A NEW STATISTICAL PROCEDURE OF INTERVAL HYPOTHESIS TESTING FOR BIOEQUIVALENCE STUDIES

Toshimitsu OHKI, a, b Tokuji SUZUKI, b and Sigeru GOTO b

Kawanishi Pharma Research Institute, Nippon Boehringer Ingelheim Co. Ltd., a 3-10-1, Yato, Kawanishi 666-01, Japan, Faculty of Pharmaceutical Sciences, Chiba University, b 1-33, Yayoi-cho, Inage-ku, Chiba 263, Japan, and Faculty of Pharmaceutical Sciences, Kyushu University, c 3-1-1, Maidashi, Higashi-ku, Fukuoka 812, Japan

A simplified procedure of interval hypothesis testing was investigated and established for bioequivalence studies. The characteristics of this procedure were compared with those of other statistical procedures.

KEYWORDS bioequivalence study; interval hypothesis; noncentral t-distribution; simplified procedure

For bioequivalence studies, statistical analysis has been one of the most important problems to be solved, 1) and several statistical methods have been proposed and applied. 2) The statistical test of the null hypothesis of no difference in a bioavailability (BA) parameter between a reference and a test preparation has often been used. 3) However, the retrospective power of the study must be estimated, and should be more than 0.80 to restrict the probability that a wrong hypothesis is accepted. 4) Another important statistical method is the confidence interval method. 5) In addition, Anderson et al. have proposed a statistical test, in which the null hypothesis states that the test preparation is not equivalent to the reference preparation, and the alternative hypothesis states that the test preparation is equivalent to the reference preparation. 6) We applied the same hypotheses as those by Anderson et al., and investigated the estimation by the noncentral t-distribution and its approximation.

THEORETICAL AND CALCULATION

Formulation of Hypotheses and Statistical Testing The following formulation of the null and the alternative hypotheses is rational for the statistical analysis in two-way crossover design studies:

\[ H_0: \mu_T - \mu_R \leq D \text{ OR } \mu_T - \mu_R \geq D \quad (D > 0) \]  \hspace{1cm} (1)

\[ H_1: \quad -D < \mu_T - \mu_R < D \]  \hspace{1cm} (2)

where \( \mu_T \) and \( \mu_R \) denote the population means of a BA parameter for the reference and the test preparations, respectively, and \( D \) denotes the limiting point of the bioequivalent range. This statistical testing is called "interval hypothesis testing". 7) To test the \( H_0 \) for bioequivalence data which are normally distributed with common variance \( \sigma^2 \), the statistic \( t \), shown by Eq. 3, is considered according to the definition: 8)

\[ t = \sqrt{N/2} \frac{\overline{X}_T - \overline{X}_R}{Ve} \]  \hspace{1cm} (3)

where \( \overline{X}_R \) and \( \overline{X}_T \) denote the sample means of the reference and the test preparations, respectively, \( Ve \) is the variance of error calculated from the analysis of variance, and \( N \) is the number of subjects. This statistic \( t \) belongs to the noncentral t-
distribution with the degree of freedom \( v = N - 2 \) and the non-centrality \( \lambda \) shown by Eq. 4:\(^8\)

\[
\lambda = \sqrt{N/2} (\mu_T - \mu_R) / \sigma
\]

One can judge whether the two preparations are equivalent or not by calculating the probability that \( H_0 \) is true from the estimated \( t \) value based on the noncentral t-distribution. For actual calculation, the non-centrality can be obtained from Eq. 4 by replacing \( (\mu_T - \mu_R) \) and \( \sigma \) with \( 0.2 \bar{X}_R \) and \( \sqrt{\sigma^2} \), respectively.

**Calculation of Probability of Noncentral t-Distribution** In the \( H_0 \), \( \lambda \) is negative when \( \mu_T - \mu_R < -D \), and is positive when \( \mu_T - \mu_R > D \). The probability of the \( H_0 \) for the value of \( t \) is shown by Eq. 5:

\[
P \leq \int_{-\infty}^{t'} f(t'; v, \lambda) \, dt' + \int_{t'}^\infty f(t'; v, -\lambda) \, dt'
\]

where \( f(t'; v, \lambda) \) denotes the probability density function of the noncentral t-distribution. The probability of the \( H_0 \) can be calculated when \( \mu_T - \mu_R \) is equal to \(-D\) or \( D \), and in that case the probability becomes the maximum value. One can test the null hypothesis statistically by using the calculated maximum probability.

