Bayesian Analysis of Population Bioequivalence Using the Independence Chain Algorithm in PROC MIXED
Richard McNally, Colorado State University, Fort Collins, CO

ABSTRACT
The statistical test for population bioequivalence given in the current FDA guidance document on this matter ignores the dependence between summary statistics. It is very conservative in cases where there is a high degree of correlation of subject responses between test and reference formulations, which is usually the case in bioequivalence studies. Adapting the ideas of Kass and Wolfinger (2000), we use the PRIOR statement in PROC MIXED to approximate the distribution of the population bioequivalence parameter $\Theta_{PBE}$, from which we can construct tests based either on the probability of the upper tail region (i.e., $P(\Theta_{PBE} \geq \Theta_0)$), or on the one-sided upper 95% confidence bound of $\Theta_{PBE}$. We show that this test can conclude population bioequivalence when the FDA method does not. We also examine the statistical properties of this approach. The Kass and Wolfinger approach is also compared to another alternative given by McNally et al which is also more powerful than the FDA method, but is slightly anti-conservative.

INTRODUCTION
On 31 January 2001, the US Food and Drug Administration (FDA) issued a new guidance document Statistical Considerations for Bioequivalence Studies (2001). This document codified the requirement that bioequivalence studies included the assessment of population and individual bioequivalence, in addition to average bioequivalence. Two formulations of the same drug product are average bioequivalent if their mean difference is within allowed limits with reasonable assurance. More specifically, two products are average bioequivalent if the 90% confidence interval of $\mu_{T-R}$ is completely inside the interval $(-0.22314,0.22314)$, assuming that the original data has been log-transformed. This test procedure is equivalent to the Two One-Sided Tests proposed by Schuirmann (1987), the formal statement of this is contained in the 1992 FDA guidance document on statistical analyses of bioequivalence studies. But even as it was being formalized, the adequacy of average bioequivalence (ABE) alone to assure that patients whose prescriptions are switched from a brand name drug to a generic product would receive the same therapeutic effect was being questioned. Also questioned was whether ABE was adequate to assure that a patient who received a generic product as a new prescription would receive the same therapeutic benefit as with the brand-name product. In response to these concerns, the respective concepts of individual bioequivalence (IBE) and population bioequivalence (PBE) were developed. The recent FDA guidance document on bioequivalence gives criteria and suggested testing methods for IBE and PBE. The test for IBE is from Hyslop, Hsuan, and Holder (2000), who constructed a method for calculating an upper 95% confidence bound based on the FDA criteria. In the guidance document, the agency adapted this procedure to construct a test for PBE. Unfortunately, some of the summary statistics used to construct the FDA test are not statistically independent, whereas in Hyslop et al, the summary statistics are independent. It appears that, when observations within a subject are highly correlated, which is usually the case in bioequivalence studies, the FDA test is very conservative. In fact, it is possible that this test will not conclude for the same data that provides evidence of IBE. This appears to the conceptual hierarchy that IBE is a higher standard than PBE, which in turn is a higher standard than ABE.

One approach that deals with the dependency problem was developed by McNally, Iyer, and Mathew (2001). This approach applied the generalized $p$-value (GPV) methodology introduced by Tsui and Weerahandi (1989). Using the GPV, McNally et al were able to develop a more powerful test for the PBE criterion, although this test appeared at times to be slightly anti-conservative. The present investigation examines a second approach. We extend the work of Kass and Wolfinger (2000), who applied a variant of Monte Carlo Markov chain (MCMC) methodology, the independence chain (IC) algorithm, to construct confidence intervals (which Bayesians would call credible regions) for individual variance components and general functions of these components. Best of all, the IC algorithm can be implemented using the PRIOR statement in PROC MIXED. Applying the IC algorithm to the PBE criterion produces a test that appears to have properties similar to those of the GPV, and is much more powerful than the FDA test.

