SAMPLE SIZE AND OPTIMAL DESIGNS IN STRATIFIED COMPARATIVE TRIALS TO ESTABLISH THE EQUIVALENCE OF TREATMENT EFFECTS AMONG TWO ETHNIC GROUPS

Yi-Hau Chen1,* and Mey Wang2

1Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, ROC
2Center for Drug Evaluation, Taipei, Taiwan, ROC

ABSTRACT

When a new investigational medicine is intended to be applied to populations with different ethnic backgrounds, a stratified comparative phase III trial using ethnic groups as strata may be conducted to assess the influence of ethnic factors on clinical outcomes of this new medicine. In this paper, based on a binomial model with odds ratio as the measure of the treatment effect, we derive the score test and the associated sample size formula for establishing the equivalence/noninferiority of the treatment effects of a medicine among two ethnic groups. A simplified test together with its sample size formula are also given. Taking into account the sample size, cost, and power of testing, respectively, we derive the optimal design parameters, i.e., the allocation among treatment groups and ethnic groups, based on the simplified test.

Key Words: Clinical trial; Sample size; Score test; Stratified trial
Evaluating the impact of ethnic factors upon the effect of a new investigational drug is crucial for the future application of this new medicine to populations with different ethnic backgrounds. In accordance with the International Conference on Harmonization (ICH) tripartite guideline entitled *Ethnic Factors in the Acceptability of Foreign Clinical Data* (ICH, 1997), an ethnic sensitive compound is the one whose pharmacological characteristics suggest the potential for clinically significant impact by ethnic factors on clinical outcomes. During the drug developing stage, e.g., a phase III trial, showing an “ethnic sensitive” drug unlikely to behave differently in different populations would make it easier to allow for its application in different ethnic populations.

Suppose that we wish to show that a medicine’s clinical effectiveness in one ethnic population is equivalent or noninferior to that in the other population. Then a stratified comparative phase III trial using ethnic groups as strata can be conducted, and the equivalence of the treatment effects among the two ethnic populations can be established by rejecting a null hypothesis that the treatment effects in the two ethnic populations differ by at least a prescribed tolerance limit. Note that unlike the usual equivalence/noninferiority tests, where equivalence/noninferiority is to be determined among two treatments, in the present setting the equivalence/noninferiority is to be established among two strata formed by ethnic groups.

This paper considers sample size requirement and optimal design for examining the equivalence/noninferiority of the treatment effects among two ethnic groups in a stratified comparative phase III trial. This work is motivated by an anti-bacteria agent for the treatment of community-acquired pneumonia. In order to seek approval from different countries and a simultaneous global launch of the new drug, the company has planned to conduct a stratified comparative phase III trial with ethnic groups (i.e., Caucasian and Asian) as strata in the late drug development stage. The important design issues for such a study are then the sample size determination and the optimal allocation of treatment to different ethnic groups.

Focusing on the binary outcome, in second section we introduce the assumed underlying model and derive the score test and its sample size formula for establishing the equivalence/noninferiority of treatment effects among two ethnic groups. A simplified test is suggested in the third section. We show that, under a contiguous alternative, the test statistic of the simplified test is asymptotically equivalent to that of the score test. Sample size formula based on the simplified test is also provided. Fourth section displays a numerical example to illustrate the use of the sample size formulae given in the second and third sections. Taking into account the sample size, cost, or statistical power, in fifth section we derive optimal designs based on the simplified test. In sixth section we conclude this paper with several remarks.
THE SCORE TEST AND ITS SAMPLE SIZE FORMULA

Let \( x_{11} \) and \( x_{10} \) represent the respective numbers of responses of the treatment and control groups in ethnic group 1 (i.e., the population in the original region), and \( x_{21} \) and \( x_{20} \) the respective numbers of responses of the treatment and control groups in ethnic group 2 (i.e., the population in the new region). Assume that \( \{x_{10}, x_{11}, x_{20}, x_{21}\} \) are independent binomial variates with sample sizes \( \{n_{10}, n_{11}, n_{20}, n_{21}\} \) and response rates \( \{p_{10}, p_{11}, p_{20}, p_{21}\} \). Denote by OR\(_i\) the odds ratio in ethnic group \( i \), i.e., OR\(_i\) = \( p_{1i}q_{i0}/(p_{0i}q_{i1}) \), where \( q_{ij} = 1 - p_{ij}, j = 0, 1 \). Let \( \theta = \text{OR}_2 \) and \( \rho = \text{OR}_1/\text{OR}_2 \).

The log-likelihood is

\[
L(\rho, \theta, p_{10}, p_{20}) = \sum_{i,j} L_{ij}(\rho, \theta, p_{10}, p_{20}),
\]

where \( L_{ij} = x_{ij} \ln p_{ij} + (n_{ij} - x_{ij}) \ln q_{ij}, \ i = 1, 2, \ j = 0, 1, \) and \( p_{11} = \rho \theta p_{10}/\{1 + (\rho \theta - 1)p_{10}\}, \ p_{21} = \theta p_{20}/\{1 + (\theta - 1)p_{20}\} \).

