PHARMACEUTICAL STATISTICS
Pharmaceut. Statist. 2007; **6**: 181–192
Published online 25 July 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/pst.295

WILEY InterScience® DISCOVER SOMETHING GREAT

Partition testing in dose–response studies with multiple endpoints

MAIN PAPER

Yi Liu[1,*,†], Jason Hsu[1] and Stephen Ruberg[2]
[1] *Statistics, The Ohio State University, Columbus, OH, USA*
[2] *Eli Lilly and Company, Indianapolis, IN, USA*

Dose–response studies with multiple endpoints can be formulated as closed testing or partition testing problems. When the endpoints are primary and secondary, whether the order in which the doses are to be tested is pre-determined or sample determined lead to different partitioning of the parameter space corresponding to the null hypotheses to be tested. We use the case of two doses and two endpoints to illustrate how to apply the partitioning principle to construct multiple tests that control the appropriate error rate. Graphical representation can be useful in visualizing the decision process. Copyright © 2007 John Wiley & Sons, Ltd.

Keywords: ***dose–response; multiple endpoints; partition testing; primary endpoint; secondary endpoint; error rate***

1. HISTORICAL ASPECT OF DOSE–RESPONSE TRIALS IN PHARMACEUTICS

Pharmaceutical drug development has long been divided into four phases with Phase II being the part of the development cycle where dose–response studies are conducted. The goal is usually to identify the best dose to use in confirmatory Phase III trials. While this has been the framework for several decades, there are still many drugs that fail to confirm efficacy and safety in Phase III (estimates are in the range 30–50%) with a substantial number of failures attributed to improper dose selection. Thus, the importance of the design and analysis of dose–response studies is as relevant today as it has ever been.

In the past, efficacy of a new drug was typically demonstrated by showing its superiority to a placebo (called a *negative* control). However, in recent years, when treatments known to be effective exist, and the disease does not cause mortality or irreversible morbidity, efficacy of a new drug might be defined as superiority or non-inferiority to a known effective treatment (called an *active* control). In the case of non-inferiority trials, the determination of non-inferiority might be based on what is a clinically meaningful difference, see [1]. Even in the case of superiority trials against a negative control, the definition of

*Correspondence to: Yi Liu, Statistics, The Ohio State University, 1958 Neil Avenue, Cockins Hall, Room 404, Columbus, OH 43210-1247, USA.
†E-mail: yliu@stat.osu.edu

Copyright © 2007 John Wiley & Sons, Ltd.

Pharmaceutical
STATISTICS

superiority might consider risk vs benefit, if there is toxicity concerns for the drug (see [2]). Some early discussion of these concepts can be seen in [3–6].

Dose–response studies may have multiple endpoints. A *primary endpoint* is one such that efficacy of a new drug relative to the control in this single endpoint constitutes evidence of efficacy. *Secondary endpoints* are ones where efficacy of a new treatment in any secondary endpoints supports evidence of efficacy, but by themselves (i.e. in the absence of efficacy in a primary endpoint) do not constitute evidence of efficacy.

When there are primary and secondary endpoints, inference on the secondary endpoint is given only if the compound is efficacious for the primary endpoint at that dose. This ordering guides the selection of test statistics for each intersection hypothesis in closed testing (as in [7]). We give a different perspective in this article, which is the ordering guides the partitioning of the parameter space in using the partitioning principle to construct multiple tests that control the appropriate familywise error rate (FWER).

In some (but not all) dose–response studies, it may be appropriate to pre-determine the order in which inferences on the doses are given. For example, one might start with the high dose and proceed to inference on the low dose only if the high dose shows efficacy. We show how this second 'ordering' further guides the partitioning of the parameter space. The resulting partitioning test is in the form of a decision tree which can be represented graphically.

Interestingly, the joint distribution of t-test statistics for multiple endpoints is *not* what is usually called the multivariate t-distribution. This article discusses the computation of this distribution in the bivariate case which, to avoid confusion, we call the *dual* t-distribution. We show, for example, that using algorithms for multivariate t-distributions results in slightly liberal critical values, while computing as if the t-statistics were independent results in somewhat conservative critical values.

Section 2 gives a motivating example of a dose–response study with multiple endpoints. Section 3 shows how the partitioning principle forms null hypotheses when inferences are ordered by dose, and the corresponding multiple test is a step-down test. Section 4 extends the partitioning principle of forming null hypotheses to when inferences are ordered by dose and by endpoint. The corresponding multiple test has a graphical representation. That section also contains a study of issues in the computation of critical values. Section 5 provides a numerical illustration of methods developed in this article, using the real data example in Section 2.

2. A MOTIVATING EXAMPLE FOR DOSE–RESPONSE STUDIES

Consider, for example, [8], a 26-center double-blind trial comparing the effect of five doses of the anti-psychotic drug 'Seroquel' (Quetiapine) and the placebo with parallel design on a total of 361 patients. For illustration purpose, we will focus on two of the doses (75 and 600 mg/day) with the primary endpoint being Clinical Global Impression (CGI) Global Improvement score and the secondary endpoint being CGI Severity of Illness score. Summary statistics for these two doses and endpoints are presented in Table I.

Throughout the paper, the placebo (0 mg/day), low dose (75 mg/day) and high dose (600 mg/day) groups will be indexed as $i = 0, 1, 2$. Primary and secondary endpoints will be indexed by superscripts $L = P, S$. For discussion involving only the primary endpoint, the superscript P will be dropped for convenience.

Let μ_i^L denote the mean response of dose group i for endpoint L, $i = 0, 1, 2$, $L = P, S$. Define $\theta_i^L = \mu_i^L - \mu_0^L$ as the true mean difference between dose group i and the placebo for endpoint L, $i = 0, 1, 2$, $L = P, S$. Let δ^L denote the clinically meaningful difference for endpoint L, $L = P, S$. The family of null hypotheses of interest consists of four null hypotheses with the first two in (1) concerning primary endpoint and the second two in (2) concerning secondary endpoint:

$$\begin{aligned} H_{01}^P : \theta_1^P \leqslant \delta^P \quad \text{vs} \quad H_{a1}^P : \theta_1^P > \delta^P \\ H_{02}^P : \theta_2^P \leqslant \delta^P \quad \text{vs} \quad H_{a2}^P : \theta_2^P > \delta^P \end{aligned} \tag{1}$$

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/pst

Table I. Summary statistics of two doses and two endpoints in [8].