STATISTICAL DESIGN AND MODEL
Consider the following 2-sequence/4-period crossover design

<table>
<thead>
<tr>
<th>Period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence 1</td>
<td>T</td>
<td>R</td>
<td>T</td>
<td>R</td>
</tr>
<tr>
<td>Sequence 2</td>
<td>R</td>
<td>T</td>
<td>R</td>
<td>T</td>
</tr>
</tbody>
</table>

where $T$ represents the test (generic) formulation and $R$ represents the reference (innovator) formulation. To assess population bioequivalence for data from this study design, we use the following model

$$
Y_{jk} = \mu_k + \eta_{jk} + \epsilon_{jk}
$$

where $i=1,2$ is the sequence number, $j=1..n_i$ is the $j$th subject in sequence $i$, $k=T,R$ is the treatment, and $i=1,2$ is the replicate number. That is to say, $Y_{j1k}$ is the response of subject $j$ in sequence $i$ receiving replicate 1 of treatment $k$. The fixed effect of treatment $k$ is $\mu_k$, and the fixed effect of cell $ik$ is $\gamma_{ik}$. The random subject effect is $\eta_{jk}$, and the random within-subject effect is $\epsilon_{jk}$. The random subject effect has a bivariate normal distribution for each subject with a zero mean vector and the following covariances

$$
\text{Var}(\eta_{ijT}) = \sigma^2_{\eta T} \quad \text{Var}(\eta_{ijR}) = \sigma^2_{\eta R} \quad \text{Cov}(\eta_{ijT}, \eta_{ijR}) = \rho \sigma_{\eta T} \sigma_{\eta R}
$$

The within-subject effect has a normal distribution with mean 0 and variance $\sigma^2_{\epsilon_{jk}}$ (k is the treatment). This is the statistical model used by both Hyslop et al and the FDA guidance document for analyzing PBE and IBE data. For details on this model, see Chinchilli and Esinhart (1996).

The FDA test for PBE is based on the following summary statistics

$$
U_{iT} = (Y_{iT1} + Y_{iT2})/2 \quad U_{iR} = (Y_{iR1} + Y_{iR2})/2
$$

$$
V_{iT} = (Y_{iT1} - Y_{iT2})^2/2 \quad V_{iR} = (Y_{iR1} - Y_{iR2})^2/2
$$

The problem with the FDA test arises with the joint distribution of $U_{iT}$ and $U_{iR}$; they have covariance $\rho \sigma_{\eta T} \sigma_{\eta R}$, so that they are independent only when $\rho = 0$. Now Hyslop et al (2000) computed a confidence interval for the IBE criteria, which is a linear combination of independent components. Their exposition is
based on the work of Howe (1974), who computed a confidence interval for the difference of 2 independent mean squares. But consider the PBE criterion
\[
\Theta_{PBE} = \frac{\delta^2 + \sigma_{TT}^2 + \sigma_{TR}^2}{\max(\sigma_{TR}^2, \sigma_{TT}^2)}
\]  
(2)
where \(\delta=\mu_{T-TR}^2\), \(\sigma_{TT}^2=\sigma^2_{at}+\sigma^2_{wt}\) and \(\sigma_{TR}^2=\sigma^2_{ar}+\sigma^2_{wr}\). By ignoring the dependence noted above, the FDA test appears to put much too high of a bound on linear combinations involving \(\sigma^2_{at}\) and \(\sigma^2_{ar}\), so that the FDA test is very conservative. The same is true of the test based on the confidence intervals developed by Quiroz et al. (2001). McAuliffe et al. (2001) solved this problem by decomposing the covariance matrix of the joint normal distribution of the \((U_{T},U_{TR})\) pairs, which produced a test more powerful than the FDA test. This test was based on the generalized p-value (GPV) methodology of Tsui and Weerahandi (1989). Tsui and Weerahandi noted similarities between the numerical calculations given by the GPV method for a given problem (e.g., the Behrens-Fisher problem) and by various Bayesian formulations of the same problem. These remarks raised our interest in approaching the PBE problem from a Bayesian perspective.

