CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER:

89884

STATISTICAL REVIEW(S)

JUN 6 1996

Nitroglycerin Transdermal System
Bio Control #BIO-018b
(Ref: ANDA #89-885; NTS, 0.4 mg/hr
ANDA #89-884; NTS, 0.2 mg/hr
ANDA #89-886; NTS, 0.6 mg/hr)
Reviewer: Sikta Pradhan
WP BIO96018.296

Hercon Laboratories, Co.
York, PA
Submission Date:
February 7, 1996

Review of the Statistical Analysis of the Bio-Control Document

OBJECTIVE:

The current Bio-Control Communication contains the reanalyzed statistical data submitted to the Agency for reconsideration.

The firm had previously (12/8/94) conducted a randomized 2-way crossover, single dose bioequivalence study on the 10%-reduced-size (13.5 cm^2) Hercon "face-adhesive" NTS-FA, 0.4 mg/hr patch comparing it with Transderm-Nitro^R 0.4 mg/hr patch of Ciba-Geigy. The study was conducted in healthy volunteers. The clinical study was conducted at ( ) under the supervision of ( ) Principal Investigator. The analytical study was conducted at ( ).

The bioequivalence study had been found unacceptable by the Division of Bioequivalence due to the following reasons:

1. For nitroglycerin (TNG) of test product, the 90% confidence interval for $LnAUC_{0-T}$, $LnAUC_{0-14}$, $LnAUC_{0-24}$, and LnC_{MAX} were within the 80% to 125% limit. However, the 90% confidence intervals for $LnAUC_{0-inf}$ remained outside the 80-125% limit.

2. For 1,2-dinitroglycerin (1,2-DNG) of test product, the 90% confidence intervals (CI) for $LnAUC_{0-T}$, $LnAUC_{0-14}$, $LnAUC_{0-24}$, $LnAUC_{0-inf}$ and LnC_{MAX} were within the acceptable range of 80-125%. However, for 1,3-dinitroglycerin (1,3-DNG) of test product, the 90% confidence intervals (CI) for $LnAUC_{0-T}$, $LnAUC_{0-14}$ and $LnAUC_{0-24}$ were outside the acceptable range of 80-125%.

Forty (40) volunteers (16 males and 24 females) were enrolled in the bioequivalence study. Four subjects (3 males and 1 female) withdrew from the study prematurely, and the study was completed by thirty-six (36) subjects. However, these subjects were dosed in two groups. The dosing of Group 2 in Period 1 was started (4/30/94) a week after the dosing of Group 1 in Period 2 (4/23/94). The firm did not provide any reason for this.

In the recent statistical analysis, the firm has deleted subject #121 indicating the subject is a fast metabolizer. The firm has also deleted subject #117 indicating she is a slow metabolizer. After elemination of these subjects, the study meets the 90% confidence intervals for both parent drug and its metabolites. The statistical analysis on Log-transformed data of treatment A versus treatment B was conducted using the following ANOVA MODEL:

RESPONSE = SEX + TREATMENT + SEQUENCE + GROUP + SEX*TREATMENT + PERIOD(GROUP) + TREATMENT*GROUP + SUBJECT(SEX*SEQUENCE*GROUP)

The TREATMENT*GROUP P-values and SEX*TREATMENT P-values calculated in the cases with no deletion, with deletion of only subject #121, and with deletion of both subjects #121 and #117 are presented below:

I. Including All Subjects:

Parameters		TNG	1,2-DNG	1,3-DNG	All Analytes
$LnAUC_{0-T}$	Sex*Treat	0.9653	0.6851	0.5659	0.7900
"	Treat*Group	0.0979	-----	-----	------
LnC_{MAX}	Sex*Treat	0.9727	0.6638	0.6165	0.7646
"	Treat*Group	0.0966	-----	------	------
$LnAUC_{0-inf}$	Sex*Treat	0.7612	0.3928	0.9843	0.4825
"	Treat*Group	0.1526	-----	------	------

II. Excluding Subject #121:

Parameters		TNG	1,2-DNG	1,3-DNG	All Analytes
$LnAUC_{0-T}$	Sex*Treat	0.9689	0.4646	0.9752	0.5309
"	Treat*Group	0.1108	-----	-----	------
LnC_{MAX}	Sex*Treat	0.9769	0.4502	0.8783	0.5087
"	Treat*Group	0.1096	-----	------	-----
$LnAUC_{0-inf}$	Sex*Treat	0.7187	0.3997	0.5769	0.5017
"	Treat*Group	0.1491	-----	------	------