Suppose that the treatment effect in ethnic group 1 is considered to be equivalent/noninferior to that in ethnic group 2 if OR\(_1\)/OR\(_2\) > \( \rho_0 \), where \( \rho_0 < 1 \) is a prescribed tolerance limit which represents materially important inequivalence/inferiority. Then, to establish such equivalence/noninferiority of treatment effects among two ethnic groups, a suitable set of null and alternative hypotheses is

\[
H_0 : \rho \leq \rho_0 \quad \text{vs.} \quad H_a : \rho = \rho_0 (> \rho_0).
\]

Let \( \hat{\theta}, \hat{p}_{10}, \hat{p}_{20} \) be the maximum likelihood estimators (MLE) of \( (\theta, p_{10}, p_{20}) \) with \( \rho \) fixed at \( \rho_0 \). See Appendix I for their calculations. The score of \( \rho \), i.e., the partial derivative of the log-likelihood with respect to \( \rho \), evaluated at \( (\rho_0, \hat{\theta}, \hat{p}_{10}, \hat{p}_{20}) \) is given by

\[
S(\rho_0) = (x_{11} - n_{11}\hat{p}_{11})/\rho_0.
\]

where \( \hat{p}_{11} = \rho_0 \hat{\theta} p_{10}/\{1 + (\rho_0 \hat{\theta} - 1)\hat{p}_{10}\} \). By the general theory of Bartlett,\(^{[1]}\) we can derive the variance of the score \( S(\rho_0) \) under \( H_0 \) and it can be estimated by

\[
\hat{\text{var}}_0\{S(\rho_0)\} = \left\{ \rho_0^2 \sum_{i=1,2} \left( \frac{1}{n_{1i}\hat{p}_{1i}\hat{q}_{i1}} + \frac{1}{n_{0i}\hat{p}_{0i}\hat{q}_{i0}} \right) \right\}^{-1},
\]

where \( \hat{p}_{i1} = \rho_0^{2-i} \hat{\theta} p_{i0}/\{1 + (\rho_0^{2-i} \hat{\theta} - 1)\hat{p}_{i0}\}, \ i = 1, 2 \). The score statistic for testing Eq. (1) is

\[
z = S(\rho_0)/[\hat{\text{var}}_0\{S(\rho_0)\}]^{1/2}
\]

\[
= (x_{11} - n_{11}\hat{p}_{11}) \left\{ \sum_{i=1,2} \left( \frac{1}{n_{1i}\hat{p}_{1i}\hat{q}_{i1}} + \frac{1}{n_{0i}\hat{p}_{0i}\hat{q}_{i0}} \right) \right\}^{1/2}, \quad (2)
\]
When $z > z_{(1-\alpha)}$, where $z_{(1-\alpha)}$ is the 100(1 - $\alpha$) percentile of the standard normal distribution, we would reject $H_0$ against $H_a$.

To facilitate the power calculation, let $(\hat{\theta}, \hat{p}_{10}, \hat{p}_{20})$ be the limiting value of $(\theta, \bar{p}_{10}, \bar{p}_{20})$ under $H_a$. The calculation of $(\theta, \bar{p}_{10}, \bar{p}_{20})$ is shown in Appendix II. Under $H_a$, the limiting value of $\var_0(S(\rho_0))$ is

$$V_0 = \left\{ \rho_0^2 \sum_{i=1}^{2} \left( \frac{1}{n_{i1}\bar{p}_{i1}\bar{q}_{i1}} + \frac{1}{n_{0}\bar{p}_{0}\bar{q}_{0}} \right) \right\}^{-1},$$

and the expectation and variance of the score $S(\rho_0)$ under $H_a$ are, respectively,

$$E_a(S(\rho_0)) = n_{11}(p_{11} - \bar{p}_{11})/\rho_0$$

and

$$V_a = \text{var}_a(S(\rho_0)) = \left[ \rho_0^2 \sum_{i=1}^{2} \left( \frac{p_{i1}q_{i1}}{n_{i1}(\bar{p}_{i1}\bar{q}_{i1})^2} + \frac{p_{0}q_{0}}{n_{0}(\bar{p}_{0}\bar{q}_{0})^2} \right) \right]^{-1},$$

where $p_{ij}$'s are the alternative-hypothesis values of the response rates. The asymptotic power function for the score test of significance level $\alpha$ is then given by $1 - \Phi(u)$, where $\Phi(\cdot)$ is the standard normal distribution function and

$$u = [z_{(1-\alpha)}V_0^{1/2} - E_a(S(\rho_0))] / V_a^{1/2}. $$

Based on the asymptotic power function, we can then derive the required sample size for the score test of significance level $\alpha$ to attain a given power, say $1 - \beta$. Let $N = \sum_i(n_{i1} + n_{i0})$ be the total sample size, $\lambda_i = (n_{i1} + n_{i0})/N$ the proportion of ethnic group $i$, and $\phi_i = n_{i1}/(n_{i1} + n_{i0})$ the fraction of the treatment group within ethnic group $i$, $i = 1, 2$. Let