Endpoint		Dose		
		0 mg/day (Placebo)	75 mg/day	600 mg/day
	Sample size	51	52	51
CGI Global Improvement score (Primary)	Mean	4.78	4.22	3.58
	SE*	0.23	0.22	0.23
CGI Severity of Illness score (Secondary)	Mean	5.2	4.8	4.4
	SE	1.2	1.3	1.5

*Standard error.

$$H_{01}^S : \theta_1^S \leqslant \delta^S \quad \text{vs} \quad H_{a1}^S : \theta_1^S > \delta^S$$
$$H_{02}^S : \theta_2^S \leqslant \delta^S \quad \text{vs} \quad H_{a2}^S : \theta_2^S > \delta^S \qquad (2)$$

Assume that the samples from dose group i for endpoint L, $Y_{i1}^L, \ldots, Y_{in_i}^L$, $i = 0, 1, 2$, $L = P, S$, come from a model:

$$Y_{ir}^L = \mu_i^L + \varepsilon_{ir}^L, \quad i = 0, 1, 2, \ r = 1, \ldots, n_i, \ L = P, S \qquad (3)$$

where

$$\begin{pmatrix} \varepsilon_{ir}^P \\ \varepsilon_{ir}^S \end{pmatrix} \text{ i.i.d. } \sim \text{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} (\sigma^P)^2 & \rho\sigma^P\sigma^S \\ \rho\sigma^P\sigma^S & (\sigma^S)^2 \end{pmatrix} \right),$$
$$i = 0, 1, 2, \ r = 1, \ldots, n_i$$

Define T_i^L, $i = 1, 2$, $L = P, S$, in (4) to be the t-statistic for testing the null hypothesis H_{0i}^L, $i = 1, 2, L = P, S$

$$T_i^L = \frac{\bar{Y}_{i\cdot}^L - \bar{Y}_{0\cdot}^L - \delta^L}{\hat{\sigma}^L \sqrt{1/n_i + 1/n_0}}, \quad i = 1, 2, \ L = P, S \qquad (4)$$

where

$$\bar{Y}_{i\cdot}^L = (1/n_i) \sum_{r=1}^{n_i} Y_{ir}^L,$$
$$\hat{\sigma}^L = \sqrt{\textstyle\sum_{i=0}^{2} \sum_{r=1}^{n_i} (Y_{ir}^L - \bar{Y}_{i\cdot}^L)^2 / \sum_{i=0}^{2} (n_i - 1)},$$
$$i = 0, 1, 2, L = P, S.$$

The purpose of dose–response studies is to find which doses are effective. Control of multiple testing error rate should control the probability of incorrectly inferring a dose is efficacious for some endpoint when in fact it is not. In this situation, controlling the false discovery rate (FDR) will not control this probability (see [9]), while controlling the familywise error rate (FWER) strongly for appropriately formulated null hypotheses will. We thus consider methods that strongly control FWER in this article.

With four null hypotheses in (1) and (2), a straightforward application of the closed testing principle would test $2^4 - 1 = 15$ intersection null hypotheses to control FWER. However, we will show that if inferences are ordered by both endpoints and doses, then testing only four (disjoint) null hypotheses according to the partitioning principle controls FWER. These hypotheses are tested in three steps, with multiplicity adjustment involved in only one of the steps. When there are $k > 2$ doses and $m > 2$ endpoints, a similar step-down procedure could in theory be developed following the same principle.

3. MULTIPLE TESTS CONSTRUCTION USING THE PARTITIONING PRINCIPLE

The purpose of this article is to demonstrate how to use the partitioning principle, a fundamental multiple tests constructing technique, to construct multiple tests when there are multiple doses and multiple endpoints. We illustrate the idea with the two hypotheses in (1), concerning the primary endpoint first.

3.1. Closed testing to step-down doses according to sample responses

The closed testing technique of [10] tests all possible non-empty intersections of the hypotheses

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/pst

Pharmaceutical STATISTICS

in (1), leading to the three hypotheses in (5), each at level-α:

$$\begin{aligned} H_0^{\cap}&: \ \theta_1 \leqslant \delta \text{ and } \theta_2 \leqslant \delta \quad \text{(neither dose is efficacious)} \\ H_{01}&: \ \theta_1 \leqslant \delta \quad \text{(low dose is not efficacious)} \\ H_{02}&: \ \theta_2 \leqslant \delta \quad \text{(high dose is not efficacious)} \end{aligned} \tag{5}$$

The logical implications of testing are:

- If $H_0^{\cap}$ is not rejected, no inference is given even if H_{01} or H_{02} or both are rejected, because $H_0^{\cap}$ implies H_{01} and H_{02}.
- If only $H_0^{\cap}$ and H_{01} (H_{02}) are rejected but not H_{02} (H_{01}), then infer low (high) dose is efficacious.
- If all three hypotheses are rejected, then infer both doses are efficacious.

We use t-statistics T_1 and T_2 defined in (4) (with superscript P dropped) to test the three intersection hypotheses in (5) with critical values $d_{\alpha,2,\nu}$ (the upper α quantile of Dunnett distribution with 2 and $\nu = \sum_{i=0}^{2}(n_i - 1)$ degrees of freedom) and $t_{\alpha,\nu}$ (the upper α quantile of t-distribution with ν degrees of freedom). Let (1) and (2) denote the random indices such that $T_{(1)} < T_{(2)}$, then the rejection rules can be given in Table II.

Since the critical values satisfy $d_{\alpha,2,\nu} > t_{\alpha,\nu}$, a step-down procedure with sample-determined steps exists, as follows:

- *Step 1:* If $T_{(2)} > d_{\alpha,2,\nu}$, infer dose (2) is efficacious and go to step 2; else stop.
- *Step 2:* If $T_{(1)} > t_{\alpha,\nu}$, infer dose (1) is efficacious and stop; else stop.

3.2. Partition to step-down doses

If higher dosage is expected to be more efficacious, then one may choose to always test the high dose first. If high dose is effective, then proceed to test low dose; otherwise, stop. We can formally state this pre-determined sequence of testing as follows.

Condition A (*order in doses*): The low dose cannot be claimed efficacious unless the high dose has shown evidence of efficacy.