III. Excluding Subjects #121 and #117:

Parameters		TNG	1,2-DNG	1,3-DNG	All Analytes
$LnAUC_{0-T}$	Sex*Treat	0.9376	0.2729	0.7519	0.3300
"	Treat*Group	0.1515	-----	-----	------
LnC_{MAX}	Sex*Treat	0.9291	0.2641	0.6556	0.3134
"	Treat*Group	0.1502	-----	------	------
$LnAUC_{0-Inf}$	Sex*Treat	0.7754	0.2083	0.8712	0.2824
"	Treat*Group	0.1835	-----	------	------

Comments:

1. According to the current policy of the Agency, the deletion of "outliers" on the basis of statistical arguments is not acceptable. There is no direct evidence that subject #121 is really a fast metabolizer and subject #117 is a slow metabolizer, except their plasma drug levels. Moreover, in a two-way crossover, single dose bioequivalence study, each subject will be exposed to both the test and the reference products in a similar manner.

2. The firm has carried out test for SEX*TREAT and TREAT*GROUP. If these tests are significant, then the study may have serious problems. If they are not significant, then it has been the standard procedure to drop these terms from the statistical model. Failure to drop these terms would result, using PROC GLM, in putting equal weight on the 13 males and 23 females and putting equal weight on the 20 subjects in Gruop 1 and the 16 subjects in Group 2.

3. In no case ($LnAUC_{0-T}$, $LnAUC_{0-Inf}$ and LnC_{MAX}), was SEX*TREAT anywhere near significant. However, in the case of trinitroglycerin, there was some borderline evidence of TREAT*GROUP interaction for $LnAUC_{0-T}$ (p=0.0979) and $LnAUC_{0-Inf}$ (p=0.0966) when all subjects are included in the analysis. **The firm did not do the test for TREAT*GROUP in the case on the 1,2-dinitroglycerin, 1,3-dinitroglycerin and "All Analytes" analysis (or at least, these tests are not reflected in the ANOVA Tables).**

4. As the dosing of Group 2 in Period 1 was started (4/30/94) a week after the dosing of Group 1 in Period 2 (4/23/94), the firm has to demonstrate that there is no evidence that the

difference between the products depends on the group, and for this reason, the firm was previously advised to conduct the analysis using the following statistical model:

Model Y = Seq Group Seq*Group Subj(Seq Group) Per(Group) Trt Trt*Group.

If Trt*Group is not significant (p> 0.10), Trt*Group could be dropped from the model. Then the following model could be used:

Model Y = Seq Group Seq*Group Subj(Seq Group) Per(Group) Trt;

or, Model Y = Seq Subj(Seq) Per(Group) Trt.

In either case, the period effect should be modeled as Per(Group), and not just Per.

Recommendations:

1. The statistical reanalyzed data of the in vivo bioequivalence study conducted by Hercon Laboratories on its Nitroglycerin Transdermal System Face Adhesive Patch (72 mg /13.5 cm^2) of 0.4 mg/hr, Lot #M0504NG/556 comparing it to Transderm-Nitro® patch (50 mg /20 cm^2) of 0.4 mg/hr, Lot #C5340 manufactured by Ciba Geigy is not acceptable for reasons cited in comments #2-4 above.

/S/

Sikta Pradhan, Ph. D.
Division of Bioequivalence
Review Branch I

RD INITIALED YCHUANG /S/ 5/24/96
FT INITIALED YCHUANG

Concur: /S/ Date: 6/6/96
Keith K. Chan, Ph. D.
Director, Division of Bioequivalence

cc: Bio Control #BIO-96-018 (Ref.ANDA # 89-885) (original, duplicate), HFD-652 (Huang, Pradhan), Drug File, Division File.

SP/03-5-96/X:\wpfile\Pradhan\BIO96018.296

Statistical Review: ANDA 89-886, Nitroglycerin Transdermal System, 0.6 mg/hr, Hercon Laboratories Corporation

Material reviewed: photocopies of material from ANDA 89-886, volumes 1 and 2 of 14. Data for my analyses were provided on diskette in data files sent to me by the Office of Generic Drugs.

Sikta Pradhan, Ph.D. is the Division of Bioequivalence reviewer for this submission. Most of the material in this review was previously communicated to Dr. Pradhan through electronic mail.

The issues in the review involve the sponsor's two-treatment, two-sequence, four-period replicated-crossover BE study (protocol HERC-9701) Three PK parameters (AUCt, AUCinf, and Cmax) for three analytes (trinitroglycerin [parent], 1,2-dinitroglycerin [1,2 metabolite], and 1,3-dinitroglycerin [1,3 metabolite]) were analyzed.

All PK parameters were statistically analyzed after log-transformation. These log-transformed parameters are designated as LAUCT=ln(AUCt), LAUCINF=ln(AUCinf), and LCMAX=ln(Cmax).

46 subjects (out of 49 subjects enrolled) completed the BE study.

