

**CENTER FOR DRUG EVALUATION AND
RESEARCH**

APPLICATION NUMBER:
21-166

STATISTICAL REVIEW(S)

Memorandum of Statistical Review

NDA: 21-166

Name of Drug: ESTROGEL® _____

Indication: Treatment of vasomotor symptoms

Applicant: Solvay Pharmaceuticals, Inc.

Documents: \\Cdsesub1\N21166\N_000\2003-10-23\FDA Request - Sep 2003

Clinical Reviewer: Phill Price, M.D., HFD-580

Project Manager: George Lyght, HFD-580

Dates: Received 10/23/03; User Fee 2/9/04

Reviewer: Kate Meaker, M.S., HFD-715

Biometrics Team Leader; Mike Welch, Ph.D.

Background

Two studies were submitted to evaluate the efficacy of two doses of Estrogel for the indication of treatment of vasomotor symptoms. Both studies were randomized, double-blind, parallel arm, multicenter studies in the appropriate patient population. Study CV141-001 (001) had three treatment groups: Estrogel 1.25g, Estrogel 2.5g, and placebo. Study CV141-002 (002) had four treatment groups: Estrogel 0.625g, Estrogel 1.25g, Estrogel 2.5g, and an active-control (Climara patch). Study 002 did not have a placebo-control group. The applicant has requested consideration of Estrogel 1.25g and 2.5g doses.

This NDA was original submitted to the Division in August, 1999, with a User Fee date of August, 2000. A statistical review of the original submission was completed in May, 2000. However, prior to an action being taken on the NDA, all submissions from this sponsor were put on hold under the Application Integrity Policy (AIP). The AIP hold was lifted in April, 2003, which reset the User Fee date for this NDA to February, 2004. Electronic data sets and files were submitted in October, 2003. These included additional analyses of the same studies submitted originally, but the sponsor did not submit any new studies.

This memorandum covers three issues. The first is the reanalysis of the original primary endpoint, the frequency of moderate-to-severe vasomotor symptoms for Study 001. The second is an analysis of the severity score endpoint, a requirement since added by the Division as a co-primary endpoint for this indication. Lastly, the supportive evidence from Study 002 is presented for these endpoints.

Analysis of Frequency of Moderate-to-Severe Vasomotor Symptoms (MSVS)

Frequency of MSVS was planned as the primary outcome for Study 001. The sponsor did not specify which week(s) were of primary interest. DRUDP requires comparison to placebo at Week 4 and Week 12 to support efficacy. Week 8 is reported as a secondary timepoint. The dependent variable is the change from baseline in the weekly average number of MSVS.

In the protocol, the planned analysis was an ANCOVA model with treatment, center, and treatment-by-center interaction terms, with baseline frequency of moderate-to-severe vasomotor symptoms as the covariate. Dunnett's test for multiple comparisons was planned to account for two EstroGel dose levels, each being compared to placebo. An interim analysis was planned, with a Lan and DeMets implementation of the O'Brien/Fleming approach. The interim analysis was conducted at $\alpha=0.001$ and the final analysis was done at $\alpha=0.049$.

In the original NDA submission, the planned ANCOVA model did not show statistically significant differences for the frequency endpoint for either of the EstroGel dose groups versus placebo at Week 4, but did show statistically significant differences for both EstroGel doses at Week 12. The statistical review of the original NDA submission (completed May 8, 2000) focused on visit-wise comparisons and analyses by quartiles of baseline scores.

In this review, examination of the residuals from the planned ANCOVA model showed that the normality assumption was questionable. Therefore, van Elteren's non-parametric test was used as an alternative. Instead of the planned Dunnett's test for multiple comparisons, the Hochberg method was applied to the non-parametric results. The analyses used $\alpha=0.049$, as planned in the protocol for the final analyses. Comparisons of each EstroGel dose group to placebo were made at Weeks 4, 5, 8, and 12. Weeks 4 and 12 denote the primary comparison timepoints.

The results of Study 001, using the van Elteren's non-parametric test, and Hochberg's adjustment for multiple dose comparisons at Week 4 and Week 12, are shown in Table 1. Both EstroGel dose groups are statistically significantly different from placebo in the mean change from baseline in number of MSVS.

