STATISTICAL REVIEW AND EVALUATION
CLINICAL STUDIES

		Date:
NDA#:	**20-908**	
Applicant:	**Novo Nordisk**	
Name of Drug:	**Vagifem 25 ug. (estradiol vaginal tablets)**	
Indication:	**Vaginal Atrophy**	MAR 25 1999
Documents Reviewed:	**6-1-98 NDA; 1-26-99 Electronic data; 3-17-99 Facsimile**	
Statistical Reviewer:	**Barbara Elashoff, M.S.**	
Medical Input:	**Ridgely Bennett, M.D.**	

Summary

The sponsor submitted one double-blind placebo-controlled trial (Study 9/USA) and one open-label active comparator (Premarin Cream) trial (Study 5/CAN) as primary evidence of efficacy of Vagifem 25 ug.

The symptom score results of the placebo-controlled trial demonstrated a statistically significant treatment difference in favor of Vagifem 25 ug. There was, however, a treatment-by-center interaction in the study (p=0.056) which necessitated further examination of individual center results. This reviewer concluded that the interaction was caused by the extremely positive results of Center #8 (1.22 units) in relationship to the remaining 7 centers (range: -0.37-0.73 units). The irritation symptom scores were much higher at Center #8 than the other centers. Further, the relationships between the scores of dryness, soreness and irritation were notably stronger at Center #8 relative to the other centers. Therefore, calculation of an overall average treatment effect including this center is problematic. An analysis of the remaining 7 centers showed a marginally statistically significant overall difference of about 0.26-0.28 units on a 0-3 scale between treatment groups (which was not internally consistent within various subgroups of patients). An additional problem with the quality of the data in this study was the unusual irritation scores at Center #7. All the patients in Center #7 scored "none" for irritation at every visit, including baseline. The statistically significantly different results across centers, unusual results within two centers, and the lack of robustness of the results of the primary efficacy variable, all contribute to the statistical uncertainty of the estimate of the treatment effect. This reviewer feels that a second placebo-controlled trial is necessary to validate the findings of efficacy of Vagifem and to better quantify the treatment effect.

The symptom score results of the open-label study 5/CAN should be viewed with caution due to the following limitations:

- Lack of double-blinding (particularly important for a subjective endpoint);
- Lack of a well-defined endpoint;
- Lack of an active control whose efficacy has been proven in a double-blind placebo-controlled trial.

Study 5/CAN was open-label and had a subjective endpoint (symptom scores). The effect of multiple biases (expectation of different effects, differential dropout rates) make it impossible to determine the accuracy of the estimate. In addition, the quality of the symptom score data in the open-label trial was compromised because neither the definitions of the symptoms nor the definitions of the severity ratings were defined to the patients. It is unclear what symptoms the patients assessed and even whether all the patients assessed the same symptoms. Finally, the treatment effect of the active comparator in the open-label trial has never been estimated and has never been demonstrated to be statistically significant in a placebo-controlled trial submitted to the FDA, to this reviewer's knowledge[1]. A large (statistically significant) placebo response was demonstrated in the sponsor's placebo controlled trial (Study 9/USA); the mean symptom score of the placebo group at the end of the study was statistically significantly different from the baseline mean. Therefore, the large changes from baseline in the open-label study in both treatment groups could be due to large placebo effects.

Premarin (at the highest approved dose) appeared to have a greater estrogenic effect than Vagifem. The vaginal cytology results of the open-label study 5/CAN demonstrated a statistically significant difference between Vagifem 25 ug and Premarin Cream at Weeks 2,12 and 24. At Week 2, after adjusting for baseline differences, the mean Maturation Value (MV) of the Vagifem 25 ug group increased by 33 units from baseline while the Premarin Cream

[1] Premarin Cream was permitted under the Drug Efficacy Study Implementation (DESI) review program. The reviewing medical officer of the current application, Dr. Bennett, performed a literature search (from 1965-1999) to determine if Premarin Cream had ever been compared to placebo to estimate the benefit of Premarin on symptoms associated with vaginal atrophy. No placebo-controlled trials were found.

group increased by 41 units, on average (p=0.0014). The baseline adjusted differences between treatment groups at Weeks 12 and 24 were statistically significant as well (Week 12: 4.7 units; p=0.0173; Week 24: 13.4 units; p=0.0001). Between 10% and 40% of the patients in each treatment group had missing data either at baseline or at the post-baseline visit. The percent of missing data was different between treatment groups, making these results difficult to interpret.

