DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY®
Vol. 29, No. 4, pp. 469–474, 2003

RESEARCH PAPER

An Alternative Derivation of the Distribution of the Individual Bioequivalence Metric

Sanford Bolton,[1,*] Mohan Sondhi,[2] and Charles DiLiberti[3]

[1]College of Pharmacy, University of Arizona, Tucson, Arizona, USA
[2]Consultant, Mountain Lakes, New Jersey, USA
[3]Barr Laboratories, Blauvelt, New York, USA

ABSTRACT

A method is proposed for estimating the probability density function (PDF) of the individual bioequivalence metric. We show that under the usual assumption of independence of the various random variables, the computations are quite feasible. A program to evaluate the integrals needed for the computations has been written. Experimentation with this program shows that the pass/fail decision based on the computed PDF compares favorably with that based on the method suggested in a recent FDA Guidance (Jan. 2001).

Key Words: Individual bioequivalence; Metric; Probability density function; Confidence limits; Decision rule.

INTRODUCTION

The Office of Generic Drugs (OGD) of the Food and Drug Administration (FDA) has recently published statistical guidelines for determination of bioequivalence.[1] Included in that publication is a statistical approach to determining individual bioequivalence (IBE), as recommended by Hyslop, et al.[2] In this paper, we describe an alternative approach. We determine the probability density function (PDF) of the IBE metric and use it to construct a decision rule for acceptance. The current FDA acceptance criterion for the upper 95% confidence limit for the metric is defined as 2.4948.[1] We show the derivation here for the reference-scaled metric. However, with minor modifications, our approach is also applicable to the constant-scaled denominator metric and to population bioequivalence described in the FDA Guidance.[1]

*Correspondence: Sanford Bolton, College of Pharmacy, University of Arizona, 5495 N. Via Velazquez, Tucson, AZ 85750, USA; Fax: (520) 529-9620; E-mail: bolton@acninc.net.

DOI: 10.1081/DDC-120018381
0363-9045 (Print); 1520-5762 (Online)

The reference-scaled metric is defined as

$$\phi = [(\mu_1 - \mu_r)^2 + \sigma_d^2 + \sigma_t^2 - \sigma_r^2]/\sigma_r^2 \tag{1}$$

or, equivalently as

$$\phi = \left[(\mu_t - \mu_r)^2 + \sigma_d^2 + \sigma_t^2\right]/\sigma_r^2 - 1 \tag{2}$$

Here, μ_t is the mean of the pharmacokinetic parameter for the test product; μ_r is the mean of the pharmacokinetic parameter for the reference product; σ_d^2 = subject-product interaction variance; σ_t^2 = within-subject test variance; σ_r^2 = within-subject reference variance.

For a four-period replicate design as described by Hyslop et al.[2] and in the FDA Guidance,[1] we can also define Eq. (3)

$$\sigma_i^2 = \sigma_d^2 + 0.5\sigma_t^2 + 0.5\sigma_r^2 \tag{3}$$

where σ_i^2 is the variance of $(\mu_t - \mu_r)$
Combining Eqs. (2) and (3),

$$\phi = [(\mu_t - \mu_r)^2 + \sigma_i^2 + 0.5\sigma_t^2]/\sigma_r^2 - 1.5 \tag{4}$$

The parameter estimates, $\overline{X_t}$ $\overline{X_r}$, S_i^2, S_t^2, and S_r^2, are computed using a mixed-effects linear model as described in the FDA Guidance.[1,3]

At an earlier stage in its development, a bootstrap approach was used to estimate the upper confidence limit for ϕ. This approach was time consuming and not exactly reproducible. The analysis in the recent guidance is approximate, has reasonably good properties,[1,2] and is relatively simple to calculate. It appears to agree well with the results of the bootstrap simulation approach.

This brief communication gives a more direct approach to estimating the upper confidence limit for ϕ. The idea is to derive the PDF of the metric ϕ. Once the PDF is known, the cumulative probability distribution function (CDF), the upper 95% confidence limit, as well as other parameters of interest can be easily determined.