However, the magnitude of the difference versus placebo is not as large as would be expected typically for this indication. The Division generally expects to see a difference of 2 MSVS per day over placebo. As shown in Table 1, neither dose group achieves that level by Week 4. The EstroGel 2.5g dose approaches that level at Week 5 (difference=1.98), and exceeds it through Week 12. The EstroGel 1.25g dose group never achieves that level versus placebo through the 12 weeks on treatment. The Medical Officer will need to address the issue of the effect size in his review.

Table 1: Study 001 – Frequency of Moderate-to-Severe Hot Flushes

	Placebo n=73	ESTROGEL 1.25g n=72	ESTROGEL 2.5g n=71
Baseline			
Mean	11.01	10.33	10.52
Std. Dev.	5.66	3.07	3.88
Week 4[°]			
Mean	5.95	4.43	4.28
Std. Dev.	5.17	4.13	4.37
Mean Change from baseline	-5.06	-5.91	-6.24
Std. Dev. – Change from baseline	4.91	3.68	4.40
Diff. vs. Placebo p-value *		0.85 0.019 **	1.18 0.020 **
Week 5			
Mean	5.70	3.99	3.23
Std. Dev.	5.98	3.98	4.18
Mean Change from baseline	-5.31	-6.34	-7.29
Std. Dev. – Change from baseline	4.43	3.71	4.56
Diff. vs. Placebo p-value *		1.03 0.065	1.98 < 0.001
Week 8			
Mean	5.36	3.44	2.33
Std. Dev.	5.78	4.40	4.16
Mean Change from baseline	-5.65	-6.89	-8.19
Std. Dev. – Change from baseline	4.11	3.80	5.07
Diff. vs. Placebo		1.24	2.54
Week 12[°]			
Mean	5.17	2.79	1.96
Std. Dev.	6.52	3.70	4.23
Mean Change from baseline	-5.84	-7.55	-8.56
Std. Dev. – Change from baseline	4.52	3.52	5.13
Diff. vs. Placebo p-value *		1.71 0.043 **	2.72 < 0.001 **

* p-values from Van Elteren's non-parametric test

** Statistically significantly different from placebo at $\alpha=0.049$, using Hochberg method to adjust for multiple comparisons at each time point.

[°] Primary Timepoint

Analysis of Severity Score

The severity score is calculated as follows. The numerator of the severity score is a weighted sum of vasomotor symptoms, with mild receiving a weight of 1, moderate a weight of 2, and severe a weight of 3. This is divided by the total number of mild, moderate, and severe vasomotor symptoms experienced by a patient.

This endpoint was not reviewed previously. In the protocol, this was planned to be analyzed as a secondary variable. That was acceptable at the time of the protocol. Since then, the Division has changed the evaluation of efficacy for this indication. The severity score is now considered as a co-primary efficacy variable.

The same ANCOVA model which was used for the frequency of vasomotor symptoms endpoint was also applied to this endpoint. A check of the residuals showed that the normality assumption was not met. Instead, van Elteren's non-parametric method, with the Hochberg adjustment approach, was used. The analysis used $\alpha=0.049$, as in the analysis of the frequency endpoint. Since the Division considers frequency and severity as co-primary endpoints, no additional adjustment to the alpha level was required.

The results for the severity score outcome in Study 001 are shown in Table 2. The EstroGel 1.25g dose group is statistically significantly different from placebo at Week 4. The EstroGel 2.5g dose group is at best marginally significant ($p=.053$). The Division is interested in exploring the statistical significance of the results at Week 5. At Week 5, both EstroGel dose groups are statistically significantly different from placebo. The same is true at Week 12. No expected minimum effect size has been defined for this endpoint.

APPEARS THIS WAY
ON ORIGINAL

Table 2: Study 001 – Hot Flushes Severity Score

	Placebo n=73	ESTROGEL 1.25g n=72	ESTROGEL 2.5g n=71
Baseline			
Mean	2.30	2.36	2.29
Std. Dev.	0.24	0.29	0.22
Week 4[◊]			
Mean	2.00	1.73	1.67
Std. Dev.	0.63	0.73	0.85
Mean Change from baseline	-0.31	-0.63	-0.61
Std. Dev. – Change from baseline	0.62	0.71	0.85
Diff. vs. Placebo		0.32	0.30
p-value *		0.005 **	0.053
Week 5			
Mean	1.92	1.72	1.45
Std. Dev.	0.69	0.78	0.89
Mean Change from baseline	-0.38	-0.64	-0.84
Std. Dev. – Change from baseline	0.69	0.78	0.86
Diff. vs. Placebo		0.26	0.46
p-value *		0.040 **	0.003 **
Week 8			
Mean	1.89	1.44	1.19
Std. Dev.	0.77	0.90	0.94
Mean Change from baseline	-0.41	-0.92	-1.09
Std. Dev. – Change from baseline	0.78	0.89	0.93
Diff. vs. Placebo		0.51	0.68
Week 12[◊]			
Mean	1.76	1.33	0.98
Std. Dev.	0.84	0.97	0.94
Mean Change from baseline	-0.54	-1.03	-1.30
Std. Dev. – Change from baseline	0.84	0.94	0.94
Diff. vs. Placebo		0.49	0.76
p-value *		<0.001 **	<0.001 **