$$\nu_0 = \left[ \sum_{i=1}^{2} \left\{ \frac{1}{\lambda_i\phi_i\bar{p}_{i1}\bar{q}_{i1}} + \frac{1}{\lambda_i(1 - \phi_i)\bar{p}_{0}\bar{q}_{0}} \right\} \right]^{-1},$$

and

$$\nu_a = \left[ \sum_{i=1}^{2} \left\{ \frac{p_{i1}q_{i1}}{\lambda_i\phi_i(\bar{p}_{i1}\bar{q}_{i1})^2} + \frac{p_{0}q_{0}}{\lambda_i\phi_i(\bar{p}_{0}\bar{q}_{0})^2} \right\} \right]^{-1}.$$

The sample size formula of the one-sided score test for given significance level $\alpha$ and power $1 - \beta$ is then given by

$$N_{SC} = (z_{(1-\alpha)}\nu_0^{1/2} + z_{(1-\beta)}\nu_a^{1/2})^2 / \{\lambda_1\phi_1(p_{11} - \bar{p}_{11})\}^2. \quad (3)$$
A SIMPLIFIED TEST AND ITS SAMPLE SIZE FORMULA

It can be shown that (Appendix III), at the contiguous alternative, i.e., \( r_a / r_0 < 1 \); the score statistic (2) is approximately equivalent to the statistic

\[
z_S = \left( \ln \hat{OR}_1 - \ln \hat{OR}_2 - \ln \rho_0 \right) \left\{ \sum_{i=1,2} \left( \frac{1}{n_{i1} \hat{p}_{i1} \hat{q}_{i1}} + \frac{1}{n_{i0} \hat{p}_{i0} \hat{q}_{i0}} \right) \right\}^{1/2},
\]

where \( \ln \hat{OR}_i = \ln \{ \hat{p}_{i1} \hat{q}_{i0} / (\hat{p}_{i0} \hat{q}_{i1}) \} \), \( \hat{p}_{ij} = x_{ij} / n_{ij} \), and \( \hat{q}_{ij} = 1 - \hat{p}_{ij} \), \( i = 1,2 \), \( j = 0,1 \). We then define a simplified test by the test statistic \( z_S \), and the null hypothesis in Eq. (1) would be rejected when \( z_S > z_{(1-\alpha)} \).

At the alternative, the asymptotic power function for the simplified test of significance level \( \alpha \) can be written as \( 1 - \Phi(u_S) \), where

\[
u_S = z_{(1-\alpha)} - \ln(\rho_a / \rho_0) \left\{ \sum_{i=1,2} \left( \frac{1}{n_{i1} \hat{p}_{i1} \hat{q}_{i1}} + \frac{1}{n_{i0} \hat{p}_{i0} \hat{q}_{i0}} \right) \right\}^{-1/2}.
\]

The required sample size for the simplified test at a given significance level \( \alpha \) and a given power \( 1 - \beta \) can thus be approximated by the formula

\[
N_S = \left[ (z_{(1-\alpha)} + z_{(1-\beta)}) / \ln(\rho_a / \rho_0) \right]^2 \times \left\{ \sum_{i=1,2} \left( \frac{1}{\lambda_i \hat{p}_{i1} \hat{q}_{i1}} + \frac{1}{\lambda_i (1 - \hat{p}_i) \hat{q}_{i0}} \right) \right\}. \tag{4}
\]

We close this section by the notice that the equivalence/noninferiority of treatment effects among ethnic groups in fact amounts to the noninteraction between treatment and ethnicity. Hence it is expected that the required sample size to achieve a given power for establishing the equivalence/noninferiority of treatment effects among two ethnic groups would be much larger than that needed for demonstrating the treatment effects within a single group.

A NUMERICAL EXAMPLE

We illustrate by a numerical example the application of the sample size formulae given in Eqs. (3) and (4). Consider the sample size for designing a stratified comparative trial to establish the equivalence of the respective effectiveness of an anti-bacteria agent for Caucasian and Asian. Let \( p_{i1} \) and \( p_{i0} \) be the respective success rates of the treatment and control groups in ethnic group \( i \), where \( i = 1 \) stands for the Asian group and \( i = 2 \) stands for the Caucasian group. If the treatment effects in Asian, in terms of the odds ratio \( p_{i1} q_{i0} / (p_{i0} q_{i1}) \), is above 80% of that in Caucasian, we may regard this anti-bacteria agent performs
equally well in both Asian and Caucasian. Therefore, the tolerance limit $r_0$ is set to 0.8. The alternative of interest is $r = 1$. With $\alpha = 0.05$ and $\beta = 0.3$, we employ the sample size formulae (3) and (4) for various values of response rates and design parameters. We choose $\beta = 0.3$ (70% power) rather than $\beta = 0.2$ or 0.1 (80 or 90% power) because, as mentioned above, to show equivalence of treatment effects among ethnic groups often requires much more sample size than that a usual study can afford. So the compromise may be to set a lower level for the testing power. The results are displayed in Table 1. Also shown in Table 1 are the approximated sample sizes for the likelihood ratio (LR) test, which are obtained from the formula proposed by Self et al.\cite{2} (see also Ref. [3]). To apply the method by Self et al. to the present setting, first we need to compute the difference in expected single-observation log-likelihoods evaluated, respectively, at the alternative- and null-hypothesis values of parameters, which is given by 