The partitioning principle of [11, 12] is a general principle for constructing multiple tests. Under *condition A*, [13] partition the null space, $\{\boldsymbol{\theta} \in \mathbb{R}^2 | \theta_1 \leqslant \delta \text{ or } \theta_2 \leqslant \delta\}$ into two disjoint subspaces $\{\boldsymbol{\theta} \in \mathbb{R}^2 | \theta_2 \leqslant \delta\}$ and $\{\boldsymbol{\theta} \in \mathbb{R}^2 | \theta_1 \leqslant \delta \text{ and } \theta_2 > \delta\}$ corresponding to the two hypotheses:

$$\begin{aligned} H_{02}^{\downarrow}&: \ \theta_2 \leqslant \delta \quad \text{(high dose is not efficacious)} \\ H_{01}^{\downarrow}&: \ \theta_1 \leqslant \delta \text{ and } \theta_2 > \delta \\ &\text{(low dose is not efficacious but high dose is)} \end{aligned} \tag{6}$$

The logical implications of testing are:

- If $H_{02}^{\downarrow}$ is not rejected (regardless of $H_{01}^{\downarrow}$), then no inference is given, since 'neither dose is efficacious' (which is contained in $H_{02}^{\downarrow}$) is not rejected.
- If $H_{02}^{\downarrow}$ and $H_{01}^{\downarrow}$ are rejected, then since the union of $H_{02}^{\downarrow}$ and $H_{01}^{\downarrow}$ is 'either low dose or high dose is not efficacious', the implication is 'both high dose and low dose are efficacious'.
- If $H_{02}^{\downarrow}$ is rejected but $H_{01}^{\downarrow}$ is not rejected, then one infers high dose is efficacious.

The interesting fact is, in testing $H_{0i}^{\downarrow}$, $i = 1, 2$ simultaneously, no multiplicity adjustment is needed to control FWER, the probability of rejecting any true null hypothesis. This is because the null spaces of $H_{0i}^{\downarrow}$, $i = 1, 2$, are disjoint. In other words, at most one null hypothesis can be true: it cannot be the case that high dose is ineffective (H_{02}) and that high dose is effective but low dose is ineffective (H_{01}), for example. We thus test each $H_{0i}^{\downarrow}$, $i = 1, 2$, at level-α.

Level-α tests for each $H_{0i}^{\downarrow}$, $i = 1, 2$, are of course not unique. Note, however, a level-α test for $H_{01}: \theta_1 \leqslant \delta$ is also a level-α test for $H_{01}^{\downarrow}: \theta_1 \leqslant \delta$ and $\theta_2 > \delta$. For example, a test that rejects no more than 5% of the time when low dose is ineffective, regardless of whether high dose is effective, will reject no more than 5% of the time in particular when low dose is ineffective and high dose is effective.

Table II. Decision rules for closed testing with two doses single primary endpoint.

Hypothesis	Rejection rule	Test level
$H_0^{\cap}: \theta_1 \leqslant \delta$ and $\theta_2 \leqslant \delta$	$T_{(2)} > d_{\alpha,2,\nu}$	α
$H_{01}: \theta_1 \leqslant \delta$	$T_1 > t_{\alpha,\nu}$	α
$H_{02}: \theta_2 \leqslant \delta$	$T_2 > t_{\alpha,\nu}$	α

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/pst

So the simplest level-α test for $H_{0i}^{\downarrow}$ is to use a one-sided two-sample size α t-test based on T_i, $i = 1, 2$, defined in (4) for each $H_{0i}^{\downarrow}$ as shown in Table III.

In terms of the rejection rules in Table III, the pre-determined D-steps (D stands for doses) proceed as follows.

- *Step 1:* If $T_2 > t_{\alpha,\nu}$, infer high dose is efficacious and go to step 2; else stop.
- *Step 2:* If $T_1 > t_{\alpha,\nu}$, infer low dose is efficacious and stop; else stop.

Note that, even though the method above controls FWER regardless of whether the shape of the true response function, it is recommended only when the response is expected to be monotonically increasing (for otherwise, it might stop too soon and miss an efficacious dose).

4. PARTITION TESTING WITH TWO DOSES AND TWO ENDPOINTS

Suppose the dose–response study has both a primary endpoint and a secondary endpoint, so that the family of null hypotheses of interest includes both those in (1) and (2). We will show that, in addition to stepping through doses, the partitioning principle can also be used to derive multiple tests that step through endpoints in a pre-determined sequence.

4.1. Partition to step-down endpoints

Multiple testing procedures for dose–response studies with a primary endpoint and hierarchically ordered secondary endpoints were developed in [7]. In the case of two doses and two endpoints (one secondary endpoint), the procedure has to satisfy the condition that for the same dose, the secondary endpoint is tested only if primary endpoint is claimed to be efficacious, which is formally stated as follows.

Table III. Decision rules for partition testing to step-down doses.

Hypothesis	Rejection rule	Test level
$H_0^{\cap} : \theta_2 \leqslant \delta$	$T_2 > t_{\alpha,\nu}$	α
$H_{01} : \theta_1 \leqslant \delta$ and $\theta_2 > \delta$	$T_1 > t_{\alpha,\nu}$	α

Condition B (*order in endpoints*): For the same dose, a secondary endpoint cannot be claimed efficacious unless its primary endpoint has been shown to be efficacious.

Within each endpoint, Dunnett's method was used in [7] to adjust multiplicity for multiple doses. The FWER of the whole procedure was controlled at α using the principle of closed testing. Figure 1 is a graphical representation of their procedure with two doses and two endpoints.

The procedure in [7] can be reproduced by partitioning the parameter space with constraints on ordered endpoints for each dose, similar to partitioning to step down through the doses technique as illustrated in Section 4. For example, for high dose, we would test the following two hypotheses in a step-down fashion (pre-determined E-steps):

$H_{02}^{P\downarrow}$: $\theta_2^P \leqslant \delta^P$ (primary endpoint is not effective)

$H_{02}^{S\downarrow}$: $\theta_2^S \leqslant \delta^S$ and $\theta_2^P > \delta^P$

(Secondary endpoint is not efficacious, but primary endpoint is)

4.2. Partitioning to step-down both doses and endpoints

The procedure developed in [7] steps through the endpoints in a pre-determined sequence, satisfying *condition B*, but not necessarily steps through the doses in a pre-determined sequence.