The two treatments studied were:

treatment T - Nitroglycerin TDS 0.6 mg/hr, Hercon Laboratories Lot#LO597NG/613

treatment R - Transderm-Nitro® 0.6 mg/hr, Novartis Lot#1F193881

Forty-nine subjects, all male, were dosed in two groups. Three subjects, numbers 33, 44, and 45, dropped out of the study, leaving 46 subjects completing the study. The experimental design and subject numbers of the 46 subjects who completed the study are as follows:

	date (all dates 1997)							
	7/14	7/18	7/22	7/26	7/28	8/1	8/5	8/9
group 1								
sequence 1	**T**	**R**	**R**	**T**				
sequence 2	**R**	**T**	**T**	**R**				
group 2								
sequence 1					**T**	**R**	**R**	**T**
sequence 2					**R**	**T**	**T**	**R**

subject numbers:

group 1, sequence 1: 1 3 5 7 9 10 14 16 18 20 21 23 25 27

group 1, sequence 2: 2 4 6 8 11 12 13 15 17 19 22 24 26 28

group 2, sequence 1: 30 32 34 36 37 39 41 42 46 47

group 2, sequence 2: 29 31 35 38 40 43 44R 48

The sponsor did pre-analyses for Group-by-Treatment statistical interaction. They found no evidence of such an interaction. There was no great separation in time between the groups . It appears that the separation of the subjects into two groups was done to keep the number of subjects being dosed on one occasion to a manageable size. For this reason, I have not included Group-by-Treatment interaction in any of my statistical models.

Statistical Models

The statistical model assumed initially for the analyses was:

$$Y_{ijklm} = \mu + \xi_i + \alpha_j + s_{(ij)k} + \gamma_{(I)l} + \tau_m + (s\tau)_{(ij)km} + \varepsilon_{ijklm}$$

$$s_{(ij)k} \sim NID(0, \sigma_s^2) \quad (s\tau)_{(ij)km} \sim NID(0, \sigma_{s\tau}^2) \quad \varepsilon_{ijklT} \sim NID(0, \sigma_{WT}^2) \text{ and } \varepsilon_{ijklR} \sim NID(0, \sigma_{WR}^2)$$

where

Y_{ijklm} = the response (e.g. ln(AUCt), ln(Cmax), or ln(AUCinf)) for subject k in group I, sequence j receiving treatment m in period l

μ = a general mean

ξ_i = the effect of group I

α_j = the effect of sequence j

$s_{(ij)k}$ = the random effect of subject k in group i, sequence j

$\gamma_{(I)l}$ = the effect of period l in group I

τ_m = the effect of treatment m

$(s\tau)_{(ij)km}$ = the subject-by-treatment interaction for subject m in group I, sequence j receiving treatment m

ε_{ijklm} = the within-subject random error for subject m in group I, sequence j receiving treatment m in period l

Statistical analyses using this model were carried out using SAS PROC MIXED. For analyses without carryover effects, the SAS statements used initially were:

```
PROC MIXED MAXITER=500;
CLASSES GRP SEQ SUBJ PER TRT;
MODEL <y> = GRP SEQ GRP*SEQ PER PER*GRP TRT;
RANDOM SUBJ(GRP*SEQ) SUBJ*TRT(GRP*SEQ);
REPEATED/GRP=TRT SUB=SUBJ;
ESTIMATE 'T vs. R' TRT 1 -1/CL ALPHA=0.10;
```

where <y> is the particular response (LAUCT, LCMAX, LAUCINF) being analyzed. These SAS statements allow for possible subject-by-treatment interaction and also allow the within-subject variances of T and R to differ. If the estimated variance component associated with the subject-by-treatment interaction was zero, the statistical model was modified to the following:

$$Y_{ijklm} = \mu + \xi_i + \alpha_j + s_{(ij)k} + \gamma_{(I)l} + \tau_m + \varepsilon_{ijklm}$$

from which the $(s\tau)_{(ij)km}$ term has been deleted. All other terms are as described previously.

The SAS statements used to carry out analyses under this modified model were:

```
PROC MIXED MAXITER=500;
CLASSES GRP SEQ SUBJ PER TRT;
MODEL <y> = GRP SEQ GRP*SEQ PER PER*GRP TRT;
RANDOM SUBJ(GRP*SEQ);
REPEATED/GRP=TRT SUB=SUBJ;
ESTIMATE 'T vs. R' TRT 1 -1/CL ALPHA=0.10;
```

For analyses with carryover effects in the model, an additional factor was included in the CLASS and MODEL statements to reflect the carryover effects.