$$g = \sum_{i=1,2} \left( \lambda_i \phi_i \{ p_{i1} \ln p_{i1}^* + (1 - p_{i1}) \ln (1 - p_{i1}^*) \} + \lambda_i (1 - \phi_i) \{ p_{i0} \ln p_{i0}^* + (1 - p_{i0}) \ln (1 - p_{i0}^*) \} \right)$$

is the expected single-observation log-likelihood evaluated at hypothesized parameter values $\{ p_{ij}^* \}$. Let $\eta$ be the noncentrality parameter of the noncentral $\chi^2$ variate whose probability of exceeding $\chi^2(1 - \alpha)$, the 100(1 - $\alpha$) percentile of

<table>
<thead>
<tr>
<th>$\lambda_1$</th>
<th>$\phi_1$</th>
<th>$\phi_2$</th>
<th>Score</th>
<th>Simplified</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11} = p_{21} = 0.67, p_{10} = p_{20} = 0.5, \rho = 0.8$</td>
<td>0.5</td>
<td>0.5</td>
<td>6.461</td>
<td>6.426</td>
<td>8.428</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.3</td>
<td>7.833</td>
<td>7.831</td>
<td>10.270</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.7</td>
<td>7.580</td>
<td>7.470</td>
<td>9.797</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.4</td>
<td>6.664</td>
<td>6.694</td>
<td>8.767</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.3</td>
<td>7.868</td>
<td>7.650</td>
<td>10.005</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.3</td>
<td>9.352</td>
<td>9.322</td>
<td>12.164</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.7</td>
<td>8.972</td>
<td>8.893</td>
<td>11.651</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.4</td>
<td>7.992</td>
<td>8.007</td>
<td>10.454</td>
</tr>
<tr>
<td>$p_{11} = p_{21} = 0.82, p_{10} = p_{20} = 0.7, \rho = 0.8$</td>
<td>0.5</td>
<td>0.5</td>
<td>8.824</td>
<td>8.802</td>
<td>11.545</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.3</td>
<td>11.233</td>
<td>11.242</td>
<td>14.749</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.7</td>
<td>9.818</td>
<td>9.716</td>
<td>12.743</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.4</td>
<td>9.152</td>
<td>9.169</td>
<td>11.974</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.5</td>
<td>10.521</td>
<td>10.479</td>
<td>13.681</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.3</td>
<td>13.466</td>
<td>13.383</td>
<td>17.370</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.7</td>
<td>11.615</td>
<td>11.567</td>
<td>15.198</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.4</td>
<td>11.095</td>
<td>11.075</td>
<td>14.387</td>
</tr>
</tbody>
</table>
the central $\chi^2$ distribution, is $1 - \beta$. Then the approximated sample size required for the LR test with significance level $\alpha$ and power $1 - \beta$ is given by

$$N_{LR} = \eta/g.$$ 

Two points can be noted from Table 1. First, to establish the equivalence/noninferiority of the treatment effects among two ethnic groups, the requisite sample sizes are quite large. This is because here we are testing whether the interaction between treatment and ethnicity exists. Second, as expected, the sample sizes obtained by the score method (3) are similar to those by the simplified method (4), but neither one of them would be uniformly less than or equal to the other over all possible values of response rates $p_{ij}$'s and design parameters $\lambda_i$'s and $\phi_i$'s. On the other hand, the LR method requires an excess of about one-third samples needed by the score and simplified methods.

**OPTIMAL DESIGNS BASED ON THE SIMPLIFIED TEST**

In this section we focus on design issues for the simplified test. First we consider the scenario concerning the determination of design parameters $\lambda_i$'s and $\phi_i$'s such that the required total sample size for the simplified test with a given significance level $\alpha$ and a given power $1 - \beta$ is minimized. Consider the approximated sample size formula $N_S$ given in Eq. (4). Recall that $\lambda_1 + \lambda_2 = 1$, thus by Cauchy–Schwartz inequality the minimal sample size is given by

$$N^*_S = \left\{ \frac{\left( z_{1-\alpha} + z_{1-\beta} \right)}{\ln(p_a/p_0)} \left( \sum_{i=1,2} \left( p_{11}q_{11} \right)^{-1/2} + \left( p_{00}q_{00} \right)^{-1/2} \right) \right\}^2,$$

which corresponds to the design with

$$\lambda_i = \frac{\left( \sum_{j=0,1} \left( p_{ij}q_{ij} \right)^{-1/2} \right)}{\left( \sum_{i,j} \left( p_{ij}q_{ij} \right)^{-1/2} \right)},$$

$$\phi_i = \left( p_{11}q_{11} \right)^{-1/2} \left\{ \sum_{j=0,1} \left( p_{ij}q_{ij} \right)^{-1/2} \right\}, \quad i = 1, 2. \quad (5)$$

Refer to Appendix IV for detailed derivations. The solution (5) thus provides a set of optimal design parameters yielding the minimum sample size needed to attain a given power $1 - \beta$ for the simplified test of significance level $\alpha$.