DERIVATION AND RESULTS

In principle, the PDF of ϕ can be determined if the joint distribution of the random variables $\overline{X_t}, \overline{X_r}, S_i^2, S_t^2$, and S_r^2 is known. In general, this would be a formidable task. However, under the usual assumption of statistical independence of these variables,[2] it is quite feasible to compute the PDF of ϕ. We will make the further assumption that the random variables $\overline{X_t}$ and $\overline{X_r}$ are Gaussian and that the variances, S_i^2, S_t^2, and S_r^2, are distributed as $\sigma^2\chi^2_{\mathrm{df}}/\mathrm{df}$, where df is the appropriate degrees of freedom. With these assumptions, which are similar to those made by Hyslop et al.,[2] the PDF of ϕ can be derived as shown below. In the derivation, we have used the formulae for computing the PDF of the sum of two independent variables and the PDF of the ratio of two independent variables. These may be found in Ref.[4]

For ease of notation, define the following random variables:

$$Y = (\overline{X}_t - \overline{X}_r)^2$$
$$Z = S_i^2$$
$$U = 0.5\ S_t^2$$
$$V = S_r^2$$

In terms of these, define further the intermediate variables:

$$W = Y + Z$$
$$G = W + U$$

The metric may then be expressed as

$$\phi = \frac{G}{V} - 1.5$$

Since $\overline{X_t}$ and $\overline{X_r}$ are both Gaussian, their difference is also Gaussian. Let the mean and standard deviation of $(\overline{X_t} - \overline{X_r})$ be μ and σ, respectively. Then the PDF of Y, $p(y)$, is given by

$$p(y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{y+\mu^2}{2\sigma^2}\right)\frac{1}{\sqrt{y}}\ \cos h\left(\frac{\mu\sqrt{y}}{\sigma^2}\right), \quad \text{for } y \geq 0 \tag{5}$$

Equation (5) is derived in Appendix 1.

Returning now to the derivation of $a(m)$, let $q(z)$ be the PDF of Z. Since Y and Z are independent, the PDF of W, $r(w)$, is given by the convolution of $p(y)$ and $q(z)$. Thus

$$r(w) = \int_0^w p(y)q(w-y)\,dy$$

Similarly, if $s(u)$ is the PDF of U, then the PDF of the variable G, $f(g)$, is given by

$$f(g) = \int_0^g r(w)s(g-w)\,dw$$

Finally, let $a(m)$ be the PDF of ϕ. If $t(v)$ is the PDF of V, then

$$a(m) = \int_0^\infty vt(v)f[(m+1.5)v]\,dv \tag{6}$$

The derivation of Eq. (6) is shown in Appendix 2.

A program was written in MATLAB[5] to evaluate $a(m)$ using numerical integration to compute the various integrals in the above derivation. The program code is shown in Appendix 3. (With a personal computer, the computation is complete in less than 10 sec. The code will be e-mailed to readers upon request at bolton@acninc.net.) If the parameters defining the distributions of $\overline{X_t}$, $\overline{X_r}$, etc., were known, this would be an exact solution. In the absence of such knowledge, an approximate solution is obtained by using the observed values of the means and variances as the parameter values. Clearly, this solution would approach the exact solution with large sample sizes. With the sample sizes usually used in BE studies, $n \geq 12$, we expect that the solution should be reasonably good. A preliminary spot check of the results and decisions comparing our approach to that of Hyslop et al.[2] is shown in Table 1. Examples are shown where the decisions are borderline, and the agreement for passing and failing is good. When the pass/fail decision is obvious in either direction, there should be 100% agreement based on the close agreement when the pass/fail decision is borderline. In general, the agreement is excellent, and it improves with larger sample sizes as expected; both derivations will give close to an exact solution with large sample sizes. Figure 1 shows an example of the cumulative distribution of ϕ for the specified parameter values shown. Note that the parameters used to compare the present method to Hyslop's method were prespecified and not generated. Both methods calculate a pass/fail decision based only on an estimate of parameters, not on simulations.

We are presently investigating the effects of sample size and parameter size on the properties of the results obtained by our method.

