $P(\hat{\gamma}_{JP} < 0)$. 

Jpn J Biomet Vol. 38, No. 2, 2017
B. R code for numerical integration

```r
HR <- 0.6  # hazard ratio
e <- 161   # event number
eJP <- 30  # event number in Japanese
eNJ <- e-eJP # event number in Non-Japanese
w <- eJP/e # fraction of Japanese
theta <- 0.5 # allocation fraction
Pi <- 0.5 # criteria for consistency
gamma <- log(HR) # log hazard ratio (gamma)
sd <- sqrt(1/c(eJP,eNJ)/theta/(1-theta)) # SD of gamma estimates

integrand <- function(gammaJP){
  lower <- log((exp(gammaJP)+Pi-1)/Pi/exp(w*gammaJP))/(1-w)
  q <- dnorm(gammaJP, mean=gamma,sd=sd[1])*
       (1-pnorm(lower,mean=gamma,sd=sd[2]))
}
outer <- integrate(Vectorize(integrand),log(1-Pi),Inf)$value
(p <- outer+pnorm(log(1-Pi),mean=gamma,sd=sd[1]))
```

The execution of the above R code returns 0.8009655.

C. SAS code for numerical integration

```sas
proc iml;
hr = 0.6; * hazard ratio;
e = 161; * event number;
eJP = 30; * event number in Japanese subpopulation;
eNJ = e-eJP; * event number in Non-Japanese subpopulation;
w = eJP/e; * Japanese fraction;
theta = 0.5; * allocation fraction;
pi = 0.5; * criteria for consistency;
gamma = log(HR); * log hazard ratio (gamma);
sdJP = sqrt(1/eJP/theta/(1-theta)); * SD of gamma estimate in JP;
sdNJ = sqrt(1/eNJ/theta/(1-theta)); * SD of gamma estimate in NJ;

* ingetrand;
start ingetrand(gammaJP) global(gamma,sdJP,sdNJ,w,pi);
  lower = log((exp(gammaJP)+pi-1)/pi/exp(w*gammaJP))/(1-w);
  q = pdf("NORMAL",gammaJP,gamma,sdJP)
      *(1-cdf("NORMAL",lower,gamma,sdNJ));
  return(q);
finish;

* ingetration;
range = log(1-pi)||.P;
call quad(outer, "ingetrand",range) peak=0;
p = outer+cdf("NORMAL",log(1-pi),gamma,sdJP);
print p;
```

Jpn J Biomet Vol. 38, No. 2, 2017
The execution of the above SAS code returns 0.8009655. This is the exact same value that was obtained using the R code for the numerical integration shown in Appendix B. In the same parameter setting, the method of Quan et al gives 0.838 and the alternative approximation method gives 0.799. This numerical example suggests that the method of Quan et al gives an optimistic value of consistency probability in comparison with the other methods.

Note: In this SAS code, option ‘peak = 0’ is specified after the CALL QUAD statement. We have confirmed empirically that this value allows the integration using CALL QUAD to converge in most cases, but it does not ensure the convergence in all cases. If the integration does not converge, other values should be tried.