In Table 2, we demonstrate results of optimal values of $\lambda_1$, $\phi_1$, and $\phi_2$ corresponding to some values of response rates $p_{ij}$'s. The minimum sample size $N^*_S$, together with the sample size $N^e_S$ under an equal sample design, i.e.,
n_{11} = n_{10} = n_{21} = n_{20}, needed by a simplified test to achieve a 70% power are also shown in Table 2. Compared with the equal sample size design, the decrease in sample size by the optimal design with the minimum sample size ranges from 0.1 to 5%.

Further, we take the cost into account. Let $C_{ij}$ be the cost for each subject in ethnic group $i$ and treatment group $j$, $i = 1, 2$, $j = 0, 1$. Hence $C = N\sum_{i,j}[\lambda_i \phi_i C_{11} + \lambda_i (1 - \phi_i) C_{10}]$ is the total cost. Suppose that the objective is to find $\lambda_i$’s and $\phi_i$’s so that the total cost $C$ required to achieving a power $1 - \beta$ for the simplified test of significance level $\alpha$ is minimized. It can be obtained that (see Appendix IV) the minimal cost is

$$C^* = \left\{ z(1 - \alpha) + z(1 - \beta) \right\} / \ln(p_0 / p_0) \right[ \sum_{i,j} \{ C_{ij} / (p_{ij} q_{ij}) \}^{1/2} \right]^{2},$$

which corresponds to the design with

$$\lambda_i = \frac{\left\{ \sum_{j=0,1} (C_{ij} p_{ij} q_{ij})^{-1/2} \right\}}{\left\{ \sum_{i,j} (C_{ij} p_{ij} q_{ij})^{-1/2} \right\}},$$

$$\phi_i = (C_{i1} p_{i1} q_{i1})^{-1/2} / \left\{ \sum_{j=0,1} (C_{ij} p_{ij} q_{ij})^{-1/2} \right\}, \quad i = 1, 2.$$  \hspace{1cm} (6)
The associated total sample size is
\[ \left\{ \left[ \left( 1 - \alpha \right) + \left( 1 - \beta \right) / \ln \left( \rho_0 / \rho_0 \right) \right]^2 \left[ \sum_{i,j} \left\{ C_{ij} / \left( p_{ij} q_{ij} \right) \right\}^{1/2} \right] / \left[ \sum_{i,j} \left\{ C_{ij} / \left( p_{ij} q_{ij} \right) \right\}^{1/2} \right] \right\}. \]

The solution (6) thus gives the optimal allocation among ethnic groups and treatment groups, in the sense that the total cost is minimized for the simplified test with a fixed significance level and a fixed power. It can be seen that the optimal values of the design parameters \( \lambda_1, \lambda_2 (= 1 - \lambda_1), \phi_1, \) and \( \phi_2 \) depend on \( C_{ij} \)'s only through their relative magnitudes.

For various specifications of the values for the response rates \( p_{ij} \)'s and the relative costs \( C_{10} / C_{20}, C_{11} / C_{10}, \) and \( C_{21} / C_{20} \), the optimal values of \( \lambda_1, \phi_1, \) and \( \phi_2 \), and the associated total sample size \( N \) are computed and shown in Table 3. Also shown in Table 3 is the relative total cost (RC) which is the ratio of the total costs for the optimal design with the minimum cost and the equal sample size design. We note from Table 3 that the reduction in total cost by the optimal design can be rather remarkable when the costs among treatment groups and/or ethnic groups vary drastically.

In certain situations the total cost may be fixed in advance so that the investigators may wish to determine the values of \( \lambda_i \)'s and \( \phi_i \)'s such that, with the total cost \( C \) fixed, the power of the simplified test of significance level \( \alpha \) can be maximized. It can be seen that this problem is definitely equivalent to the previous one concerning the minimum cost subject to a fixed power. Hence the solution is also given by Eq. (6). With the specifications in Eq. (6), the corresponding total sample size is
\[ N = C \left\{ \sum_{i,j} \left\{ C_{ij} / \left( p_{ij} q_{ij} \right) \right\}^{-1/2} \right\} / \left[ \sum_{i,j} \left\{ C_{ij} / \left( p_{ij} q_{ij} \right) \right\}^{1/2} \right], \]