This reviewer does not feel that the open-label active-control trial results of Study 5/CAN for the subjective endpoint (symptom scores) provide reliable statistical evidence of the efficacy of Vagifem. While the results of the secondary objective endpoint, cytology, are not confirmatory, the data suggest a greater estrogenic effect of Premarin over Vagifem.

In conclusion, the sponsor has submitted one study (9/USA) that is supportive evidence of the efficacy of Vagifem. It was not possible to estimate the overall treatment effect due to the treatment-by-center interaction. This reviewer feels that a second placebo-controlled trial is necessary to validate the findings of efficacy and to quantify the treatment effect.

Vagifem appeared to have a low estrogenic effect and hence could be safer than Premarin and other estrogen products. It is a clinical decision whether the degree of benefit of Vagifem (unknown at this time) warrants exposing patients to the potential risk of hyperplasia. This reviewer feels that an estimate of the benefit of this and other estrogen products would be useful to the consumer and physician. The proposed Vagifem label does not state any estimate of benefit at this time. A second placebo-controlled trial would be needed in order to identify a reliable estimate of the benefit of Vagifem to put in the label.

Table of Contents

1 Introduction

The clinical development program for Vagifem included eight studies: one open-label active-comparator study in Canada, two double-blind placebo-controlled studies (one in the United States and one in Denmark), and five other studies (2 long-term open-label safety extension studies, 2 clinical pharmacology studies and 1 long-term dose regimen trial). An additional 11 foreign studies were included in the NDA as "secondary source data". Only the safety data of these studies was included in the NDA because these studies were not conducted under the Sponsor's IND and information was limited or incomplete for these trials. Only the open-label active comparator study performed in Canada and the double-blind placebo-controlled study performed in the US will be summarized in this review. Data from the Denmark study had two problems: demographic information was not collected and changes in data on the source documents were made without being dated and initialed. Therefore, this study is not summarized in this review.

Table 1 below summarizes the designs of the US placebo-controlled study and the Canadian open-label comparative study.

Table 1: Study Designs

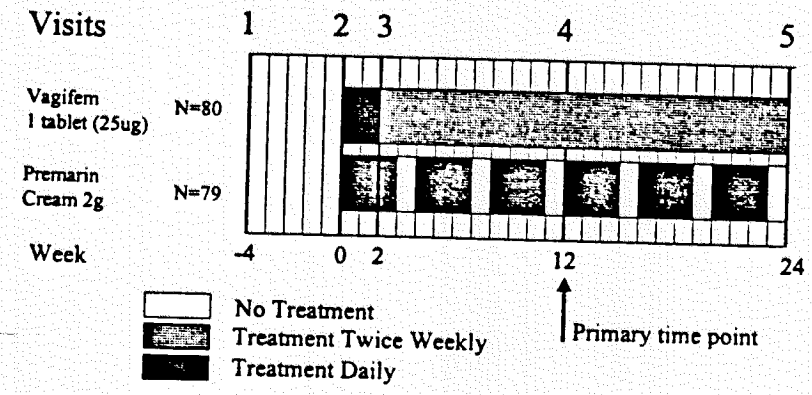
	5/CAN	9/USA
Dates Conducted	4/23/93 - 1/29/95	8/18/94 – 11/2/95
N Randomized N Treated N ≥ 1 post-baseline measure	160 159 149	230 230 224
Blinded?	Open-label	Double-blind
Active Treatment Group(s)	**Vagifem 25 ug** daily for 2 weeks, thereafter twice weekly with ≥ 3 days between each application	**Estradiol 10 ug & Vagifem 25 ug** daily for 2 weeks, thereafter twice weekly with ≥ 3 days between each application
Control Group	**Premarin Cream 2 g** daily for 21 days, withheld for 7 days, repeat cyclically (3 weeks on, 1 week off) through 24 weeks	**Placebo** daily for 2 weeks, thereafter twice weekly with ≥ 3 days between each application
Run-in	4 weeks	4 weeks
Treatment Phase	24 weeks	12 weeks
Primary Endpoint	Change from baseline of Vaginal Symptoms (Dryness, Soreness & Irritation) at **12 weeks**	Change from baseline of Vaginal Symptoms (Dryness, Soreness & Irritation) at **7 weeks**
Procedure for assessing symptoms	Investigator interviewed patients; each symptom was assigned a severity rating *after* the questioning	Definitions of symptoms and differences between none, mild, moderate and severe were explained to the patient by the investigator; the patient evaluated each symptom by checking the CRF where applicable
Definitions of symptoms	None	Dryness: no lubrication or secretions noted on perineum or after wiping; for sexually active patient, loss of lubrication with coitus. Soreness: throbbing, pressure, fullness sensation in vagina that causes discomfort. Irritation: Sand paper type feeling, uncomfortable with clothing or undergarments touching the perineum.
Definitions of Severity Ratings	None; Instructions were provided to the investigators as to how to assign severity scores based on the answers to the interview questions.	None: no sensation of symptom; Mild: feels symptom episodically; symptom does not interfere with daily activities; Moderate: sensation of symptom most of the time, symptom does not interfere with activities of daily living; Severe: sensation of symptom most of the time, symptom interferes with activities of daily living.
Entrance Criteria	1. Patients ≥ 40 and ≤ 80 yrs; 2. Patients with estrogen deficiency-derived atrophic vaginitis; ≥ 2 symptoms as moderate or severe; 3. Intact uteri; 4. At least 1 year of amenorrhea; 5. FSH≥40 IU/L at Visit 1; 6. ≥30% parabasal cells; 7. E2≤110 pmol/L (29pg/mL) at Visit 1	1. Patients ≥ 45; 2. Moderate or severe vaginal dryness and vaginal soreness; 3. Patients with intact urteri must have endometrial thicknesses of ≤5 mm; 4. At least 1 year of amenorrhea in nonhysterectomized patients; 5. ≤5% superficial cells as assessed by vaginal cytology evaluation; 6. FSH≥40 mIU/L; 7. Serum estradiol ≤25 pg/mL
Protocol Amendments	Entrance criteria change: 7/21/93 Patients must classify two or more symptoms (vaginal dryness, soreness, irritation, or dyspareunia) as moderate or severe.	Entrance criteria change: 12/2/94 FSH no longer needed to be measured therefore FSH did not need to be ≥40 mIU/L; Entrance criteria change: 2/10/95 Serum estradiol ≤20 pg/mL Additional visit (Week 4): 3/31/95