Analyses without Carryover Effects

If carryover effects are not included in the statistical model used in the analysis of this study, I obtain the following 90% confidence intervals:

	AUCt	Cmax	AUCinf
parent	99.60% , 113.39%	97.58% , 115.68%	98.70% , 112.94%
1,2 metabolite	99.75% , 109.06%	102.60% , 112.06%	100.19% , 109.33%
1,3 metabolite	94.50% , 102.46%	95.43% , 102.89%	95.59% , 103.37%

As may be seen, all of these 90% confidence intervals fall within the limits of 80% to 125%.

Analyses with Carryover Effects

The most surprising thing about the statistical models considered by the sponsor is that they included a term in their model for simple first-order carryover effects. There is no explanation in the text of the sponsor's Study Summary as to why this possibility was considered. In my experience, we have not considered the possibility of such carryover effects in the review of past Nitroglycerin TDS submissions.

If the possibility of carryover effects needs to be considered in the analysis of this study, the sponsor did not go far enough in examining the possibility of unequal carryover effects. Even if we assume (as we generally do) that only first-order carryover effects need to be considered - that is, that any carryover effect only lasts into the next period of the study (so we don't have to worry, for example, that a treatment given in the first period of the study has an effect on the response to a treatment given in the third period of the study) - we still have four possible first-order carryover effects:

the effect that administration of T has on the response to T given in the next period.

the effect that administration of T has on the response to R given in the next period.

the effect that administration of R has on the response to T given in the next period.

the effect that administration of R has on the response to R given in the next period.

The sponsor's model assumes that the effect of T on T is the same as the effect of T on R, and that the effect of R on T is the same as the effect of R on R.

If these carryover effects are present in the study, but we use a statistical model that does not include carryover effects, then the resulting estimate of the average difference between T and R may be biased. I carried out a statistical test to determine if this bias is statistically significant (alpha = 0.10). The p-values for the test of this bias are as follows:

	p-values for bias		
	AUCt	Cmax	AUCinf
parent	0.0001	0.0001	0.0001
1,2 metabolite	0.0003	0.0005	0.0001
1,3 metabolite	0.0003	0.0003	0.0001

In all cases, the p-value was less than 0.10 (all analyses were done after log transformation).

In examining the observations in this study, it does not appear that these highly statistically significant p-values for bias are due to a few "outliers". Nor can the significant p-values be attributed to *direct carryover* of drug or metabolite from one period of the study into the next, since the time-zero concentrations of parent and both metabolites were reported as zero for all subjects in periods 2, 3, and 4.

The question of whether the possibility of unequal carryover effects must be considered in the analysis of this study is a medical/biological/pharmacokinetic question, not a statistical question. As such, the final decision as to whether the possibility of carryover effects must be considered in the analysis of this study must be made by the Division of Bioequivalence. If carryover effects (as described above) are included in the statistical model, the 90% confidence intervals are:

	AUCt	Cmax	AUCinf
parent	63.24% , 84.87%	55.66% , 82.14%	60.79% , 81.95%
1,2 metabolite	76.52% , 93.73%	79.57% , 97.23%	75.89% , 92.52%
1,3 metabolite	75.60% , 90.09%	77.13% , 91.10%	75.07% , 89.03%

As may be seen, with carryover effects included in the statistical model the resulting 90% confidence intervals do not fall within the limits of 80% to 125% for any PK parameter, for any of the three analytes.

Summary

1. In the sponsor's own statistical analysis, they included a term in their statistical model for simple first-order carryover effects. In their study report the sponsor does not explain why they did this. To the best of my knowledge, the possibility of unequal carryover

effects has not been considered previously in the statistical review and analysis of Nitroglycerin TDS products BE studies.

2. If the possibility of first-order carryover effects does have to be considered, the sponsor's model is inadequate to the task, since it does not consider all of the possible first-order carryover effects that could be present in this replicated-crossover BE study.

3. In testing for the possibility of bias due to unequal carryover effects, highly statistically significant bias (p-values ranged from 0.0001 to 0.0005) was found in for all nine combinations of PK parameter (AUCt, Cmax, and AUCinf) and analyte (parent, 1,2 metabolite, and 1,3 metabolite).

4. The final decision as to whether the possibility of unequal carryover effects needs to be considered in the review of this BE study must be made by the Division of Bioequivalence.

5. If carryover effects are not included in the statistical model, the 90% confidence intervals fall within the limits of 80% to 125% for all three PK parameters and for all analytes.

/S/

Donald J. Schuirmann
Expert Mathematical Statistician
Quantitative Methods & Research staff

/S/ 2/16/99

Concur: Stella Green Machado, Ph.D.
Director, Quantitative Methods & Research staff

cc:
Original ANDA 89-886
HFD-652 Sikta Pradhan
HFD-652 Yih Chain Huang
HFD-651 Rabindra N. Patnaik
HFD-650 Dale P. Conner
HFD-600 Douglas L. Sporn
HFD-615 Harvey A. Greenberg
HFD-705 QMR Chron
HFD-705 Stella G. Machado
HFD-705 Donald J. Schuirmann

89-886 4-1

Statistical Review: ANDA 89-885, Nitroglycerin Transdermal System 0.4 mg/hr, Hercon Laboratories

Material reviewed: three orange-colored volumes of ANDA 89-885 - volume 7.1, volume 7.2, and an unnumbered volume with a cover letter dated November 22, 1996 - plus a copy of a data diskette provided by the sponsor.