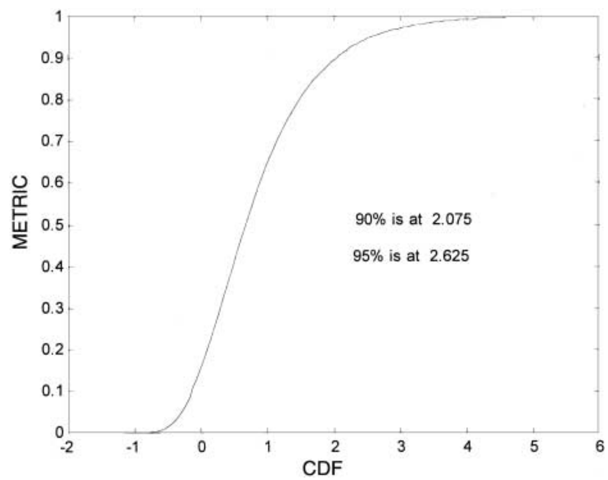

Figure 1. Cumulative distribution function estimated for $N = 26$ $(\mu_t - \mu_r)^2 = 0.0025$, $\sigma_i^2 = 0.08$, $\sigma_t^2 = 0.049$ and $\sigma_r^2 = 0.05$. X axis label METRIC; Y axis label CDF.

Table 1. Comparison of results of proposed method to Hyslop method for assessing bioequivalence for various parameter estimate values.

N [a]	Mean	S_i^2	S_t^2	S_r^2	Hyslop[b]	Proposed method[c]
122	0	0.02	0.02	0.0125	−0.0286 (P)[d]	1.635 (P)[d]
	0	0.02	0.02	0.01	−0.0007 (P)	2.46 (P)
	0	0.02	0.03	0.01	+0.0046 (F)	2.79 (F)
	0.2	0.12	0.12	0.065	+0.0226 (F)	2.90 (F)
26	0.05	0.12	0.1	0.085	−0.0153 (P)	2.46 (P)
	0.0	0.02	0.02	0.015	−0.0037 (P)	2.35 (P)
	0.05	0.08	0.049	0.05	+0.0060 (F)	2.625 (F)
	0.2	0.12	0.12	0.095	+0.0136 (F)	2.955 (F)
16	0.07	0.04	0.02	0.04	−0.0223 (P)	1.91 (P)
	0.05	0.05	0.05	0.05	−0.0173 (P)	2.24 (P)
	0.05	0.04	0.02	0.03	+0.0022 (F)	2.845 (F)
	0.05	0.03	0.02	0.02	+0.0128 (F)	3.725 (F)
12	0	0.04	0.02	0.0475	−0.0324 (P)	1.69 (P)
	0.05	0.03	0.01	0.03	−0.0087 (P)	2.295 (P)
	0	0.05	0.04	0.0475	+0.0004 (F)	2.85 (F)
	0.07	0.05	0.04	0.0475	+0.0082 (F)	3.175 (F)

[a]Two sequences of equal size ($N/2$).
[b]Hyslop et al. method[2] passes for negative values.
[c]Proposed method passes for values less than 2.4948.
[d]P = pass IBE confidence limit criterion. F = fail IBE confidence limit criterion.

SUMMARY

An alternative approach to determining confidence limits for the IBE metric is proposed. This approach may have equally good or better properties than the solution presently recommended.[1] It has the advantage of being a more general and direct approach to the problem, as well as defining the PDF. The properties of this method for defining the upper confidence limit need to be evaluated under different conditions of sample size and parameter estimates.