**Approximate Calculation** The probability of the \( H_0 \) can be estimated by the approximation with the normal distribution. The approximate formula of Eq. 6 was derived using Fisher's approximation:\(^9\)

\[
P = \Phi \left( (|t| - \lambda) / \sqrt{1 + t^2/(2v)} \right) + \Phi \left( (-|t| - \lambda) / \sqrt{1 + t^2/(2v)} \right)
\]

The probability \( P \) of the \( H_0 \) can be estimated by Eq. 6, where \( \Phi \) denotes the lower probability integral of the normal distribution.

**Accuracy of the Approximation** Probabilities were calculated by the noncentral t-distribution, and by the approximation for assumed data in which \( N \) is 10, \( \bar{X}_R \) is 100, \( \bar{X}_T \) is 100, 95, 90 or 85, and the coefficient of variation of intrasubject error (CV) which is given by Eq. 7 is 10, 20 or 30%.

\[
CV = 100\sqrt{\sigma / \bar{X}_R}
\]

**Comparison between the Proposed Procedure and Other Statistical Methods** The regions of the CV versus \( \bar{X}_T / \bar{X}_R \), which is judged to be bioequivalent, were calculated using a computer for the conventional hypothesis testing (\( H_0 : \mu_T = \mu_R \) at \( \alpha = 0.05 \)) and its retrospective power (1 - \( \beta = 0.80 \) when \( D = 0.2 \bar{X}_R \)) [Method A],\(^10\) the conventional confidence interval (\( \alpha = 0.05 \) and \( D = 20\% \) of \( \bar{X}_R \)) [Method B], and the approximation of the noncentral t-distribution (\( \alpha = 0.05 \) and \( D = 20\% \) of \( \bar{X}_R \)) proposed here [Method C] for the two-way crossover design with ten subjects.

**RESULTS AND DISCUSSION**

**Accuracy of Our Proposed Approximation** The results are shown in Table I. The approximation has sufficient accuracy in the calculation of the noncentral t-distribution, and is a little conservative in the judgment of bioequivalence in comparison with the exact calculation of the noncentral t-distribution.

**Comparison between the Proposed Procedure and Other Statistical Methods** The results are shown in Fig. 1. In case of Method A, when the difference in the sample means of the two preparations becomes large, CV should become large accordingly in order not to indicate the significant difference, and CV are limited to a certain level regardless of the difference in the sample means from the condition of retrospective power. In case of Method B, the maximum CV is allowed when the sample means are equal, and CV must be proportionally small when the
difference in the sample means becomes larger. In case of Method C proposed by us, larger CV is allowed than in other methods when sample means are similar in bioequivalence studies, though some restriction exists near the limiting points of the bioequivalent range. The restriction is relaxed when N and/or the significance level are increased.

The great advantage in Method C is that the probability that a non-equivalent test preparation is accepted as equivalent by mistake is limited by the significance level $\alpha$ of the testing. In actual studies, one cannot always succeed in indicating bioequivalence at the significance level of 0.05. In such cases, one can declare bioequivalence at the significance level of 0.10.

<table>
<thead>
<tr>
<th>$\bar{X}_R$</th>
<th>$\bar{X}_T$</th>
<th>CV</th>
<th>Noncentral t Approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>100</td>
<td>10</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>0.156</td>
</tr>
<tr>
<td>95</td>
<td>10</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.050</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.163</td>
<td>0.165</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>0.023</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.134</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.240</td>
<td>0.246</td>
</tr>
<tr>
<td>85</td>
<td>10</td>
<td>0.173</td>
<td>0.196</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.286</td>
<td>0.303</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.354</td>
<td>0.366</td>
</tr>
</tbody>
</table>

a) Unable to be calculated.

REFERENCES AND NOTES


(Received October 11, 1994; accepted December 12, 1994)

Fig. 1. Comparison of the Bioequivalent Region of CV Versus $\bar{X}_T/\bar{X}_R$ for Two-way Crossover Studies with Ten Subjects

A reversed triangle of three straight lines: Method A, two dotted lines: Method B, and a bold curve: Method C.