Sikta Pradhan, Ph.D. is the Division of Bioequivalence reviewer for this submission. The material in this review was previously communicated to Dr. Pradhan through electronic mail.

The issues in this review involve the sponsor's original two-treatment, two-period crossover bioequivalence (BE) study (Protocol 567794, carried out in two groups of subjects) and a retest of subject #121 from the original study. Three PK parameters (AUCt, AUCinf, and Cmax) for three analytes (trinitroglycerin [parent compound], 1,2 dinitroglycerin, and 1,3 dinitroglycerin) were reviewed, for a total of nine outcome variables.

Gender and Group Differences

Of the 36 subjects who completed the original BE study (out of 40 subjects who were entered into the study), 23 were females and 13 were males. In addition, the original study was carried out in two groups of subjects. In group 1, 20 subjects (12 females and 8 males) completed the study, and in group 2, 16 subjects (11 females and 5 males) completed the study. Period 1 of group 2 began one week after period 2 of group 1.

The sponsor has carried out analyses to assess various effects of gender and group. In SAS PROC GLM terminology, the sponsor's initial full statistical model (as presented in the SAS outputs of Appendix C.1, volume 7.1) was

```
CLASS SEX TREAT GROUP SEQUENCE SUBJECT PERIOD;
MODEL response = SEX + TREAT + GROUP + SEQUENCE +
      SEX*TREAT + GROUP*SEQUENCE + TREAT*GROUP +
      PERIOD(GROUP) + SUBJECT(SEX*SEQUENCE*GROUP);
```

In testing for the main effects of gender and group, the sponsor has used the wrong error term. Because gender and group are between-subject factors, the appropriate error term is the SUBJECT(SEX*SEQUENCE*GROUP) mean square, not the "Error" mean square. Nevertheless, conclusions when the proper error term is used agree qualitatively with the conclusions reached by the sponsor. For log-transformed parameters (lauct=ln(AUCt), laucinf= ln(AUCinf), lcmax=ln(Cmax)) there is no evidence of a difference between the two groups ($p > 0.2$ for all parameters and analytes), but strong evidence for a difference between males and females (p less than or equal to 0.002 for all parameters and analytes).

The sponsor has also looked at gender-by-treatment and group-by-treatment interaction. There was no evidence of gender-by-treatment interaction ($p > 0.43$ for all parameters and analytes). That is, there was no evidence that the relative performance of the test product (T) and the reference product (R) was not consistent between males and females. For this reason, the SEX*TREAT term was removed from the statistical model for final inferences concerning bioequivalence.

The test for group-by-treatment interaction is a test to see if there is evidence that the relative performance of the test product (T) and the reference product (R) was not consistent between the two groups of subjects. In the past, this test has sometimes been recommended for cases where bioequivalence studies have been carried out in two groups. In cases where the two groups are studied at different clinical sites, or at the same site but widely separated in time, there may be some question as to whether the phenomenon of interest - the relative performance of T and R - is the same between the two groups. When the question arises, I have recommended carrying out the test for group-by-treatment interaction at the 0.10 level of significance. If it is not significant, the decision would be to remove the group-by-treatment term from the statistical model and assess equivalence using the data from both groups. If it is significant, the validity of the study might be called into question, and it might be required to demonstrate equivalence in both groups separately.

In the original Hercon study, the test for group-by-treatment was significant ($p < 0.10$) in two cases: lauct and laucinf for the parent compound. However, the sponsor states that the separation of the subjects into two groups was done for purely logistical reasons. Period 1 of the second group began a week after period 2 of the first group. Furthermore, at a meeting with the sponsor on August 23, 1996 it was stated that the entire study (both groups) were completed before the blood samples were chemically assayed. This raises the question of whether concern that the relative performance of T and R was different in the two groups is a reasonable concern based on *a priori* scientific considerations. If it is not a reasonable concern, then it would be appropriate to delete the group-by-treatment term from the statistical model regardless of the p-value.