Wolfinger and Kass (2000) approached the problem of calculating confidence intervals for general functions of variance components using a Monte Carlo Markov Chain (MCMC) method called the independence chain algorithm (IC). The idea is that in certain mixed models (e.g., the one-way random-effects model with balanced group sizes) the mean squares have well-defined (i.e., Chi-squared) distributions. Therefore, we can implement MCMC by sampling directly from the posterior distribution of the random effects. Once a set of random effects is generated, they can be plugged into a multivariate normal likelihood to generate random realizations of the fixed effects. In other words, the posterior of the fixed effects is the likelihood times a flat prior \(f(\beta)=1\). The base distribution (i.e., the distribution that generates MCMC candidate points) for the random-effects is exact, so that MCMC samples are rejected only if they are not in the parameter space, which may occur, for example, if an individual variance component is estimated to be negative. Even for modestly unbalanced data, the base distribution is usually a good approximation to the true prior. This algorithm is implemented in SAS using the PRIOR statement in PROC MIXED.

The FDA guidance document recommends that sponsors use method of moments (MoM) estimators for the mean difference for IBE and PBE. The MoM relies heavily on the assumption of normality, whereas MoM does not. But Chinchilli and Eshihart showed that, for the model given above for the 4-period crossover design, the MoM and the restricted REML estimates are the same (the unrestricted REML estimates for the random-subject effects are those given by using type=un in the RANDOM statement). Therefore, the distinction between REML and MoM may seem unimportant here. This could be considered auspicious, since the posterior that we are using to generate MCMC samples is conditioned on the REML estimates. The MCMC samples are also generated within the context of normal linear models theory, which may appear to be at odds with regulatory reluctance to assume normality. But the only inferential method proposed for PBE that does not use normality for inference is the nonparametric bootstrap.

**INDEPENDENCE CHAIN ALGORITHM**

The independence chain (IC) algorithm is a special case of rejection sampling. See Tierney (1994) for a good overview of MCMC methodology, including rejection sampling and, in particular, the independence chain.

---

**Model (1)** is a special case of the general linear mixed model
\[
y=X\beta + Z\gamma + \epsilon
\]  
(3)
where \(y\) is a vector of observed responses, \(X\) is the design matrix corresponding to the fixed-effects parameter vector \(\beta\), \(Z\) is the design matrix corresponding to the random-effects parameter vector \(\gamma\), and \(\epsilon\) is the residual error vector. We assume that \(\gamma\) and \(\epsilon\) are independent and have multivariate normal distributions with mean 0 vectors and covariance matrices \(G\) and \(R\), respectively.

In the variance component setup, \(G\) and \(R\) are diagonal matrices (i.e., the effects associated with the variance components are independent), but for the population bioequivalence problem we will need \(G\) to be a general covariance matrix. Let \(\theta\) be the vector of variance components, i.e., the components of \(G\) and \(R\). Since we are working in the Bayesian framework, the analysis revolves around the posterior distribution of the model parameters. The joint posterior density of \((\beta,\gamma,\theta)\) is
\[
f(\beta,\gamma,\theta|y)=f(\beta,\gamma|\theta, y)f(\theta|y)
\]  
(4)
The basic procedure is to simulate the density \(f(\theta|y)\) of the variance components using an independence chain, then, conditioning on \(\theta\), to simulate from \(f(\beta,\gamma|\theta, y)\) as a multivariate normal distribution. We assume, with Wolfinger and Kass, that \(\beta\) has a uniform prior \(f(\beta)=1\). Now
\[
f(\theta|y)=f(\theta|y)f(\theta)
\]  
(5)
is the posterior on \(\theta\). The prior distribution \(f(\theta)\) that we use is the Jeffreys’ prior (Jeffreys’ prior is actually a class of uninformative and often improper prior distributions).

Wolfinger and Kass summarized the IC algorithm for variance components in the following way:

1. Transform the variance components to approximate independence using the ratios established in the MIVQUE(0) coefficient matrix.
2. Determine inverted-gamma densities for each transformed parameter by performing a linear regression on a series of grid evaluations of the log-marginal posterior of the variance components, \(f(\theta|y)\). This posterior is computed by taking the product of the restricted likelihood and some prior.
3. Form the base sampling density for the transformed variance components, \(g(\theta|y)\), by taking the product of the inverted gamma densities from step 2.
4. Use an independence chain to generate a pseudo-random sample from \(f(\theta|y)\).
5. Given sampled values for the variance components, use the multivariate normal distribution to generate observations for the mean [i.e., fixed] parameters.