* p-values from Van Elteren's non-parametric test

** Statistically significantly different from placebo at $\alpha=0.049$, using Hochberg method to adjust for multiple comparisons at each time point.

[◊] Primary Timepoint

Results of Study 002 as Supportive Evidence

This study included treatment groups for the two desired dose levels of EstroGel (1.25g and 2.5g), along with a group which received a lower dose of EstroGel (0.625g) and an active-control group which received an open-label Climara patch. The original statistical review compared the two desired EstroGel dose groups to the 0.625 EstroGel group, and found no statistically significant differences. No comparisons to the active-control group were done. However, these comparisons do not address the decision on the efficacy of the desired two EstroGel doses.

Study 002 did not have a placebo group, so the comparisons of EstroGel to placebo needed to make conclusions about efficacy cannot be done. The results from this study do provide supportive evidence regarding the level of efficacy in terms of the observed effect size. As shown in Tables 3 and 4, the treatment effect sizes for both frequency of MSVS and severity score are similar to study 001 across all time points.

Table 3: Study 002 – Frequency of Moderate-to-Severe Hot Flushes

	ESTROGEL 1.25g n=90	ESTROGEL 2.5g n=84
Baseline		
Mean	11.75	11.84
Std. Dev.	5.37	4.74
Week 4[◇]		
Mean	5.08	4.27
Std. Dev.	4.63	5.00
Mean Change from baseline	-6.67	-7.57
Std. Dev. – Change from baseline	5.21	5.33
Week 5		
Mean	4.37	3.61
Std. Dev.	4.35	4.25
Mean Change from baseline	-7.38	-8.23
Std. Dev. – Change from baseline	4.75	5.57
Week 8		
Mean	3.09	2.58
Std. Dev.	3.69	3.85
Mean Change from baseline	-8.66	-9.26
Std. Dev. – Change from baseline	5.63	5.56
Week 12[◇]		
Mean	2.89	2.29
Std. Dev.	3.63	3.80
Mean Change from baseline	-8.86	-9.55
Std. Dev. – Change from baseline	5.74	5.63

[◇] Primary Timepoint

Table 4: Study 002 – Hot Flushes Severity Score

	ESTROGEL 1.25g n=90	ESTROGEL 2.5g n=84
Baseline		
Mean	2.41	2.32
Std. Dev.	0.35	0.31
Week 4^o		
Mean	1.90	1.52
Std. Dev.	0.81	0.89
Mean Change from baseline	-0.52	-0.81
Std. Dev. – Change from baseline	0.73	0.91
Week 5		
Mean	1.71	1.44
Std. Dev.	0.90	0.92
Mean Change from baseline	-0.70	-0.88
Std. Dev. – Change from baseline	0.83	0.92
Week 8		
Mean	1.43	1.21
Std. Dev.	1.04	0.97
Mean Change from baseline	-0.98	-1.12
Std. Dev. – Change from baseline	0.95	0.98
Week 12^o		
Mean	1.39	1.00
Std. Dev.	1.09	0.96
Mean Change from baseline	-1.02	-1.32
Std. Dev. – Change from baseline	1.04	0.99

^o Primary Timepoint

Summary

The results of Study 001 support the efficacy of both the Estrogel doses for the indication of the treatment of vasomotor symptoms. For the frequency of moderate-to-severe vasomotor symptoms (MSVS) endpoint, both the 1.25g and 2.5g dose groups are significantly different from placebo at Week 4 and Week 12. For the severity score endpoint, the 1.25g dose group is statistically significantly different from placebo at Week 4 and Week 12. While the Estrogel 2.5g dose group did not achieve clear statistical significance at Week 4 ($p=.053$), it was statistically significantly different from placebo at Week 5 and Week 12.