2 Study 5/CAN

2.1 *Study Design and Objectives*

The Canadian study (5/CAN) was a 24-week open-label study comparing Vagifem 25 ug to Premarin Cream 2 g (highest approved dose). One-hundred fifty nine (159) patients were randomized to either Vagifem 25 ug, given daily for 2 weeks then twice weekly for the remaining 22 weeks, or Premarin Cream 2g, given in 3-week cycles: (3 weeks on, 1 week off).

Figure 1: Study Design 5/CAN



2.2 *Primary Endpoint*

The primary endpoint analyzed in the study report was the change from baseline in average score of vaginal symptoms (dryness, soreness, and irritation) at 12 weeks. The investigator interviewed the patient regarding vaginal symptoms. Vaginal symptoms were rated from 0 to 3 (none=0, mild=1, moderate=2, severe=3). Neither the investigators nor the patients were given any definitions of symptoms or differences with regard to severity. This is in contrast to the US study which included a very detailed standardized explanation of each symptom and each category of each symptom. The investigators were provided instructions as to how to assign severity scores based on the answers to the interview questions.

Reviewer Comment

The protocol originally stated that the primary endpoint was the difference in parabasal cell count at the end of 12 weeks. A sample size of 120 was calculated based on a 10% difference, standard deviation of 0.15 and an alpha-level of 0.05. An amendment to the protocol dated 7/21/93 changed the primary endpoint to four efficacy parameters: vaginal dryness, soreness, irritation and dyspareunia. A new sample size of 150 was calculated based on the percentage of patients who had mild or no symptoms. Finally, as a result of three meetings between FDA and the sponsor (3/6/96, 4/9/96 and pre-NDA 4/29/97), the primary variable was changed to a composite score (average) of three vaginal symptoms: dryness, soreness and irritation.[2] The primary time point was also changed from Week 24 to Week 12. The reason for this change was to demonstrate early relief of vaginal symptoms. When the change in primary efficacy variable was made, the primary analysis was changed from a Mantel-Haenszel Chi-Square test on the percent of patients with mild or no symptoms in each treatment group to a 95% confidence interval around the observed mean difference in composite score. Based on the given sample size (80 per group) and the revised objective, there was to have been a 90% probability that the confidence interval would lie within (-0.4, 0.4), assuming that the expected difference was zero.

---

[2] The sponsor stated in the NDA that dyspareunia was not evaluable for all patients, (Volume 68, page 24). Therefore, it is not summarized in this review.