To step through both doses and endpoints, we test the following four partition hypotheses in a step-down fashion (pre-determined DE-steps):

- *Step 1*: Test $H_{02}^{P\downarrow} : \theta_2^P \leqslant \delta^P$ (high dose is not effective for primary endpoint).
  - If rejected, infer high dose for the primary endpoint is effective and proceed to step 2; else stop.
- *Step 2*: Test two hypotheses: $H_{01}^{P\downarrow} : \theta_1^P \leqslant \delta^P$ and $\theta_2^P > \delta^P$ (low dose is not effective, but high dose is effective for primary endpoint).

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/pst

Pharmaceutical
STATISTICS

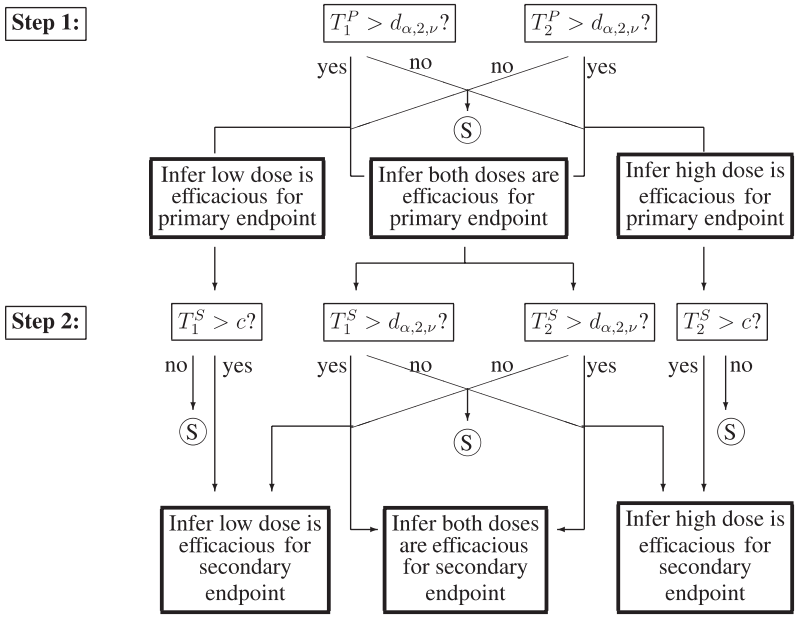

Figure 1. Decision process in [7]. S means 'Stop', c is calculated based on sample correlation between endpoints.

$H_{02}^{S\downarrow}: \theta_2^S \leqslant \delta^S$ and $\theta_2^P > \delta^P$ (high dose is not effective for secondary endpoint, but is effective for primary endpoint).

- If both hypotheses are rejected, infer both doses are effective for primary endpoint, high dose is also effective for secondary endpoint and proceed to the next step.
- Otherwise, if only $H_{01}^{P\downarrow}$ is rejected, infer both doses are effective for primary endpoint and stop.
- If only $H_{02}^{S\downarrow}$ is rejected, infer high dose is effective for both endpoints and stop.

- *Step 3*: Test $H_{01}^{S\downarrow}: \theta_1^S \leqslant \delta^S$ and $\theta_2^S > \delta^S$ and $\theta_1^P > \delta^P$ and $\theta_2^P > \delta^P$. (Low dose is not effective for secondary endpoint, but is effective for primary endpoint and high dose is effective for both endpoints.)
  - If rejected, infer both doses are effective for both endpoints and stop. Otherwise stop.

The direction of these steps is presented in Table IV.

Table IV. Direction of the pre-determined DE-steps.

Endpoint/dose	Low		High
Primary	Step 2*	←	Step 1
	↓		↓
Secondary	Step 3	←	Step 2*

*Proceed only if both hypotheses in step 2 are rejected.

4.2.1. Decision tree.

The rejection rules for these hypotheses in each step are presented in Table V in the form of usual t-statistics T_i^L, $i = 1, 2$, $L = P, S$, as defined in (4).

For steps 1 and 3, the simplest choice of critical value c_1 is $t_{\alpha,\nu}$. For step 2, since the null spaces of $H_{01}^{P\downarrow}$ and $H_{02}^{S\downarrow}$ are not disjoint, we need to adjust for multiplicity to make the FWER for the whole procedure controlled at level α.

The decision tree of this step-down procedure with pre-determined DE-steps is presented in Figure 2.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/pst

4.2.2. Computing the critical value c_2.

According to Table V, we need to find the critical value c_2 accounted for multiplicity for the two hypotheses in step 2. That is, one needs to focus on the joint distribution of (T_1^P, T_2^S). Notice that between doses, the observations are independent, but within a dose, observations between endpoints are dependent, the numerators as well as the denominators of T_1^P and T_2^S are correlated through endpoints for the same dose. For one-sided tests, with the rejection rule of the form in (7), the supremum of the probability is obtained at the boundary values (involving θ_1^P and θ_2^S only, since neither T_1^P nor T_2^S involves θ_2^P). The critical value c_2 needs to satisfy:

$$
\begin{aligned}
1-\alpha &\leqslant 1 - \sup_{\theta_1^P \leqslant \delta^P \text{ and } \theta_2^S \leqslant \delta^S} \Pr_\rho\{T_1^P > c_2 \text{ or } T_2^S > c_2\} \\
&= \Pr_\rho\left\{ \frac{\bar{Y}_{1\cdot}^P - \bar{Y}_{0\cdot}^P - \theta_1^P}{\hat{\sigma}^P\sqrt{1/n_1 + 1/n_0}} \leqslant c_2 \quad \text{and} \right. \\
&\qquad \left. \frac{\bar{Y}_{2\cdot}^S - \bar{Y}_{0\cdot}^S - \theta_2^S}{\hat{\sigma}^S\sqrt{1/n_2 + 1/n_0}} \leqslant c_2 \right\} \\
&= \Pr_\rho\left\{ \frac{Z_1}{\hat{\sigma}^P/\sigma^P} \leqslant c_2 \text{ and } \frac{Z_2}{\hat{\sigma}^S/\sigma^S} \leqslant c_2 \right\} \qquad (7) \\
&= \int_0^\infty \int_0^\infty \Pr_\rho\{Z_1 \leqslant c_2 s_1 \text{ and } Z_2 \leqslant c_2 s_2\} \\
&\quad \times \gamma_{\rho,\nu}(s_1, s_2)\, \mathrm{d}s_1\, \mathrm{d}s_2 \qquad (8)
\end{aligned}
$$

Table V. Critical values for the pre-determined DE-steps.