### Table 3. Optimal Designs with Minimum Total Cost Based on the Simplified Test

<table>
<thead>
<tr>
<th>( C_{10} / C_{20} )</th>
<th>( C_{11} / C_{10} )</th>
<th>( C_{21} / C_{20} )</th>
<th>( \lambda_1 )</th>
<th>( \phi_1 )</th>
<th>( \phi_2 )</th>
<th>( N )</th>
<th>RC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_{11} = 0.70, p_{10} = 0.51, p_{21} = 0.6, p_{20} = 0.4, \rho^x = 0.8 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.51</td>
<td>0.52</td>
<td>0.50</td>
<td>6456</td>
<td>99</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.42</td>
<td>0.52</td>
<td>0.50</td>
<td>6651</td>
<td>97</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>1.5</td>
<td>0.42</td>
<td>0.47</td>
<td>0.45</td>
<td>6725</td>
<td>97</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.59</td>
<td>0.52</td>
<td>0.50</td>
<td>6651</td>
<td>97</td>
</tr>
<tr>
<td>0.5</td>
<td>1.5</td>
<td>1.5</td>
<td>0.59</td>
<td>0.47</td>
<td>0.45</td>
<td>6715</td>
<td>96</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0.37</td>
<td>0.52</td>
<td>0.50</td>
<td>6955</td>
<td>94</td>
</tr>
<tr>
<td>0.33</td>
<td>1</td>
<td>1</td>
<td>0.64</td>
<td>0.52</td>
<td>0.50</td>
<td>6955</td>
<td>93</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>2</td>
<td>0.43</td>
<td>0.47</td>
<td>0.41</td>
<td>6715</td>
<td>97</td>
</tr>
<tr>
<td>0.5</td>
<td>1.5</td>
<td>2</td>
<td>0.61</td>
<td>0.47</td>
<td>0.41</td>
<td>6872</td>
<td>93</td>
</tr>
</tbody>
</table>

\( N \) is the associated sample size required by the simplified test with 5% significance level and 70% power to establish the equivalence to treatment effects. RC is the relative cost of the minimum cost relative to the cost under an equal sample size design.
and the resultant power of testing is

\[
\Phi \left( \ln(\rho_a/\rho_0)C^{1/2}/ \left[ \sum_{i,j} \{ C_{ij}/(p_iq_j)\}^{1/2} \right] - z_{1-\alpha} \right).
\]

**DISCUSSION**

Conducting a race-stratified trial facilitates the scientific assessment of the influence of ethnic factors on a medicine. In this paper, under a binomial model with odds ratio as the measure of the treatment effect, we have provided sample size formulae for designing a trial to establish the equivalence/noninferiority of a medicine’s effectiveness among two ethnic groups. The proposed sample size formulae are based, respectively, on the score and its simplified tests. To apply these formulae, we need to specify a tolerance limit referring to materially significant inequivalence, as well as the response rates of all treatment-ethnicity groups. Usually the determination of the tolerance limit depends on the subject matter and the aim of the study, while the determination of the response rates is based on prior information from a pilot study and/or relevant references.

Based on the simplified test, we also provide formulae for the optimal design parameters, i.e., allocation among treatment and ethnic groups, under several scenarios. Again, the determination of the optimal values for design parameters requires a priori knowledge of the response rates. Therefore, in practice we have to estimate these response rates in advance and can obtain a close to optimal design by applying these estimates of response rates to the formulae of optimal design parameters.

Although in this paper we consider one-sided tests only, the provided sample size formulae can be extended to the two-sided test

\[
H_0 : \rho < \rho_0 \text{ or } \rho > \rho_1 (\rho_0 < \rho_1) \text{ vs. } H_a : \rho = \rho_0 (\rho_0 \leq \rho_a \leq \rho_1)
\]

by replacing \(z_{(1-\beta)}\) with \(z_{(1-\beta/2)}\) in the formulae.

Through the numerical example we have noticed that the remarkable expansion of the required sample size to establish the equivalence of the treatment effects among two ethnic groups. One possible strategy to overcome this difficulty is to adopt a Bayesian approach to the utilization of prior information on the treatment effects. For example, the treatment effect in one ethnic group may be available from previous trials so that such previous knowledge can be incorporated into the analysis to reduce the required sample size. A Bayesian approach to the design and analysis in the context of active control trials has been established by Simon.\(^{[4]}\) In the spirit of his method, development of the Bayesian approach in the setting considered in this paper deserves further studies.
APPENDIX I: CALCULATION OF MAXIMUM LIKELIHOOD ESTIMATORS OF $\theta$, $p_{10}$, AND $p_{20}$ UNDER $H_0 : \rho = \rho_0$