If the decision of the Division of Bioequivalence is that the possibility of group-by-treatment interaction must be considered for the original study, then we must consider the question of what to do about lauct and laucinf for the parent compound. The approach the sponsor has used is to retain the group-by-treatment term in the statistical model and then use the data from both groups to assess equivalence of T and R (indeed, the sponsor has kept group-by-treatment in the model for all of the parameters and analytes). This has the effect of putting equal weight on the estimates of T vs. R from the two groups, even though the sample sizes were unequal (20 subjects in group 1, 16 subjects in group 2). Is this the appropriate weighting? An alternative approach would be to assess equivalence separately in the two groups. Because of the reduced sample sizes involved, this would not lead to a conclusion of equivalence.

In summary, it seems to me that there is a good argument for dismissing the possibility of group-by-treatment interaction based on the fact that the groups were not widely separated

in time (period 1 of group 2 began a week after period 2 of group 1) and the blood samples from the two groups were not chemically assayed until after the completion of both groups. Unless there was some evidence that the subject populations from which the subjects were recruited were different for the two groups, I don't see a plausible explanation for a group-by-treatment interaction. However, the final decision on this issue must be made by the Division of Bioequivalence.

Outlier Issue

The sponsor maintains that subject #121 in the original BE study was an outlier, and that his data should be deleted from statistical analyses to assess equivalence.

Subject #121 had the lowest AUCt and AUCinf for either product (the test product, T or the reference product, R) on all three analytes. Subject #121 also had the lowest Cmax for either product on the parent compound and the 1,2 metabolite, and the lowest Cmax for the reference product on the 1,3 metabolite (Subject #121 had the third lowest Cmax for the test product 1,3 metabolite). This in itself is not remarkable - in any dataset, one of the observations has to be the lowest, and in a crossover study where within-subject correlations may be high, the subject who has the lowest observation on some variables might well have the lowest observation on all or most variables.

To see if any observations were statistical outliers on a between-subject basis, I looked at the log-transformed parameters for each product. Testing was done separately in males and females, since (as discussed above) the data showed significantly higher levels of all three analytes in females relative to males.

For the test product, both lauct and laucinf for subject #121 would be declared an outlier for the 1,3 metabolite based on the maximum absolute studentized residual, testing at the 0.05 level of significance and using critical values as described by Lund (Lund, R.E. (1975) Tables for an approximate test for outliers in linear models, *Technometrics*, 17(4):473-476). No other parameter or analyte for the test product, and no parameter or analyte for the reference product would be declared an outlier based on this test.

A more important consideration is whether there were any within-subject outliers. In the case of a standard two-treatment, two-period crossover study, the tests for bioequivalence that we use depend on the within-subject test-minus-reference differences of the log-transformed parameters. If we look at the studentized residuals from the statistical model used in the bioequivalence analyses, two subjects show up as within-subject outliers based on Lund's test - subject #115 for lauct and laucinf for the parent compound, and subject #121 for lauct and laucinf for the 1,3 metabolite.

So subject #121 does show up as a statistical outlier for lauct and laucinf of the 1,3 metabolite, both on a within-subject basis and, for the test product, on a between-subject

basis. However, the Center has stated quite clearly in the July 1992 GUIDANCE - STATISTICAL PROCEDURES FOR BIOEQUIVALENCE STUDIES USING A STANDARD TWO-TREATMENT CROSSOVER DESIGN

> " In principle, however, outliers cannot be dropped from the analysis of the data solely on the basis of a statistical test. Sponsors who have identified one or more outliers should provide scientific evidence or explanations to justify the exclusion of the subject(s) data from statistical analysis."

The question then is: WHY did subject #121 show such discordant results? This question cannot be answered based on statistical analysis of the original study results.

Retest

The sponsor has carried out a retest of subject #121. In the retest, the test product (T) and the reference product (R) were administered to the subject in the order R-T-T-R.

In my opinion, there are two facts that limit the usefulness of the retest data:

1. The products studied in the retest were different lot numbers from the products studied in the original BE study. According to page vii of the sponsor's study report on the original BE study (in Section VI A & B, "Statistical Re-Analysis", of volume 7.1, stamped 0011 at the bottom of the page), the study medications were:

   Treatment A (Test): NTS FA-13.5,
   Hercon Laboratories Lot# M0504NG/556

   Treatment B (Reference): Transderm-Nitro 20 square-cm,
   Ciba-Geigy Lot# C5340

   However, in the retest, according to page 2 of the sponsor's study report on the retest (in section VIA. & VIB., "ATTACHMENT #1 Supplemental Report", of the unnumbered volume, stamped 0007 at the bottom of the page), the study medications were:

   Treatment A (Test): NTS FA-13.5,
   Hercon Laboratories Lot# S1396NG/583

   Treatment B (Reference): Transderm-Nitro 20 square-cm,
   Ciba-Geigy Lot# 1M010807

2. The sponsor did not retest any subjects other than the suspected outlier, subject #121. It has long been argued by the Division of Bioequivalence that cross-study

comparisons are not valid. If a few additional subjects - not suspected to have been outliers in the original study - in addition to #121 had been retested, then it might have been possible to contrast the relative performance of these subjects in the retest with the relative performance of these subjects in the original study. With only subject #121 being retested, such within-study comparisons are not possible.