Step	Hypothesis	Rejection rule	Test level
1	$H_{02}^{P\downarrow}: \theta_2^P \leqslant \delta^P$	$T_2^P > c_1$	α
2	$H_{01}^{P\downarrow}: \theta_1^P \leqslant \delta^P$ and $\theta_2^P > \delta^P$	$T_1^P > c_2$	MA*
	$H_{02}^{S\downarrow}: \theta_2^S \leqslant \delta^S$ and $\theta_2^P > \delta^P$	$T_2^S > c_2$	MA
3	$H_{01}^{S\downarrow}: \theta_1^S \leqslant \delta^S$ and $\theta_2^S > \delta^S$ and $\theta_1^P > \delta^P$ and $\theta_2^P > \delta^P$	$T_1^S > c_1$	α

*Multiplicity adjustment needed.

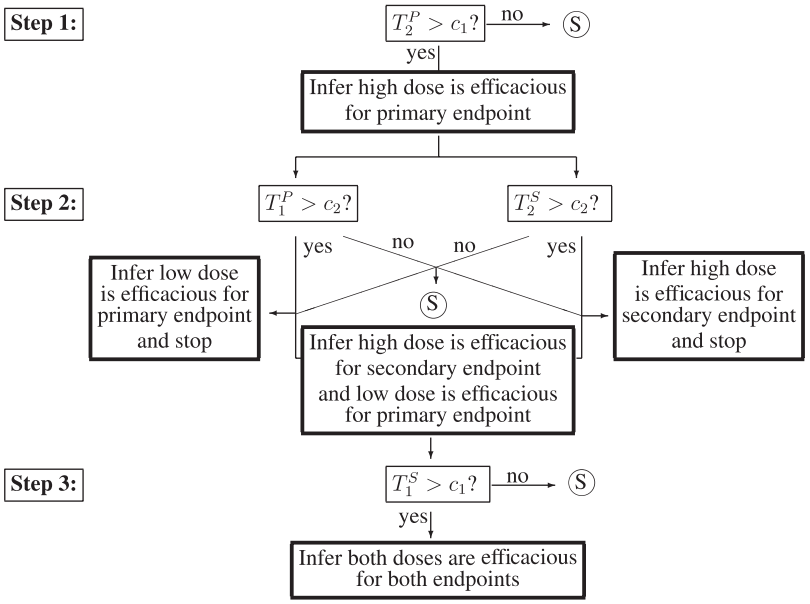

Figure 2. Decision tree for pre-determined DE-steps.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/pst

Pharmaceutical STATISTICS

where

$$(Z_1, Z_2) = \left(\frac{\bar{Y}^P_{1\cdot} - \bar{Y}^P_{0\cdot} - \theta^P_1}{\sigma^P \sqrt{1/n_1 + 1/n_0}}, \frac{\bar{Y}^S_{2\cdot} - \bar{Y}^S_{0\cdot} - \theta^S_2}{\sigma^S \sqrt{1/n_2 + 1/n_0}} \right)$$

with mean $\mathbf{0} = (0, 0)$, variance 1 and covariance $\rho/(\sqrt{(n_0/n_1 + 1)(n_0/n_2 + 1)})$ $(= \rho/2$ in the balanced case), $\gamma_{\rho,v}(s_1, s_2)$ is the joint density of $(\hat{\sigma}^P/\sigma^P, \hat{\sigma}^S/\sigma^S)$.

Note that the definition of a bivariate *t*-distribution, as originally introduced by [14] and computed by the ProbMC function in SAS and the qmvnorm function in the R package mvtnorm, requires the denominators in (7) to be the same random variable $\hat{\sigma}^P \equiv \hat{\sigma}^S$, which is not the case here. In fact, computing (7) as if the distribution were bivariate *t* overestimates the probability, resulting in somewhat liberal critical values, as we will demonstrate. To avoid confusion, we call the distribution of (T^P_1, T^S_2) the *dual t*-distribution. (This distribution was studied by Siddiqui [15].)

Probability (8) involves the unknown parameter ρ. In the following, we describe five ways of approximating (8). The first four methods (normal, bivariate *t*, dual *t* and independent standard errors) require ρ be known or approximated. The last method computes by assuming that T^P_1 and T^S_2 are independent, which we prove results in a conservative critical value.

Normal approximation: If the sample sizes n_1, n_2 are large, by Slutsky's theorem, $\hat{\sigma}^P/\sigma^P$ and $\hat{\sigma}^S/\sigma^S$ converge in probability to 1, then the joint distribution of (T^P_1, T^S_2) (under the true mean difference) is asymptotically bivariate normal as (Z_1, Z_2).

Bivariate t distribution: Assume $\hat{\sigma}^P/\sigma^P$ and $\hat{\sigma}^S/\sigma^S$ are perfectly correlated, or they are the same, the situation coincides with the bivariate *t* case. Specifically, the probability in (7) reduces to

$$\int_0^\infty \Pr_\rho\{Z_1 \leqslant c_2 s \text{ and } Z_2 \leqslant c_2 s\} \gamma_v(s)\, \mathrm{d}s$$

where $\gamma_v(s)$ is the density of $\sqrt{\chi^2_v/v}$, and $v = \sum_{i=0}^{2}(n_i - 1)$ is the degrees of freedom.

Dual t distribution: It can be shown that $\hat{\sigma}^P/\sigma^P$ and $\hat{\sigma}^S/\sigma^S$ are diagonal elements of a Wishart random matrix (see Appendix). We use simulations from a Wishart distribution (which depends on ρ, the correlation between endpoints from a single individual) and Monte Carlo methods to approximate the probability based on exact dual *t*-distribution of the test statistics (T^P_1, T^S_2).