In the independence chain algorithm (step 4), MCMC proposals are drawn from a fixed based sampling density \(g(\theta|y)\). Since the proposal density is exactly the posterior for model (1), the acceptance probability of a MCMC proposal should be close to 1. What complicates the situation for population bioequivalence is that the functions of the variance components in which we are interested are not independent. So we modify the above method by allowing an inverted-Wishart density to be used in step 2. This should give tests constructed based on the IC algorithm an advantage over tests that do not take statistical dependence into account.

Since we are interested in tests of hypothesis, we use the posterior distribution to construct tests for population bioequivalence. For a general function \(T(\beta, \theta)\) of the mixed model parameters, if \(H_0: T(\beta, \theta) > T_0\) is the null hypothesis, then we can
compute an estimate of the 95-percentile of this function and compare it to the pre-specified upper bound $T_0$. The values less than this percentile would form a one-sided 95% credible region, the Bayesian equivalent of a confidence interval, for $T([\beta, \theta])$. An alternative is to compute the area of the extreme or tail region, i.e., $D(X) = Pr(T(\beta, \theta) > T_0)$, where $D(X)$ is a function of the data, called the discrepancy of the observed data (Meng, 1994). This procedure is similar to the calculation of a frequentist p-value, and is even more similar to the generalized p-value of Tsui and Weerahandi (1989), the methodology on which McNally et al. based their test of population bioequivalence. Calculations of this type are not generally accepted by Bayesians, but Rubin (1984) defended this type of test procedure as being Bayesiantly justifiable under certain circumstances, and cited the use of tail probabilities by Box and Tiao to evaluate model differences. Rubin subsequently gave this measure of discrepancy a name, the posterior-predictive p-value. Meng develops some of the theory of posterior-predictive p-values, and he also notes their similarity to generalized p-values.

**IMPLEMENTING THE IC ALGORITHM IN PROC MIXED**

The following is the SAS® code used to implement the IC algorithm.

```sas
data pbeparmtrans;
  input covp1-covp5;
  cards;
  2 0 0 1 0 2 0 0 0 2 0 1 0 0 0 1 0 0 0 0 1
;;

proc mixed data=pbemph;
  class treatment;
  model lnauc=treatment gamma1-gamma6 / ddfm=satterth;
  random treatment/subject=subject type=un g;
  estimate 'F' treatment -1 1 / cl alpha=0.1;
  repeated/group=treatment;
  ddfm=satterth;
  model lnauc=treatment gamma1-gamma6 /
    type=un
  ;;;
run;

```

```sas
data pbe; set sample;
  s=0.04;
  theta0=(log(1.25))*2 + 0.02)/s_0;
  theta=(est1*est21-covp1+covp3-
    covp4+covp5)/max(covp1+covp4,s_0);
  psi=(thetap gt theta0);
run;

proc univariate data=pbe;
  var thetap psi;
run;
```

The first data step is needed so that PROC MIXED uses the proper transformation of the covariance parameters. The default action is for the procedure to generate this transformation automatically. For balanced random-effects models, the coefficients of this transformation are those of the estimated mean squares. In general, these transformations are the MIVQUE(0) equations, normalized by dividing through by the coefficient of the residual effect. But in the population bioequivalence model, the residual effects depend on treatment, so the procedure cannot properly generate this transformation. Fortunately, the `tdata=parmtrans` option in the PRIOR statement allows the user to specify the correct transformation.

The **PROC MIXED** call fits the model in equation (1) above to the log-transformed pharmacokinetic data. Here the pharmacokinetic parameter is area under the plasma-concentration/time curve (AUC). The `gamma1-gamma6` are indicator variables that must be created in a data step. Since the model (1) has a mean effect for each treatment, there are estimability conditions on the gammas. Chinchilli and Esihant have the details on these conditions. Using `type=un` in the RANDOM statement gives the method of moments estimates for the variance components in this design. The `group=treatment` option in the REPEATED statement allows us to have separate residual terms for each treatment group. The use of the `group` option is what necessitates the use of the `tdata` option in the PRIOR statement. The treatment contrast in the ESTIMATE statement is needed so that estimates of the $\delta$ parameter in (2) will be included in the MCMC samples. Here the ESTIMATE statement options request a 90% confidence interval for the estimate of $\delta$. This is considered equivalent to the two one-sided tests for average bioequivalence of Schuirmann (1987).