The MLEs of $\theta$, $p_{10}$, and $p_{20}$ under $H_0 : \rho = \rho_0$ are obtained by solving the simultaneous score equations

$$\frac{\partial \ln L(\rho_0, \theta, p_{10}, p_{20})}{\partial \theta} = 0,$$

$$\frac{\partial \ln L(\rho_0, \theta, p_{10}, p_{20})}{\partial p_{10}} = 0,$$

$$\frac{\partial \ln L(\rho_0, \theta, p_{10}, p_{20})}{\partial p_{20}} = 0.$$

After direct calculations, the score equations reduce to

$$r \theta^2 + s \theta + t = 0, \quad a_1 p_{10}^2 + b_1 p_{10} + c_1 = 0,$$

$$a_2 p_{20}^2 + b_2 p_{20} + c_2 = 0,$$

where

$$r = \rho_0 p_{10} p_{20} (n_{11} q_{11} + n_{21} q_{21}), \quad s = \rho_0 p_{10} q_{20} (n_{11} - x_{11} - x_{21}) + p_{20} q_{10} (n_{21} - x_{11} - x_{21}), \quad t = -q_{10} q_{20} (x_{11} + x_{21}), \quad a_1 = n_0 (\rho_0^2 - \theta - 1), \quad b_1 = \rho_0^2 - \theta n_{11} + n_0 - (x_{11} + x_{21}) (\rho_0^2 - \theta - 1), \quad c_1 = -x_{11} - x_{21}, \quad i = 1, 0$$

Note that by expressing the score equations in the quadratic-equation form (7) we can solve the MLEs of $\theta$, $p_{10}$, and $p_{20}$ iteratively as follows. Start with some initial values for $\theta$, $p_{10}$, and $p_{20}$. Given the current values $\theta'$, $p'_{10}$, and $p'_{20}$ for $\theta$, $p_{10}$, and $p_{20}$, respectively, the updated values of $\theta$, $p_{10}$, and $p_{20}$ are given by

$$\theta = \{ - s' + (s'^2 - 4r' r')^{1/2} \} / (2r'),$$

$$p'_{10} = \{ - b'_1 + (b'_1^2 - 4a'_1 c'_1)^{1/2} \} / (2a'_1), \quad i = 1, 2,$$

where

$$r' = \rho_0 p_{10}' p_{20}' (n_{11} q_{11}' + n_{21} q_{21}'), \quad s' = \rho_0 p_{10}' q_{20}' (n_{11} - x_{11} - x_{21}) + p_{20}' q_{10}' (n_{21} - x_{11} - x_{21}), \quad t' = -q_{10}' q_{20}' (x_{11} + x_{21}), \quad a'_1 = n_0 (\rho_0^2 - \theta' - 1), \quad b'_1 = \rho_0^2 - \theta' n_{11} + n_0 - (x_{11} + x_{21}) (\rho_0^2 - \theta' - 1), \quad c'_1 = -x_{11} - x_{21}.$$  

The iteration is finished when arriving at the convergence, and the MLEs $\tilde{\theta}$, $\tilde{p}_{10}$, and $\tilde{p}_{20}$ for $\theta$, $p_{10}$, and $p_{20}$ under $H_0 : \rho = \rho_0$ are obtained by the solution of the last iteration.

APPENDIX II: CALCULATION OF THE LIMITING VALUES FOR MAXIMUM LIKELIHOOD ESTIMATORS OF $\theta$, $p_{10}$, AND $p_{20}$ UNDER THE ALTERNATIVE HYPOTHESIS

The limiting values $\bar{\theta}$, $\bar{p}_{10}$, and $\bar{p}_{20}$ for $\tilde{\theta}$, $\tilde{p}_{10}$, and $\tilde{p}_{20}$ under the alternative of interest are solutions of the simultaneous equations

$$r \bar{\theta}^2 + s \bar{\theta} + t = 0, \quad a_1 \bar{p}_{10}^2 + b_1 \bar{p}_{10} + c_1 = 0, \quad a_2 \bar{p}_{20}^2 + b_2 \bar{p}_{20} + c_2 = 0,$$

where $r$, $s$, $i$, $a_i$, $b_i$, and $c_i$, $i = 1, 2$, are defined as $r$, $s$, $t$, $a_i$, $b_i$, and $c_i$ in Appendix I,
respectively, except that all the involved observations \( x_{ij} \), \( i = 1, 2 \), and \( j = 0, 1 \), are replaced by their expectations \( n_{ij} \hat{p}_{ij} \), where the \( p_{ij} \)'s are given by the alternative hypothesis of interest. The limiting values \( \theta, \hat{p}_{10} \), and \( \hat{p}_{20} \) can thus be computed in the iterative way stated in Appendix I.

**APPENDIX III: THE APPROXIMATE EQUIVALENCE OF THE SCORE AND SIMPLIFIED TEST STATISTICS**

When \( \rho_0 = \rho_0 \), in large samples we would have \( \hat{p}_{ij} = \bar{p}_{ij} \) since their limiting values are approximately the same. By the Taylor expansion we have

\[
\ln(\hat{p}_{ij}/\bar{q}_{ij}) - \ln(\bar{p}_{ij}/\bar{q}_{ij}) = (\hat{p}_{ij} - \bar{p}_{ij})/(\bar{p}_{ij}\bar{q}_{ij}), \quad i = 1, 2, \quad j = 0, 1,
\]

where \( \hat{q}_{ij} = 1 - \hat{p}_{ij} \) and \( \bar{q}_{ij} = 1 - \bar{p}_{ij} \).