The primary endpoint was flawed and therefore, the results of vaginal cytology will be used to assess comparative efficacy in this review. The reasons the primary endpoint was flawed are outlined below.

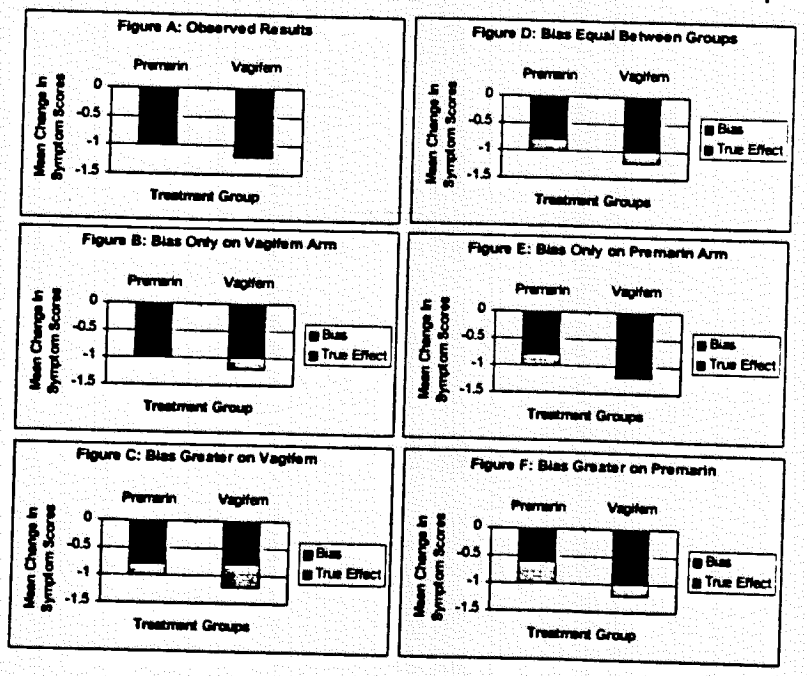
Symptom Scores Not Clearly Defined:

When questioning patients about subjective endpoints such as symptom scores, it is customary to provide all patients with the same definitions of each symptom (soreness, dryness and irritation) and with the same definitions of the severity of each category (i.e.: mild = bothersome but does not interfere with daily activities; moderate = bothersome and interferes with some daily activities; severe = interferes with all daily activities, no activity is possible). Without standardized definitions, patients and investigators may interpret the meaning of the symptoms and the severity of the symptoms differently. Some patients may not understand the distinction between soreness and irritation; some may interpret "loss of lubrication with coitus" as a dryness symptom while others may interpret it as a soreness symptom. It is unclear what symptoms the patients assessed in this study and whether or not all patients assessed the same symptoms. Interpretation of the results across patients is, at the least, difficult, and perhaps even impossible.

Symptom Scores Subject to Bias

In an open-label trial, results of subjective endpoints are subject to bias. It is possible that the investigator and/or the patient may believe that the new, investigational drug is more effective than the old drug. Suppose, for example, that Vagifem did not improve symptom scores as much as Premarin. If the Vagifem patients think the drug is as good as, or even better than, the already-approved drug, Premarin, the mean symptom scores of the Vagifem patients may decrease at least as much as the scores of the patients on Premarin, regardless of the relative effectiveness of the two drugs. Figure 2 below illustrates the observed results and several situations of bias that may have occurred. Only five situations are illustrated below, but there are numerous possibilities. The bias could be as large as the entire treatment effect and as small as zero. The potential for bias increases the uncertainty in the estimate of the difference between treatment groups. The 95% confidence interval around the estimate of the treatment difference was calculated based on the assumption of an unbiased estimate. Since we cannot calculate the confidence interval including the extra uncertainty of bias, the results are uninterpretable.

Figure 2: Increased Uncertainty of Results Due to Bias Introduced from Open-Label Design



The objective endpoint, maturation value of vaginal cells (vaginal cytology) was pre-specified in the protocol as a secondary endpoint and was assessed by a laboratory technician who was blinded to the treatment groups. This

endpoint was clearly defined and was not subject to bias. Therefore, the results of vaginal cytology will be used to assess comparative efficacy in this review.

One can argue that cytology is not necessarily related to symptom scores and that a blinded trial cannot be performed in this situation because a placebo cream might interfere with the absorption of the Vagifem tablet. This may be true, but the difficulties of doing a blinded comparative trial do not justify using an open-label trial as evidence of efficacy when a placebo-controlled trial can be performed.