The **PRIOR** statement implements the MCMC sampling routine in PROC MIXED. Here we are using the default prior distribution in MCMC sample generation, which is the Jeffreys’, or uninformative, prior. The user can specify their own prior by creating a data set with the distribution name and the parameters of the prior, which is called using `data=` option. An example of a user-specified prior is given in Wolfinger and Kass. The user can also specify a flat prior, $f(\theta)=1$, in which case the posterior used to generate the MCMC samples for the variance component is the likelihood function. We have also specified the following options on the **PRIOR** statement. The option `alg=ic` specifies that the independence chain algorithm is to be used to generate MCMC samples as described in the previous section. Technically, this is not needed as the independent chains is the default. The `nsample=1000` statement specifies that 10000 MCMC samples be generated and put into the `bayes` data set specified in the **OUT=** statement. The default number of samples is 1000, but Wolfinger and Kass recommend 10000 samples in order to get more accurate results. We follow the progress of the sample generation using the `lognote=100` option, which causes a message to be printed on the LOG for every 1000 samples generated. We have already discussed the use of `tdata=parmtrans`. Other useful options that give information on the generation of MCMC samples can be found in the SAS online help on **PROC MIXED**.

The data step that follows computes the population bioequivalence function, and both the boundary and the indicator of the tail region. In this data step `est1=delta, covp1=sigma_BR, covp2=delta_BR, covp3=sigma_BT, covp4=sigma_WR, and covp5=delta_WT`, which can be used to calculate estimates of the population bioequivalence criterion (equation (2)). We then use **PROC UNIVARIATE** to compute the percentiles of the population bioequivalence function and the posterior-predictive p-value.

**EXAMPLE: METHYLPHENIDATE**

Meyer et al. (2000) studied the individual bioequivalence of brand name and generic versions of methylphenidate (better known by the brand name Ritalin™), which was evaluated using the method of Hyslop et al. (2000), whose methodology is recommended by the FDA. McNally et al. (2001) also examined the individual and population bioequivalence of these of these drug formulations. In examining the pharmacokinetic metric Area Under the concentration-time Curve (AUC), McNally et al. used the generalized p-value (GPV) to conclude that the brand name and generic formulations were population bioequivalent, contrary to what would be concluded according to the methodology recommended in the FDA guidance document (2001), and also contrary to the conceptual hierarchy that IBE is a higher standard than PBE, since Meyer et al. concluded the formulations were
individual bioequivalent with respect to AUC. Bootstrap analysis seemed to confirm the conclusions of McNally et al. The difference between the GPV and FDA tests is that the GPV test accounts for the dependence between summary statistics, which appear to yield a more powerful test for population bioequivalence. Since the independence chain methodology also accounts for dependence (technically, in the Bayesian context the dependence is between the variance components), its performance should be similar to that of the GPV test.

The parameters for the base density are $v=16$, $\sigma_1=4.1570$, $\sigma_2=3.1467$, $\alpha_2=2.9777$ for the inverted Wishart, $\alpha=8$ and $\beta=0.4150$ for the inverted gamma corresponding to $d_{WR}^V$, and $\alpha=8$ and $\beta=0.2723$ for the inverted gamma corresponding to $d_{WR}^V$. The values of parameters of the inverted gamma distribution are, respectively, one-half the degrees of freedom for calculating the variance components and one-half the sum of squares of the corresponding variance component. Table 1 summarizes the methylphenidate AUC population bioequivalence results. The first group of numbers are upper 95% confidence bounds based, respectively, on the generalized confidence interval (GCI), the independence chain (ICCR), and the FDA method. The upper bound for both the GCI and ICCR are similar and well under the regulatory upper bound of 1.7448, so that we can conclude population bioequivalence from both methods. The FDA interval is on a different scale, which requires that the upper bound be less than 0 in order to conclude population bioequivalence, which it is not in the table. Therefore, from the FDA test we would draw a conclusion contrary to that of the other 2 methods. The next 2 numbers are the areas of tail or extreme regions (i.e., $Pr(\theta<1.7448)$), which are similar to the conventional frequentist p-value. The first is generalized p-value (GPV); the second is the posterior predictive p-value (PPP). In practice, these values can be used in the same manner as conventional p-values to assess statistical hypotheses. Again, the values are similar and both lead credence to the belief that the 2 formulations of methylphenidate are population bioequivalent. The FDA method does not yield anything comparable to a p-value.