The retest results are reported in Appendix A of the unnumbered volume. These results plus the results for subject #121 from the original study (in hr x pmol/ml for AUC's and pmol/ml for Cmax) are:

parent compound

	Cmax	AUCt	AUCinf
original study			
T, period 1	0.3003	3.2815	3.3010
R, period 2	0.3800	4.8698	4.9060
retest			
T, period 2	0.4025	7.5590	7.5996
T, period 3	0.5328	10.0598	10.0984
geometric mean	0.4631	8.7202	8.7604
R, period 1	0.4492	8.2323	8.2948
R, period 4	1.3387	13.7910	13.8193
geometric mean	0.7755	10.6551	10.7065

1,2 metabolite

	Cmax	AUCt	AUCinf
original study			
T, period 1	7.2492	86.2620	87.9620
R, period 2	8.6221	159.4200	161.0790
retest			
T, period 2	10.8188	223.5840	226.2920
T, period 3	10.1049	211.2570	213.2260
geometric mean	10.4558	217.3331	219.6619
R, period 1	10.4344	208.6670	210.3930
R, period 4	13.5098	296.4540	298.8890
geometric mean	11.8729	248.7170	250.7671

1,3 metabolite

	Cmax	AUCt	AUCinf
original study			
T, period 1	1.3894	10.9438	10.9438
R, period 2	1.5761	33.4395	34.5044
retest			
T, period 2	2.0704	44.5727	45.5385
T, period 3	1.9276	41.3408	42.1495
geometric mean	1.9977	42.9263	43.8112
R, period 1	2.2571	47.1525	47.9353
R, period 4	3.1303	66.1747	67.7750
geometric mean	2.6581	55.8597	56.9984

The following table presents the test-over-reference ratios from the original study for subject #121 and the test-over-reference ratios of geometric means from the retest of subject #121:

	Cmax	AUCt	AUCinf
parent compound			
T/R original study	0.7903	0.6738	0.6728
T/R retest	0.5972	0.8184	0.8182
1,2 metabolite			
T/R original study	0.8408	0.5411	0.5461
T/R retest	0.8806	0.8738	0.8760
1,3 metabolite			
T/R original study	0.8815	0.3273	0.3172
T/R retest	0.7516	0.7685	0.7686

Levels of all three analytes were notably higher in the retest than in the original study. There was also some indication of period effects in the retest - particularly illustrated by the much higher levels obtained from the reference product in period 4 compared to period 1. Statistically significant period effects had been seen in the analysis of the original study.

For the 1,2 and the 1,3 metabolites, the T/R ratios were more consistent between Cmax and the AUC's in the retest than in the original study. However, this was not true of the parent compound.

In the case of a retest, an outcome that might have supported the view that subject #120 was an outlier in the original study is the one where the reference product results are much like those seen in the original study but the test product results are notably higher, at least for the 1,3 metabolite. That outcome was not seen in the actual retest results. Both the test AND the reference results were notably higher (this was true for all PK parameters and all analytes, not just AUC's from the 1,3 metabolite).

It is true that the retest T/R ratios seen for the AUC's from the 1,3 metabolite, while low (about 77%), were not as extremely low as in the original study. The observed ratios of the T/R ratios (original over retest) were 0.3273/0.7685 = 42.6% for AUCt and 0.3172/0.7686 = 41.3% for AUCinf. Are these observed ratios statistically significantly lower than 100%?

If we assume that the T/R ratio seen in the retest is not biased by any period effects (there is no way to test this assumption, since the sponsor did not retest any subjects other than #121), and if we assume that the within-subject variability estimated in the original study is applicable to the retest as well, then we may compute a confidence interval for the ratio of the underlying geometric mean T/R ratio for #121 in the original study over the underlying geometric mean T/R ratio for #121 in the retest. The estimated within-subject variance (using a reduced statistical model without gender-by-treatment or group-by-treatment) was 0.029203 for 1,3 metabolite lauct and 0.029376 for 1,3 metabolite laucinf (both estimates based on 33 degrees of freedom). The resulting confidence intervals are

	observed ratio	90% confidence interval
AUCt	42.6%	25.8% to 70.3%
AUCinf	41.3%	25.0% to 68.2%

Since neither of these confidence intervals contains 100%, we would conclude that the geometric mean T/R ratio for subject #121 in the original study was statistically significantly lower than the geometric mean T/R ratio for subject #121 in the retest (subject to the validity of the assumptions we had to make in order to calculate the confidence intervals). But this, of course, does not answer the question of WHY it was lower.