Independent standard errors $\hat{\sigma}^P, \hat{\sigma}^S$: The difficulty in computing (7) is caused, in part, by pooling data across doses in estimating the standard errors σ^P, σ^S, resulting in the denominators of test statistics (T^P_1, T^S_2), $(\hat{\sigma}^P/\sigma^P, \hat{\sigma}^S/\sigma^S)$ being correlated. One can obtain a conservative upper bound of c_2 (see Theorem 1) in solving $(9) = 1 - \alpha$ for c_2 by assuming $\hat{\sigma}^P/\sigma^P$ and $\hat{\sigma}^S/\sigma^S$ are independent, where (9) can be viewed as a simplified version of (8):

$$\int_0^\infty \int_0^\infty \Pr_\rho\{Z_1 \leqslant c_2 s_1 \text{ and } Z_2 \leqslant c_2 s_2\} \times \gamma_v(s_1)\gamma_v(s_2)\, \mathrm{d}s_1\, \mathrm{d}s_2 \tag{9}$$

Theorem 1. *Probability (7) is greater when* $\hat{\sigma}^P/\sigma^P$ *and* $\hat{\sigma}^S/\sigma^S$ *are correlated than when they are independent.*

Proof. *We first condition on* Z_1 *and* Z_2*, and note* $c_2 > 0$

$$\begin{aligned}
&\Pr_\rho\left\{\frac{Z_1}{\hat{\sigma}^P/\sigma^P} \leqslant c_2 \text{ and } \frac{Z_2}{\hat{\sigma}^S/\sigma^S} \leqslant c_2\right\} \\
&= E_{(Z_1,Z_2)}\left[\Pr_\rho\left\{\frac{z_1}{\hat{\sigma}^P/\sigma^P} \leqslant c_2 \text{ and } \frac{z_2}{\hat{\sigma}^S/\sigma^S} \leqslant c_2\right| \right. \\
&\quad \left.(Z_1, Z_2) = (z_1, z_2)\}\right] \\
&= E_{(Z_1,Z_2)}\left[\Pr_\rho\left\{\frac{z_1}{\hat{\sigma}^P/\sigma^P} \leqslant c_2 \text{ and } \frac{z_2}{\hat{\sigma}^S/\sigma^S} \leqslant c_2\right\}\right] \\
&= E_{(Z_1,Z_2)}\left[\Pr_\rho\left\{\hat{\sigma}^P/\sigma^P \geqslant \frac{z_1}{c_2} \text{ and } \hat{\sigma}^S/\sigma^S \geqslant \frac{z_2}{c_2}\right\}\right]
\end{aligned} \tag{10}$$

Since $\hat{\sigma}^P/\sigma^P$ *and* $\hat{\sigma}^S/\sigma^S$ *are associated (see [16, Theorem 6.1]), for any given* Z_1 *and* Z_2

$$\begin{aligned}
&\Pr_\rho\left\{\hat{\sigma}^P/\sigma^P \geqslant \frac{z_1}{c_2} \text{ and } \hat{\sigma}^S/\sigma^S \geqslant \frac{z_2}{c_2}\right\} \\
&\quad \geqslant \Pr\left\{\hat{\sigma}^P/\sigma^P \geqslant \frac{z_1}{c_2}\right\} \Pr\left\{\hat{\sigma}^S/\sigma^S \geqslant \frac{z_2}{c_2}\right\}
\end{aligned} \tag{11}$$

The result follows. □

Independent t statistics T^P_1, T^S_2: To further break down the four-dimensional integral in (9), we add the

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/pst

assumption of independent numerators (Z_1, Z_2) of (T_1^P, T_2^S) in addition to independent denominators, which is reasonable since the correlation between Z_1 and Z_2 is only $\rho/2$ in the balanced case. This method is more conservative than the previous method shown in Theorem 2.

Theorem 2. *Probability (9) assuming $\hat{\sigma}^P/\sigma^P$ and $\hat{\sigma}^S/\sigma^S$ are independent is greater than the probability assuming test statistics (T_1^P, T_2^S) are independent. (This is equivalent to further assuming the numerators of (T_1^P, T_2^S), (Z_1, Z_2) are independent.)*

Proof. *Assuming independence of $\hat{\sigma}^P/\sigma^P$ and $\hat{\sigma}^S/\sigma^S$, we have*

$$\begin{aligned}
&\Pr_\rho\left\{\frac{Z_1}{\hat{\sigma}^P/\sigma^P}\leqslant c_2 \text{ and } \frac{Z_2}{\hat{\sigma}^S/\sigma^S}\leqslant c_2\right\}\\
&=\int_0^\infty\int_0^\infty \Pr_\rho\{Z_1\leqslant c_2 s_1 \text{ and } Z_2\leqslant c_2 s_2\}\\
&\quad\times\gamma_v(s_1)\gamma_v(s_2)\,\mathrm{d}s_1\,\mathrm{d}s_2
\end{aligned}$$

since $Cov(Z_1, Z_2)=\rho/\sqrt{(n_0/n_1+1)(n_0/n_2+1)}>0$ *By Slepian's inequality* (*Corollary A.3.1on p. 229 of* [17])

$$\begin{aligned}
&\geqslant\int_0^\infty\int_0^\infty \Pr\{Z_1\leqslant c_2 s_1\}\Pr\{Z_2\leqslant c_2 s_2\}\\
&\quad\times\gamma_v(s_1)\gamma_v(s_2)\,\mathrm{d}s_1\,\mathrm{d}s_2\\
&=\int_0^\infty \Pr\{Z_1\leqslant c_2 s_1\}\gamma_v(s_1)\,\mathrm{d}s_1\\
&\quad\times\int_0^\infty \Pr\{Z_2\leqslant c_2 s_2\}\gamma_v(s_2)\,\mathrm{d}s_2\\
&=\Pr\left\{\frac{Z_1}{\hat{\sigma}^P/\sigma^P}\leqslant c_2\right\}\Pr\left\{\frac{Z_2}{\hat{\sigma}^S/\sigma^S}\leqslant c_2\right\}\\
&=\Pr\{T_1^P\leqslant c_2\}\Pr\{T_2^S\leqslant c_2\}
\end{aligned}$$

□

The degrees of conservatism for the methods introduced above are compared in terms of critical value c_2 for different combinations of correlation between endpoints ρ and error degrees of freedom v in Table VI. Critical values for the dual t and independent standard error methods are based on 100 000 simulations.

As shown in Table VI, for each combination of ρ and v, the normal approximation is somewhat liberal and the bivariate t-approximation is slightly liberal. Assuming independence of $\hat{\sigma}^P/\sigma^P$ and $\hat{\sigma}^S/\sigma^S$ is slightly conservative, while assuming independence of t-statistics is somewhat conservative. It should be noted that, in the setting of simultaneous efficacy and safety studies, Tamhane and Logan [19] considered using the Bonferroni inequality and the bootstrap technique to compute such probabilities.