Study 002 did not have a placebo-control group, so it was not possible to make statistical comparisons for this study. However, the descriptive statistics of the effect sizes for each dose group were similar to those seen in Study 001.

The remaining clinical review issue is whether the magnitude of the difference between Estrogel and placebo in the change in frequency of MSVS represents a clinically

meaningful effect. Typically a reduction of 2 MSVS per day more than the placebo group is expected. In Study 001, the Estrogel 2.5g dose group achieved this level, but the 1.25g dose did not. An expected clinical effect for the severity score endpoint has not been established.

APPEARS THIS WAY
ON ORIGINAL

**Statistical Review and Evaluation
Clinical Studies**

NDA#: 21-166

Applicant: Unimed

Name of Drug: Estrogel (Estradiol gel)

Indication: Treatment of moderate-to-severe vasomotor symptoms

Medical Reviewer: Phill Price, M.D. (HFD-580)

Documents reviewed: Volumes 1.1-1.4, 1.40-1.81,1.83;
Amendment dated 10/28/99; Amendment dated 1/21/00;
Amendment dated 4/21/00

Background

The applicant is seeking approval for the treatment of moderate-to-severe vasomotor symptoms (MSVS). The results from two U.S. multicenter, randomized clinical studies have been submitted to support the indication. The duration of the studies was 12 weeks. One study was placebo-controlled and the other used an active, open-label control. Both studies enrolled post-menopausal women who had an average of approximately 11 to 12 MSVS per day at baseline.

In Study CV141-001, 221 women were randomized to 1.25g Estrogel, 2.5g Estrogel, or Placebo; all treatment groups were double-blinded. In Study CV141-002, 361 women were randomized to 0.625g Estrogel, 1.2g Estrogel, 2.5g Estrogel or open-label Climara patch 12.5 cm². In this study, only the Estrogel arms were blinded; the patch was open-label.

The primary endpoint in each study, as specified in the final analysis plan, was change from baseline in the frequency of moderate-to-severe hot flushes from baseline through 12 weeks. To evaluate this endpoint, the protocol specified the use of analysis of covariance (ANCOVA) and Dunnett's test for multiple comparisons. The ANCOVA models included the following factors: treatment, center, treatment-by-center interaction; the number of moderate-to-severe flushes at baseline was the covariate. Small centers (less than five patients) were combined into "meta-centers". The applicant performed additional ANCOVA models that, in addition to the factors above, included a treatment-by-baseline interaction term.

Both studies had an interim analysis of the primary efficacy endpoint. A Lan and DeMets implementation of the O'Brien/Fleming approach was used. The interim analysis was conducted at $\alpha = 0.001$ and the final analysis was done at $\alpha = 0.049$.

Descriptive analyses of the primary efficacy outcome were performed for subgroups defined by age (<55, >55) and by race (white, non-white).

Reviewer's comments

Primary endpoint

The study protocol and data analysis plan did not identify those weeks that would be the basis for a regulatory decision. Generally, FDA has approved NDAs for this indication if a change from baseline, as compared with placebo, is statistically significant at Week 4, Week 8, and Week 12. These weeks are the focus of my review of the efficacy data contained in this NDA.

Study CV141-001

Pre-specified analysis of covariance (treatment, center, treatment-by-center interaction, and frequency of MSVS at baseline as a covariate)

The key results from the pre-specified analyses of the ITT¹ efficacy sample are:

- The Week 4 comparison of Estrogel versus Placebo was not statistically significant for 1.25g (p=.124) nor for 2.5g (p=.121)².
- For the comparisons of 1.25g Estrogel versus Placebo, the ANCOVA showed statistical significance at Week 6, Week 9 and subsequent weeks.
- For the comparisons of 2.5g Estrogel versus Placebo, the ANCOVA showed statistical significance at Week 5 and subsequent weeks.

The results of these ITT analyses with LOCF imply that 2.5g Estrogel is effective at Week 5 and subsequent weeks. The results imply 1.25g Estrogel is effective at Week 9 and subsequent weeks. These results, therefore, do not meet the guideline of statistical significance at Week 4.

The results of the **visit-wise**^{3,4} comparisons give a somewhat different picture for the 1.25g treatment group. With a single exception at Week 9, there are no statistically significant differences between 1.25g Estrogel and placebo. These discrepant findings are of interest because, generally, subjects benefiting from study drug tend to remain on study; while those who are not benefiting tend to discontinue treatment. Visit-wise comparisons, therefore, are often more significant than analyses of ITT with LOCF. This was not the case for 1.25g. The results for 2.5g, however, mimic those for the ITT analyses.