APPENDIX 1

$p(y)$ may be derived as follows:

For brevity, define the random variable $X = \overline{X_t} - \overline{X_r}$

Let $g(x)$ denote the Gaussian PDF with mean μ and standard deviation σ, i.e.,

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

which was assumed to be the PDF of X. Then the cumulative distribution of Y, $C(y)$, is given by

$$\begin{aligned} C(y) &= \text{Prob}(Y \leq y) \\ &= \text{Prob}(X^2 \leq y) \\ &= \text{Prob}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} g(x)\,dx \end{aligned}$$

Now, the PDF of Y is just the derivative of $C(y)$ with respect to y. Hence,

$$P(y) = \frac{1}{\sqrt{y}}\left[\frac{g(\sqrt{y}) + g(-\sqrt{y})}{2}\right]$$

When the Gaussian function is substituted for g in this equation, we get the PDF of Y (Eq. 5)

APPENDIX 2

The PDF of ϕ may be derived as follows:

Let $A(m) = \text{Prob}(\phi \leq m)$

Then,

$$\begin{aligned} A(m) &= \text{Prob}(G \leq (m+1.5)V) \\ &= \int_0^{\infty} dv \int_0^{(m+1.5)v} t(v) f(g)\,dg \end{aligned}$$

Since the PDF of ϕ is the derivative of the cumulative distribution, we get Eq. (6), $a(m) = \int_0^{\infty} vt(v) \times f[(m+1.5)v]\,dv$.

APPENDIX 3

Matlab Program to Compute CDF of Metric

Example

```
Data from Replicate Design with 8 subjects
A. mean of xt-xr is the estimated dierence between the means of Test and Reference = -0.1715
B. std err of xt-xr is the standard error of (A) = 0.295
C. si^2 is the estimated interaction variance = 0.0871
D. sr^2 is the estimated within subject variance of the reference = 0.06605
E. st^2 is the estimated within subject variance of the test = 0.0729
```