The blood level-time profiles for subject #121, both for the original study and for the retest, are illustrated on pages 3-5 of section VIA. & VIB., "ATTACHMENT #1 Supplemental Report", of the unnumbered volume (stamped 0008-0010 at the bottom of the pages). The interesting feature of the test product profiles from the original study (as noted by the sponsor) is that the levels of the 1,3 metabolite begin to decline after the 8 hour sample, going below the limit of detection around 14 hours. The levels of the 1,2 metabolite and the parent compound also drop markedly after the 8 hour sample. This pattern is not seen in either of the test product profiles in the retest, nor was it seen in the profiles for the reference product.

What could have caused such a profile in the original study? Could it represent a product failure? If it represents a product failure, since it was seen with the test product, does it have implications for the equivalence of the products? Could it have been due to some characteristic of subject #121? If so, why wasn't it seen in the retest? Was it because a different lot number was used in the retest?

In summary, because the sponsor did not retest any subjects besides #121, it is not possible to carry out a formal statistical test to verify that the relative performance of T and R was consistent between studies for the other subjects but inconsistent for subject #121. By making certain assumptions, we were able to conclude that the geometric mean T/R ratio for subject #121 in the original study was lower than the geometric mean T/R ratio for subject #121 in the retest (for lauct and laucinf of the 1,3 metabolite), but this does not tell

us WHY it was lower. We are therefore led to making a judgment about the original results and the retest results based on the sciences of pharmacokinetics, biopharmaceutics, and medicine. Since this is not a statistical judgment, the final decision must be made by the Division of Bioequivalence.

Log-Transformed vs. Untransformed Analysis

The sponsor has submitted various simulation studies related to the issue of whether the PK parameters from this study should be analyzed after log-transformation or untransformed. If the decision is made that the data from subject #121 should be deleted from the analyses (either for all analyses or just for AUCt and AUCinf for the 1,3 metabolite), then this question is irrelevant, since the product would be declared equivalent using either approach.

We have maintained that the sample sizes (for example, n=36 in the original Hercon study) seen in typical crossover BE studies are too small to make a reliable determination as to whether the log-transformed or untransformed parameters better meet the assumptions underlying the statistical bioequivalence analysis. The decision to use the log transformation for AUC and Cmax in blood-level BE studies was therefore made mostly on the basis of theoretical pharmacokinetic arguments. This has been presented in the July 1992 Guidance.

The question arises as to whether the pharmacokinetic arguments that apply to PK parameters from orally administered dosage forms also apply to PK parameters from Transdermal Systems. We have heard no arguments from the Division of Bioequivalence that the same reasoning that leads us to use the log transformation in the case of orally administered dosage forms does not apply as well to nitroglycerin transdermal patches. If there are such arguments, the Division of Bioequivalence should inform the Quantitative Methods and Research staff as soon as possible. In the absence of such arguments, however, the material presented by the sponsor would not lead me to abandon the log transformation in this case.

Summary

1. There is no evidence of gender-by-treatment interaction in the original study.

2. There is some evidence ($p < 0.10$) of group-by-treatment interaction for parent compound lauct and laucinf in the original study. However, the way the study was conducted, with the start of the second group beginning only a week after the finish of the first group, and with the assay of the blood samples taking place after both groups were completed, raises the question of whether group-by-treatment interaction was a realistic possibility.

3. Because the sponsor did not retest any subjects other than the suspected outlier (subject #121), there is little we can do in the way of formal statistical tests for determining whether or not subject #121 did not reproduce his performance in the original study, while the other retested subjects did. Under certain assumptions, we may conclude that the T/R ratio seen in subject #121 in the original study was statistically significantly lower than the T/R ratio seen in subject #121 in the retest (for lauct and laucinf of the 1,3 metabolite). However, this does not answer the question of WHY it was lower. Furthermore, the products tested in the retest were different lot numbers than the products tested in the original study.

4. The final decision as to whether the data from subject #121 may be deleted from the statistical analyses of the original study should be based on pharmacokinetic/biopharmaceutic/
medical judgment. As such, this decision must be made by the Division of Bioequivalence.

5. We are not aware of any arguments that the reasoning underlying the use of the log transformation for AUC and Cmax from orally administered dosage forms does not also apply to nitroglycerin transdermal patches. In the absence of such arguments, data from 36 subjects is insufficient to make a judgment as to whether the assumptions underlying our statistical bioequivalence analyses are better met by the untransformed parameters rather than the log-transformed parameters, for the original Hercon study.

/S/ 8/29/97
Donald J. Schuirmann
Expert Mathematical Statistician
Quantitative Methods & Research staff

/S/ 8/29/97
Concur: Charles Anello, Sc.D.
Deputy Director, Office of Epidemiology
& Biostatistics