5. ANALYSIS OF THE ANTI-PSYCHOTIC DRUG DATA

We illustrate the step-down procedure with predetermined DE-steps using the anti-psychotic drug example introduced at the beginning of the paper with $\delta^P=\delta^S=0$. Using pooled standard error from these three groups, the t-statistics corresponding to each dose and endpoint and the error degrees of freedom are:

$$T_1^P=-1.7499,\quad T_2^P=-3.7318,$$
$$T_1^S=-0.2116,\quad T_2^S=-0.4212\quad\text{and}\quad v=151$$

If we want to control the FWER at 5% level, then $c_1=t_{0.05,151}=1.6550$, while $c_2=1.9702$, based on the independent t method (which is appropriate because the correlation is unknown). Note in this particular example, a lower score indicates better drug effect. To make it consistent with our hypotheses setup, we take the negatives of the t-statistics and then apply our procedure.

- *Step 1*: Is $3.7318>1.655$? Yes, go to step 2.
- *Step 2*: Is $1.7499>1.9702$ or $0.4212>1.9702$? No, stop.

High dose (600 mg/day) is inferred to be efficacious for primary endpoint only, and no evidence of efficacy can be made for low dose (75 mg/day).

6. CONCLUSION

In this paper, we presented a systematic way of constructing null hypotheses by partitioning the

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/pst

Table VI. Comparing conservatism in terms of critical value c_2.

		ρ[a]				
v	Method	0	0.2	0.4	0.6	0.8
50	Bivariate t[b]	2.0015	1.9972	1.9913	1.9833	1.9730
	Dual t[c]	2.0026	1.9988	1.9934	1.9851	1.9741
	Ind. SE	2.0028	1.9988	1.9940	1.9866	1.9771
	Ind. T	2.0028	2.0028	2.0028	2.0028	2.0028
100	Bivariate t	1.9777	1.9737	1.9682	1.9607	1.9507
	Dual t	1.9783	1.9744	1.9692	1.9617	1.9513
	Ind. SE	1.9783	1.9745	1.9694	1.9623	1.9527
	Ind. T	1.9783	1.9783	1.9783	1.9783	1.9783
200	Bivariate t	1.9660	1.9622	1.9569	1.9495	1.9398
	Dual t	1.9659	1.9627	1.9572	1.9498	1.9401
	Ind. SE	1.9664	1.9628	1.9573	1.9503	1.9408
	Ind. T	1.9664	1.9664	1.9664	1.9664	1.9664
∞	Normal[d]	1.9545	1.9508	1.9456	1.9385	1.9289

[a] Note ρ refers to the correlation between endpoints for a single individual. The correlation between numerators (Z_1, Z_2) of the test statistics is assumed to be $\rho/2$ in the balanced case.
[b] The critical values of bivariate t are obtained by *qmvt* function in R package *mvtnorm*.
[c] Wishart is generated first using Bartlett's decomposition and then appropriately transformed (see [18]).
[d] The bivariate normal quantiles are obtained by *qmvnorm* of R package *mvtnorm*.

parameter space according to conditions that order inferences by dose (high to low), or endpoint (primary to secondary), or both. The hypotheses so constructed automatically form step-down procedures (pre-determined D-steps, E-steps, or DE-steps) and control the proper error rate at a pre-specified level. This way of test construction can be generalized to situations with $k > 2$ doses and $m > 2$ endpoints. The technique is as follows. Start by writing down the null hypotheses whose rejections correspond to desired inferences on primary and secondary endpoints. Then, following stated conditions for the decision process, make each null hypothesis in a subsequent step disjoint from null hypotheses in previous steps (by removing from it the union of the null hypotheses whose rejections lead up to it). Adjust for multiplicity within a step (only) if the null hypotheses are not disjoint, but do not adjust for multiplicity between steps.

An important issue in multiple endpoint problems is how to deal with correlations among the test statistics induced by correlations among measurements on the multiple endpoints. We made a systematic study of liberalism and conservatism of five approximation methods.

Graphical representation makes the decision process clear and easy to understand. Multiplicity adjustment may be needed for some steps but not all. There are two general rules. First, if there are two or more statistics involved in one box, then multiplicity adjustment may be needed. Second, if there are two or more branches from a box, and the hypotheses corresponding to these boxes are not disjoint, then further multiplicity adjustment may be needed among those boxes.

ACKNOWLEDGEMENTS

Jason Hsu's research is supported by Grant No. DMS-0505519 from the US National Science Foundation. Haiyan Xu motivated us to work on this problem, and we have had rather useful discussions with her, Frank Bretz, and John Lawrence. We also thank the reviewers for improving the paper with very helpful comments.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/pst

REFERENCES

1. ICH E10. *Choice of Control Groups in Clinical Trials*. CPMP (Committee for Proprietary Medical Products), EMEA (The European Agency for the Evaluation of Medical Products), London, Draft ICH (International Conference on Harmonisation) Guideline ed., 1999. http://www.ifpma.org/pdfifpma/e10.pdf
2. Temple RJ. Effect size – can the effect be too small. Presentation at FDA Center for Drug Evaluation and Research, Cardiovascular and Renal Advisory Committee meeting, 2006. http://www.fda.gov/ohrms/dockets/ac/06/slides/2006-4215S1-02-FDA.ppt
3. Ruberg SJ. Contrasts for identifying the minimum effective dose. *Journal of the American Statistical Association* 1989; **84**(407):816–822.
4. Ruberg SJ. Dose response studies. I. Some design considerations. *Journal of Biopharmaceutical Statistics* 1995; **5**(1):1–14.
5. Ruberg SJ. Dose response studies. II. Analysis and interpretation. *Journal of Biopharmaceutical Statistics* 1995; **5**(1):15–42.
6. Ruberg S, Cairns V. Providing evidence of efficacy for a new drug. *Statistics in Medicine* 1998; **17**(15–16):1813–1823.
7. Dmitrienko A, Offen W, Wang O, Xiao D. Gatekeeping procedures in dose–response clinical trials based on the Dunnett test. *Pharmaceutical Statistics* 2006; **5**:19–28.
8. Arvanitis L, Miller B. The Seroquel Trial 13 Study Group. Multiple fixed doses of 'Seroquel' (Quetiapine) in patients with acute exacerbation of Schizophrenia: a comparison with Haloperidol and placebo. *Biological Psychiatry* 1997; **42**:233–246.
9. Huang Y, Hsu JC. Hochberg's step-up method: cutting corners off Holm's step-down method. *Biometrika* 2007; **94**, in press.
10. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 1976; **63**:655–660.
11. Stefansson G, Kim WC, Hsu JC. On confidence sets in multiple comparisons. In *Statistical Decision Theory and Related Topics IV*, Gupta SS, Berger JO (eds), vol. 2. Springer: New York, 1988; 89–104.
12. Finner H, Strassburger K. The partitioning principle: a powerful tool in multiple decision theory. *Annals of Statistics* 2002; **30**:1194–1213.
13. Hsu JC, Berger RL. Stepwise confidence intervals without multiplicity adjustment for dose–response and toxicity studies. *Journal of the American Statistical Association* 1999; **94**(446):468–482.
14. Dunnett CW, Sobel M. A bivariate generalization of Student's *t*-distribution, with tables for certain special cases. *Biometrika* 1954; **41**:153–169.
15. Siddiqui MM. A bivariate *t* distribution. *Annals of Mathematical Statistics* 1967; **38**(1):162–166.
16. Karlin S, Rinott Y. Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities. *Annals of Statistics* 1981; **9**(5):1035–1049.
17. Hsu JC. *Multiple comparisons: theory and methods*. Chapman and Hall: London, 1996.
18. Johnson ME. *Multivariate Statistical Simulation*. Wiley: New York, 1987.
19. Tamhane AC, Logan BR. Multiple test procedures for identifying the minimum effective and maximum safe doses of a drug. *Journal of the American Statistical Association* 2002; **97**(457):293–301.
20. Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis* (5th edn). Prentice-Hall: Englewood Cliffs, NJ, 2002.