The power of the IC test for population bioequivalence was evaluated for $\delta=0.05$, $\sigma_{BR}=0.3,0.46,0.6,0.1,0.15,0.23,0.3,0.5$, and $p=0.9,0.95$. For each set of parameter values 1000 sets of individual observations were generated and assessed for population bioequivalence in the same manner as for the evaluation of test size. The power of the is the number of replicated data sets out of 1000 that were judged to be population bioequivalent, based on evaluating the posterior predictive p-value at the $\alpha=0.05$ level. The power summary in Table 3 shows the minimum number of subjects per sequence needed to achieve 80% power, and the corresponding power.

If we compare these sample size requirements and power numbers to those in McNally et al, we find that the IC test produces similar power results to that of the generalized p-value test (GPV), and has much more power than the FDA test. The GPV test appears to be slightly more powerful than the IC test, but the IC algorithm appears to produce a much more acceptable test for population bioequivalence than the methodology currently recommended by the FDA.

### DISCUSSION

In the debate over the new bioequivalence standards now mandated by the FDA, individual bioequivalence has been given most of the attention; population bioequivalence has largely been ignored. The test for individual bioequivalence proposed by

---

**Table 1.** Population Bioequivalence of Methylphenidate

<table>
<thead>
<tr>
<th>95% Upper Bound</th>
<th>Area of Critical Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICCR</td>
<td>GCI</td>
</tr>
<tr>
<td>1.122</td>
<td>1.101</td>
</tr>
</tbody>
</table>

*Values in bold indicate the conclusion of population bioequivalence.*

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\sigma_{BR}=0.3$</th>
<th>$\sigma_{BR}=0.46$</th>
<th>$\sigma_{BR}=0.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>6</td>
<td>0.805</td>
<td>6</td>
</tr>
<tr>
<td>0.95</td>
<td>6</td>
<td>0.843</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\sigma_{BR}=0.6$</th>
<th>$\sigma_{BR}=1.0$</th>
<th>$\sigma_{BR}=0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>6</td>
<td>0.826</td>
<td>6</td>
</tr>
<tr>
<td>0.95</td>
<td>6</td>
<td>0.856</td>
<td>6</td>
</tr>
</tbody>
</table>

---

The posterior predictive p-value has an upper bound of $2\alpha$, where $\alpha$ is the nominal significance level of the test procedure.
Hyslop et al and recommended by the FDA is relatively straightforward and works very well for that parameter. It would be perfectly sensible to apply basically the same test to population bioequivalence, as the FDA has done, if not for the statistical dependence between the summary statistics. The new bioequivalence criteria are meant to address the needs of patients who could be exposed to a generic version of a drug product under the assumption that they will derive the same therapeutic effect as they would from the innovator product. The most critical scenario to be considered is whether a patient who is already receiving the innovator product will derive the same therapeutic benefit if they are switched to the generic product, which leads to individual bioequivalence. This 'switchability' is especially crucial when a patient has been titrated to a specific dose. Less critical but still important is whether a patient who is prescribed the generic product de novo will derive the same therapeutic benefit as he would had the innovator product been prescribed. This is the idea behind population bioequivalence. Although the bioequivalence criteria do not have any structural hierarchy, there is still a conceptual hierarchy that individual bioequivalence is a higher standard than population bioequivalence, which in turn is a higher standard than average bioequivalence. This is the seeming inconsistency in the methylphenidate example. Meyer et al were able to use the Hyslop method to conclude that the test and reference formulations of methylphenidate were individual bioequivalence with respect to AUC. McNally et al showed that the same conclusion could be reached using the GPV method and even the confidence interval methodology of Quiroz et al (2001). It is troubling that we cannot used the FDA test to conclude population bioequivalence for the methylphenidate, and it seems unlikely that this would be an isolated case, given what seems to be overwhelming evidence (i.e., very low p-values and upper bounds much lower than the upper limit of \( \Theta \)).