Recall that the MLEs \( \hat{p}, \hat{p}_{10} \), and \( \hat{p}_{20} \) satisfy the score equations given by

\[
\left\{ (x_{11} - n_{11}\hat{p}_{11}) + (x_{21} - n_{21}\hat{p}_{21}) \right\}/\hat{\theta} = 0, \quad i = 1, 2, \quad j = 0, 1,
\]

where \( \bar{p}_{11} = \rho_0^{-1}(1 + (\rho_0^{-2} - \bar{p}_{10})/(1 - \bar{p}_{0})}. \) Note that

\[
\{\ln(\hat{p}_{11}/\bar{q}_{11}) - \ln(\hat{p}_{10}/\bar{q}_{10})\} - \{\ln(\hat{p}_{21}/\bar{q}_{21}) - \ln(\hat{p}_{20}/\bar{q}_{20})\} = \ln \rho_0.
\]

From Eqs. (8) and (9) we have

\[
n_{11}(\hat{p}_{11} - \bar{p}_{11}) \approx (\ln \hat{O}_1 - \ln \bar{Q}_2 - \ln \rho_0) / \left\{ \sum_{i=1,2} \left( \frac{1}{n_{11}\hat{p}_{11}\bar{q}_{11}} + \frac{1}{n_{00}\hat{p}_{00}\bar{q}_{00}} \right) \right\},
\]

thus

\[
z = (\ln \hat{O}_1 - \ln \bar{Q}_2 - \ln \rho_0) / \left\{ \sum_{i=1,2} \left( \frac{1}{n_{11}\hat{p}_{11}\bar{q}_{11}} + \frac{1}{n_{00}\hat{p}_{00}\bar{q}_{00}} \right) \right\}^{1/2}
\]

\[
= (\ln \hat{O}_1 - \ln \bar{Q}_2 - \ln \rho_0) / \left\{ \sum_{i=1,2} \left( \frac{1}{n_{11}\hat{p}_{11}\bar{q}_{11}} + \frac{1}{n_{00}\hat{p}_{00}\bar{q}_{00}} \right) \right\}^{1/2} = z_S.
\]

**APPENDIX IV: DERIVATIONS OF OPTIMAL DESIGN PARAMETERS**

Let \( C = N \sum_i (\lambda_i \phi_i C_{i1} + \lambda_i (1 - \phi_i) C_{i0}) \) be the total cost, where the \( C_{ij} \) is the cost per subject from ethnic group \( i \) and treatment group \( j \). By setting \( C_{ij} = 1 \) for each \( i, j \) we have \( C = N \). This implies that the optimal design with the minimum
total sample size \( N \) can be derived as a special case of the optimal design with the minimum total cost \( C \). Hence the following derivations deal with cases with general \( C_{ij}s \).

For given \( l_i\)s and \( f_i\)s, recall that the required sample size for the simplified test of significance level \( \alpha \) to achieve a power \( 1 - \beta \) is

\[
\left\{ \frac{z_{(1-\alpha)} + z_{(1-\beta)}}{\ln(\rho_{\alpha}/\rho_0)} \right\}^2 \left[ \sum_{i=1,2} \frac{1}{\lambda_i \phi_i \hat{p}_{i1} \hat{q}_{i1}} + \frac{1}{\lambda_i (1 - \phi_i) \hat{p}_{i0} \hat{q}_{i0}} \right] \].

Further, by Cauchy–Schwartz inequality we have

\[
C = N \left[ \sum_i \{ \lambda_i \phi Ci_{i1} + \lambda_i (1 - \phi_i) Ci_{i0} \} \right] \\
= W \left[ \left\{ \sum_{i=1,2} \frac{1}{\lambda_i \phi_i \hat{p}_{i1} \hat{q}_{i1}} + \frac{1}{\lambda_i (1 - \phi_i) \hat{p}_{i0} \hat{q}_{i0}} \right\} \right] \\
\times \left[ \sum_i \{ \lambda_i \phi Ci_{i1} + \lambda_i (1 - \phi_i) Ci_{i0} \} \right] \geq W \left[ \sum_{i,j} \{ C_{ij}/(p_{ij}q_{ij}) \}^{1/2} \right]^2,
\]

where \( W = \left\{ \frac{z_{(1-\alpha)} + z_{(1-\beta)}}{\ln(\rho_{\alpha}/\rho_0)} \right\}^2 \) and the last equality holds when

\[
\lambda_1 \phi_1 (C_{11} p_{11} q_{11})^{1/2} = \lambda_1 (1 - \phi_1) (C_{10} p_{10} q_{10})^{1/2} = \lambda_2 \phi_2 (C_{21} p_{21} q_{21})^{1/2} = \lambda_2 (1 - \phi_2) (C_{20} p_{20} q_{20})^{1/2},
\]

which together with the relation \( \lambda_1 + \lambda_2 = 1 \), yield the solution given in Eq. (6). The solution (5) for the case of minimal total sample size can be obtained immediately by setting \( C_{ij} = 1 \).

**ACKNOWLEDGMENT**

The authors would like to thank Dr. Jen-Pei Liu for his helpful comments on this manuscript.

**REFERENCES**