¹ The ITT analyses used LOCF for missing data.

² Source: Addendum Table 1.1, Amendment dated 4/21/00

³ The visit-wise analyses use only the data available at a specific visit; no imputation is done.

⁴ Source: Addendum Table 1.2; Amendment dated 4/21/00

Assessment of whether the magnitude of treatment effect (active gel versus placebo gel) depends on frequency of MSVS at baseline

ANCOVA models assume a linear relationship between the primary endpoint and the covariate. An additional assumption is the slopes of these lines are the same for each of the three treatment groups. For a particular timepoint (e.g., Week 4), conceptually, this means

- plotting change from baseline versus baseline for 1.25g, 2.5g, and Placebo
- fitting a separate line through these data for each of the treatment arms
- determining whether these three lines are parallel

The applicant assessed the underlying assumption of parallel slopes across treatment groups by examining treatment-by-baseline interactions for each of the pre-specified ANCOVA models. Adding a treatment-by-baseline interaction term to the ANCOVA models and testing whether the interaction term was statistically significant achieved the goal of assessing the existence of parallel slopes.

The NDA identified those weeks for which the assumption of equal slopes was rejected, as assessed by the treatment-by-baseline interaction term. The assumption did not hold for Week 5 and subsequent weeks. For those weeks, the comparisons of treatment means versus placebo means were done at the 25th, 50th, and 75th percentiles of the frequency of MSVS at baseline. These percentiles correspond to 7.93, 9.39, and 11.50 MSVS, respectively.

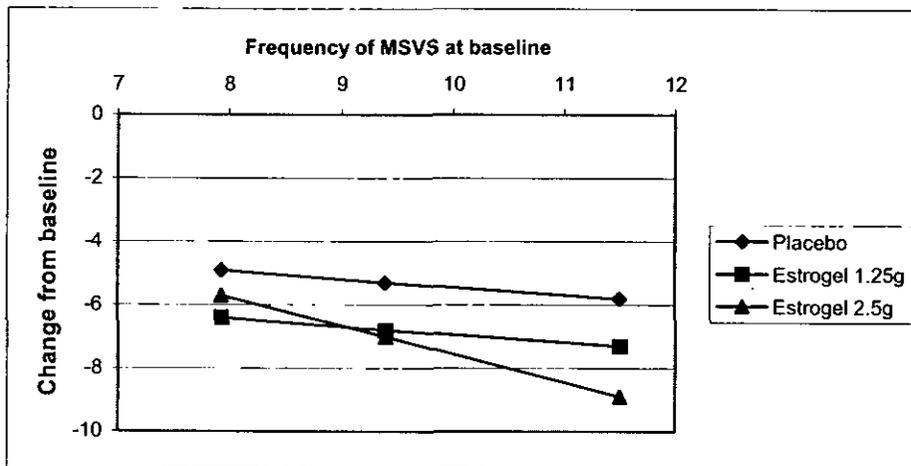
To explore the nature of the treatment-by-baseline interactions, I prefer looking at estimates of the coefficients for the terms in the ANCOVA models. These estimates, however, were not included in the submission. Instead, I constructed graphs for Weeks 8 and 12 using the estimated change from baseline in MSVS at the 25th, 50th, and 75th percentiles of the frequency of MSVS at baseline⁵; see next page.

These graphs show that the lines for placebo and Estrogel 1.25g appear approximately parallel. This suggests the difference between Estrogel 1.25g and Placebo was essentially constant over the range between the 25th and 75th percentiles of the frequency of MSVS at baseline. This does not appear to be the case for Estrogel 2.5g. Instead, the difference between Estrogel 2.5g and Placebo is smallest at the 25th percentile and greatest at the 75th percentile.