APPENDIX: MODEL FOR DOSE–RESPONSE STUDIES WITH TWO ENDPOINTS

Suppose we have samples $Y'_{0r} = (Y^P_{0r}, Y^S_{0r})$, $r = 1, \ldots, n_0$ from a placebo group, and $Y'_{1r} = (Y^P_{1r}, Y^S_{1r})$, $r=1, \ldots, n_1$, $Y'_{2r}=(Y^P_{2r}, Y^S_{2r})$, $r=1, \ldots, n_2$ from two dose groups, measuring two endpoints. Let $\boldsymbol{\mu}_0 = (\mu^P_0, \mu^S_0)'$, $\boldsymbol{\mu}_1 = (\mu^P_1, \mu^S_1)'$, $\boldsymbol{\mu}_2 = (\mu^P_2, \mu^S_2)'$ be the group means,

$$\Sigma = \begin{pmatrix} (\sigma^P)^2 & \rho\sigma^P\sigma^S \\ \rho\sigma^P\sigma^S & (\sigma^S)^2 \end{pmatrix}$$

be the covariance matrix of ε_n, the nth row of $\boldsymbol{\varepsilon}$, and $\mathbf{1}_n$ be a column vector of 1s of length n. The model can be written in the following form:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y_0} \\ \mathbf{Y_1} \\ \mathbf{Y_2} \end{pmatrix} = \begin{pmatrix} Y^P_{01} & Y^S_{01} \\ \vdots & \vdots \\ Y^P_{0n_0} & Y^S_{0n_0} \\ Y^P_{11} & Y^S_{11} \\ \vdots & \vdots \\ Y^P_{1n_1} & Y^S_{1n_1} \\ Y^P_{21} & Y^S_{21} \\ \vdots & \vdots \\ Y^P_{2n_2} & Y^S_{2n_2} \end{pmatrix}$$

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/pst

Pharmaceutical STATISTICS

$$= \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_0^P & \mu_0^S \\ \mu_1^P & \mu_1^S \\ \mu_2^P & \mu_2^S \end{pmatrix} + \boldsymbol{\varepsilon} = \mathbf{X} \begin{pmatrix} \mu_0' \\ \mu_1' \\ \mu_2' \end{pmatrix} + \boldsymbol{\varepsilon}$$

The projection matrix

$$\mathbf{H} = \begin{pmatrix} \frac{1}{n_0}\mathbf{1}_{n_0}\mathbf{1}_{n_0}' & 0 & 0 \\ 0 & \frac{1}{n_1}\mathbf{1}_{n_1}\mathbf{1}_{n_1}' & 0 \\ 0 & 0 & \frac{1}{n_2}\mathbf{1}_{n_2}\mathbf{1}_{n_2}' \end{pmatrix}$$

$$= \begin{pmatrix} H_0 & 0 & 0 \\ 0 & H_1 & 0 \\ 0 & 0 & H_2 \end{pmatrix}$$

$$\mathbf{I} - \mathbf{H} = \begin{pmatrix} I - H_0 & 0 & 0 \\ 0 & I - H_1 & 0 \\ 0 & 0 & I - H_2 \end{pmatrix}$$

and $N\hat{\boldsymbol{\Sigma}} = N\mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} = N\sum_{i=0}^{2} \mathbf{Y}_i'(I - H_i)\mathbf{Y}_i$.

By Result 7.10. on p. 390 of [20], $N\hat{\boldsymbol{\Sigma}}$ follows the Wishart distribution $W_2(N-3, \boldsymbol{\Sigma})$, where $N = \sum_{i=0}^{2} n_i$, and $N-3$ is the degrees of freedom.

Notice, the diagonal elements of $N\hat{\boldsymbol{\Sigma}}$ are $(N-3)(\hat{\sigma}^P)^2$ and $(N-3)(\hat{\sigma}^S)^2$, where $\hat{\sigma}^P = \sqrt{\sum_{i=0}^{2}\sum_{r=1}^{n_i}(Y_{ir}^P - \bar{Y}_{i\cdot}^P)^2/(N-3)}$ and $\hat{\sigma}^S = \sqrt{\sum_{i=0}^{2}\sum_{r=1}^{n_i}(Y_{ir}^S - \bar{Y}_{i\cdot}^S)^2/(N-3)}$. To obtain the joint distribution of $(\hat{\sigma}^P/\sigma^P, \hat{\sigma}^S/\sigma^S)$, let $U = \mathrm{diag}\{(\sigma^P)^{-1}, (\sigma^S)^{-1}\}$, then

$$NU'\hat{\boldsymbol{\Sigma}}U \sim W_2\left(N-3, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

with diagonal elements $(N-3)(\hat{\sigma}^P/\sigma^P)^2$ and $(N-3)(\hat{\sigma}^S/\sigma^S)^2$.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/pst