CONCLUSION

Our conclusions with respect to the power and size of the 2 tests parallel those of Chiang (2001), who compared the method of surrogate variables to the IC algorithm for variance components (the method of surrogate variables is essentially the same as the methodology for Weerahandi’s generalized confidence intervals). These conclusions may appear to be rash, especially as regards the size of the IC test, since they are based relatively small simulations for the IC test. Indeed, a more extensive investigation of the statistical properties of the IC test for population is probably warranted, similar to what McNally et al have done for the GPV test. An extensive examination of the use of informative priors in population and individual bioequivalence may also be useful, although informative priors may be regarded as problematic in a regulatory setting.

The main advantage of the IC test for population over the GPV test is that, thanks to the PRIOR statement in PROC MIXED, it is relatively straightforward to implement. There is a caveat, that the user needs to know what covariance parameters correspond to what treatments. This depends on how the treatments are ordered in SAS. On the other hand, the GPV test for population bioequivalence involves using a very complicated test function. This is in contrast to the GPV test for individual bioequivalence, which is also somewhat straightforward to program using SAS code for the GPV tests for individual and population bioequivalence is available from the author.

CONCLUSION

The independence chain (IC) algorithm provides a relatively straightforward methodology for assessing population bioequivalence. Although the IC test may not be the best method available, it can easily be implemented using the PRIOR statement in PROC MIXED, and it is superior to the test currently recommended by the FDA.

### Table 4. Power of GPV test for Population Bioequivalence (from McNally et al) \( \delta=0.05 \)

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \sigma_{BT}^{\text{GPV}}=0.3 )</th>
<th>( \sigma_{BT}^{\text{GPV}}=0.46 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_{WT}^{\text{GPV}}=0.15 )</td>
<td>( \sigma_{WT}^{\text{GPV}}=0.23 )</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Power</td>
<td>N</td>
</tr>
<tr>
<td>0.9</td>
<td>6</td>
<td>0.854</td>
</tr>
<tr>
<td>0.95</td>
<td>5</td>
<td>0.847</td>
</tr>
<tr>
<td>( \sigma_{BT}^{\text{GPV}}=0.6 )</td>
<td>( \sigma_{BT}^{\text{GPV}}=1.0 )</td>
<td></td>
</tr>
<tr>
<td>( \sigma_{WT}=0.3 )</td>
<td>( \sigma_{WT}=0.5 )</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Power</td>
<td>N</td>
</tr>
<tr>
<td>0.9</td>
<td>6</td>
<td>0.852</td>
</tr>
<tr>
<td>0.95</td>
<td>5</td>
<td>0.839</td>
</tr>
</tbody>
</table>

Naturally the question arises which is the better test for population bioequivalence, the GPV or the IC algorithm? Both Tsui and Weerahandi (1989) and Meng (1994) cite examples where the GPV gives the same answer as the Bayesian approach with a non-informative prior, but this does not seem to be the case here. Our tendency is to favor the GPV test, mainly for the following reasons.

1. The GPV test appears to have better power (see Table 4 for sample size requirements and power for the GPV test).
2. Despite seeming to be slightly less powerful, the IC algorithm does not appear uniformly to maintain the nominal level of significance.
REFERENCES
Quiroz J, Ting N, Wei GCG, Burdick RK. Alternative confidence intervals for the assessment of bioequivalence in four-period crossover designs. Statistics in Medicine, to appear.

ACKNOWLEDGMENTS
I would like to thank Marvin Meyer and Arthur Straughn of the University of Tennessee, Memphis for providing the methylphenidate data. I would also like to thank Hari Iyer of Colorado State University for his helpful comments.