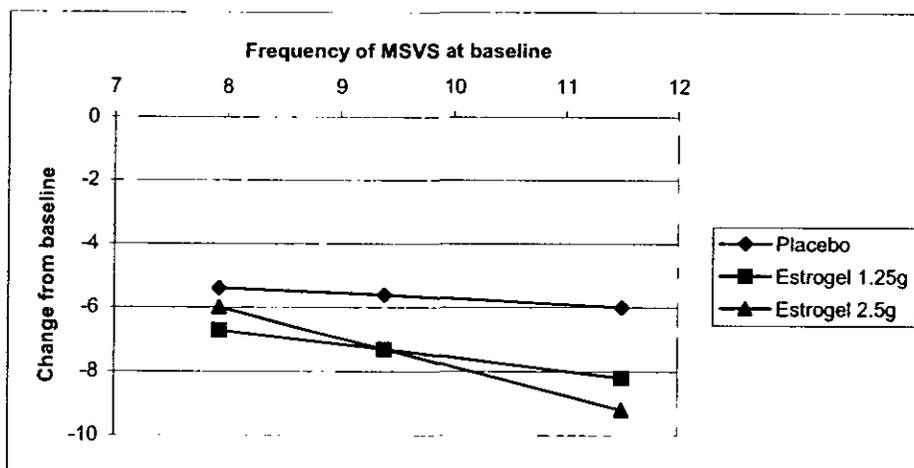
Because of the statistically significant interaction, the NDA presents results of comparisons between Estrogel and Placebo at the 25th, 50th, and 75th percentiles; see Volume 1.41, Table 10.4.2.1. Focusing on Weeks 4, 8, and 12, the only statistically significant differences between Estrogel 1.25g and Placebo were at Week 4 for the 25th percentile. For Estrogel 2.5g, statistically significant differences were observed at Weeks 8 and 12 at the 75th percentile, and at Week 8 for the 50th percentile.

⁵ Week 4 is not included because a treatment by baseline interaction was not present at this timepoint.

Study CV141-001: Change from baseline in frequency of MSVS versus 25th, 50th, and 75th percentiles⁶ of frequency of MSVS at baseline, by treatment group; ITT sample population⁷.



Results at Week 8



Results at Week 12

⁶ The percentiles are 7.93 (25th), 9.39 (50th), and 11.50 (75th).

⁷ Source of data: Table 10.4.2.1, Volume 1.41

Subgroup analyses

The NDA notes women who were non-white tended to have greater changes from baseline than women who were white⁸. The difference between the Estrogel arms and Placebo, however, appeared to be about the same for white women and non-white women. The treatment effects for women <55 were about the same as those for women >55.

Summary

The results suggest that 2.5g Estrogel is statistically different from Placebo starting at Week 5, provided that the treatment-by-baseline interaction is ignored. When accounting for the interaction, the results are less compelling. The magnitude of the treatment effect depended on the number of MSVS at baseline. Only women with a relatively large number of MSVS at baseline appeared to benefit from 2.5g Estrogel relative to Placebo. There did not appear to be an important difference for women who had fewer numbers of MSVS at baseline.

The results for 1.25g Estrogel were inconsistent. The analysis of the ITT sample implied efficacy at Weeks 9 and greater. The visit-wise analyses, however, implied essentially no statistically significant differences from placebo.

Study CV141-002

Estrogel .625g was the comparator arm in all analyses. Using the pre-specified ANCOVA models, the comparisons with 1.25g and 2.5g were not statistically significant different at any time point⁹, nor did there appear to be a dose response among the three doses of Estrogel. The active control treatment group demonstrated statistically significant greater decreases from baseline than did the Estrogel .625g group for Weeks 2 through Week 6

These results were supported by the analyses evaluated at the 25th, 50th, and 75th percentiles of MSVS at baseline¹⁰, which were done at Week 2, and at Weeks 5 through 9. For Weeks 5 and 6, significant differences between the active control and .625g were seen for the 75th percentile only.

Reviewer's conclusions

This submission contains two studies with conflicting results. The placebo-controlled study provided weak evidence for the 2.5g dose of Estrogel, and little for Estrogel 1.25g. The study with the active control did not yield any statistically significant comparisons among the Estrogel arms, nor were any of the arms statistically better than the active control. The results of the two controlled clinical studies, therefore, are not persuasive in supporting the efficacy of Estrogel 1.25g or Estrogel 2.5g.

⁸ Source: Tables 10.4.5.2.1 and 10.4.5.2.2, Volume 1.41

⁹ Source: Table 10.4.2.2.1, Volume 1.54; Addendum Table 1.1.1, Amendment dated 4/21/00

¹⁰ For this study, the 25th, 50th, and 75th percentiles are 8.8, 9.9 and 12.6, respectively.

Lisa A. Kammerman, Ph.D.
Team Leader, DB II

concur: E. Nevius, Ph.D. (HFD-715)

cc:

Archival NDA# 21-166

HFD-580/PPrice, SSlaughter, DSpell-LeSane, SAllen

HFD-715/ENevius, LKammerman