Matlab Program

```
clear
global degfree
global mean
%These variables are used to compute chisq/df distributions
%using chidf.m
%
mtr = -.01715       % mean of xt-xr
str = 0.295         % std err of xt-xr
msi = 0.0871        % si^2
msr = 0.06605       % sr^2
mst = 0.0729        % st^2
%mit = msi+.5*mst % mean of si^2 + .5 st^2
%sit = sqrt(ssi^2 + (.5*sst)^2)
%
%Note: si^2, sr^2 and st^2 are distribu-
       ted as sig^2 chi^2/d.f
%      where the degrees of freedom
       are 6.
%      Therefore the mean = sig^2
%      sig^2 is, of course, unknown. We
       approximate it by the sample
%      estimate
```

```
%       we need the distribution of
%       [(xt-xr)^2 + si^2 + .5st^2 ] /
         sr^2 -1.5
%prob density of (xr-xt)^2 is ptr2
%prob density of si^2 is psi
  (deg of free =6)
%prob density of sr^2 is psr
%prob density of st^2 is pst
%prob density of si^2+.5st^2 is pit
%prob density of denominator is pden
    (= prob density of sr^2)
%prob density of numerator is pnum
%prob density of metric is pmetric
%
xtr2max=0.15
xchimax=0.4
deltr=.00005
ntr2=1+round(xtr2max/deltr)
nchi=1+round(xchimax/deltr)
xtr=linspace(0.,xtr2max,ntr2);
xchi=linspace(0.,xchimax,nchi);
ctr=(1/(str*sqrt(2*pi)))*
    exp(-mtr^2/(2*str^2));
ptr2(1)=2*ctr/sqrt(deltr);


for i=2:ntr2
  xdel=xtr(i);
  xsqrt=sqrt(xdel);
  ptr2(i)=ctr*exp(-xdel/(2*str^2))*
 (1/xsqrt)*cosh(mtr*xsqrt/str^2);
end
sumtr2=deltr*ones(1,ntr2)*ptr2'
ptr2=ptr2/sumtr2;
plot (xtr,ptr2)
title('prob density of (xt - xr)^2')
text(.03,250,'xt-xr is gaussian')
text(.03,225,['mu=  Eq  'num2str(mtr),
', sigma= ', num2str(.129)])
iplot=0


if(iplot==1)
  print -dwin
end
%
mean=msi
degfree=6
psi=chidf(xchi);
sumpsi=deltr*ones(1,nchi)*psi'
plot (xchi,psi)
title('prob density of si^2')
text(.3,5,['sig^2= 'num2str(msi)])
text(.3,4,['df = 'int2str(degfree)])
if(iplot==1)
  print -dwin
end
mean=msr
degfree=6
psr=chidf(xchi);
sumpsr=deltr*ones(1,nchi)*psr'
plot(xchi,psr)
title('prob density of sr^2')
text(.3,8,['sig^2= ' num2str(msr)])
text(.3,7,['df = 'int2str(degfree)])

if(iplot==1)
  print -dwin
end
mean=mst*.5
degfree=6
pst=chidf(xchi);
sumpst=deltr*ones(1,nchi)*pst'
plot(xchi,pst)
title('prob density of .5*st^2')
text(.3,8,['sig^2= 'num2str(mst)])
text(.3,7,['df = 'int2str(degfree)])

if(iplot==1)
  print -dwin
end
%
psist= deltr*conv(psi,pst);
[junk,npl]=size(psist)
xpl=linspace(0.,2.*xchimax,npl);
plot(xpl,psist)
title('prob density of si^2+.5st^2')
sumsist=deltr* ones(1,npl)*psist'

if(iplot==1)
  print -dwin
end
pnum=deltr*conv(psist,ptr2);
[junk,nplnum]=size(pnum)
xnum=linspace(0.,2.*xchimax+
  xtr2max,nplnum);
for i=1:nplnum
  cumnum(i)=deltr*ones(1,i)*
    pnum(1:i)';
end

plot(xnum,pnum)
sumnum=deltr*ones(1,nplnum)*pnum'
title('prob density of numerator')

if(iplot==1)
  print -dwin
```

```
end
maxplot=15.
nplot=301
xmetric = linspace(-1.5,maxplot,
    nplot);
for i=1:nplot
  factor=xmetric(i)+1.5;
  for k=1:nchi
    index=round((factor*xnum(k)/
      deltr)+1);
    if(index > nplnum)
      index=nplnum;
    end
    integrand(k)=xnum(k)*
      psr(k)*pnum(index);
    cumint(k)=psr(k)*cumnum(index);
  end
  pmetric(i)=deltr*ones(1,nchi)*
      integrand';
  cumplot(i)=deltr*ones(1,nchi)*
      cumint';
end
plot(xmetric,pmetric)
title('prob density of metric')
summetric=cumplot(nplot)
%summetric=(maxplot/(nplot-1))*
    ones(1,nplot)*pmetric'

if(iplot==1)
  print -dwin
end
nstop95=1
nstop90=1
for i=1:nplot
%   cumplot(i)=(maxplot/(nplot-1))*
          ones(1,i)*pmetric(1:i)';
    if cumplot(i)>.90 & nstop90 > 0
      cum90 = xmetric(i);
      nstop90=0
    end
    if cumplot(i)>.95 & nstop95 > 0
      cum95=xmetric(i);
      nstop95=0
    end
end
iplot=1
plot(xmetric,cumplot)
title('cumulative distribution of
ratio')
text(10,.5,['90% is at 'num2str
    (cum90)])
text(10,.4,['95% is at 'num2str
    (cum95)])

if(iplot==1)
  print -dwin
end
%distribution of chi^2/d.f. when mean is
not unity
function y=chidf(x)
global degfree
global mean
n=degfree;
n2=n/2;
n2m=n2-1;
u=degfree*x/mean;
y=(n/mean)*   (1/((2^n2)*gamma(n2)))*
u.^n2m.*exp(-u/2);
end
```

REFERENCES

1. *Guidance for Industry*, *Statistical Approaches to Establishing Bioequivalence*. Food and Drug Administration, CDER, January, 2001.
2. Hyslop, T.; Hsuan, F.; Holder, D.J. A small sample confidence interval approach to assess individual bioequivalence. Statistics in Medicine **2000**, *19*, 1885–2897,
3. Ekbohm, G.; Melander, H. The subject-by-formulation interaction as a criterion for interchangeability of drugs. Biometrics **1989**, *45*, 1249–1254.
4. Rice, J.A. *Mathematical Statistics and Data Analysis*; Wadsworth and Brooks/Cole Statistics/Probability Series: Pacific Grove, CA, 1988.
5. *Matlab*; The Mathworks, Inc.: Natick, MA.