2.3 Demographics

Demographics across treatment groups were similar (see Table 2, below).

Table 2: Study 5/CAN Demographics

	Premarin Cream	Vagifem (25 ug)
Total n treated	79	80
Age (yrs)		
Mean (SD)	57.2 (7.8)	57.3 (7.1)
Min-Max	42-85	45-76
N (%) Caucasian	77 (97)	77 (96)
Time since last menses (yrs)		
Mean (SD)	7.6 (7.2)	7.9 (7.0)
Min-Max	1-31	1-37

2.4 Results

2.4.1 Compliance

A record of all study medication dispensed and returned was kept by the investigator. Compliance was measured using the "Accountability of Unused Study Drug Record". Each investigator sent drug accountability records to the sponsor. These records contained: the date and quantity of medication received, dates, and quantity dispensed, patient identification numbers, amount of drug used and returned. According to the sponsor, these "forms were found to be unreliable in some cases," page 49, Volume 84. Nevertheless, the sponsor calculated the number of patients who used at least 80% of study medication and the number who were at each visit, as recorded on these forms.

Table 3: Study 5/CAN Number of Compliant Patients* At Each Visit Out of Total Number of Patients At Each Visit

Week	Vagifem	Premarin	Total
2	78 / 78 (100)	70 / 76 (92)	148 / 154 (96)
12	72 / 76 (95)	59 / 65 (91)	131 / 141 (93)
24	71 / 74 (96)	54 /56 (96)	125 / 130 (96)

* Compliant: patients used at least 80% of study medication

Reviewer Comment

The percentages were slightly lower in the Premarin treatment group at Weeks 2 and 12.

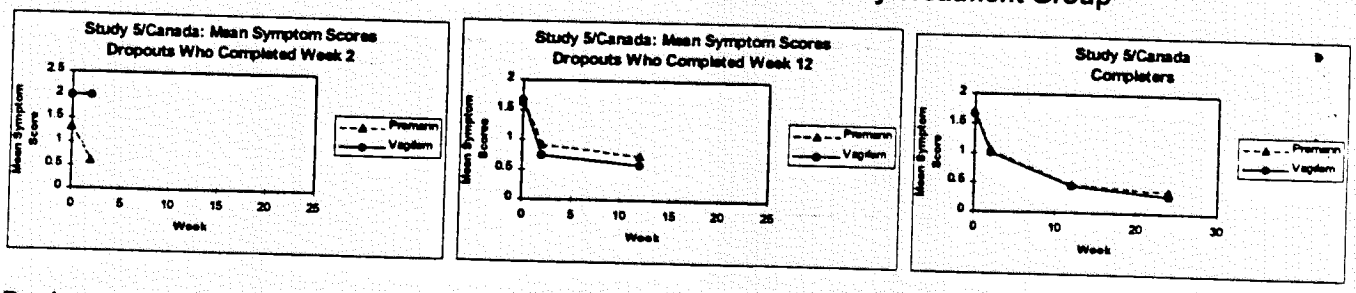
2.4.2 Patient Discontinuations

About 10% of the patients (n=16) discontinued the study before Week 12, the primary time point. There was a large difference between percentages discontinuing in the Premarin group (18%) and the Vagifem group (2.5%). Eight out of the 14 dropouts on the Premarin arm dropped out at the baseline visit immediately after randomization. With the exception of the 1 Vagifem patient who discontinued after the Week 2 visit, the pattern of symptom score over time was similar between the dropouts and the completers (see Figure 3 below).

Table 4: Study 5/CAN Patient Discontinuations

TRT	Last Visit	N	Age Mean	Age Std	Age Median	# of Years Since Amenorr. Mean	# of Years Since Amenorr. Std Dev	# of Years Since Amenorr. Median	Baseline Avge Sx Score Mean	Baseline Avge Sx Score Std Dev	Baseline Avge Sx Score Median	Change From Baseline Mean	Change From Baseline Std Dev	Change From Baseline Median
Premarin	0	8	57	10	54	8	9	5	1.6	1	1.3	0.0	0	0.0
Vagifem	0	1	63	-	63	12	-	12	1.3	-	1.3	0.0	-	0.0
Premarin	2	6	56	4	57	6	6	5	1.3	1	1.5	-0.7	0	-0.7
Vagifem	2	1	45	-	45	1	-	1	2.0	-	2.0	0.0	-	0.0
Premarin	12	9	58	9	57	9	7	7	1.6	1	1.7	-0.9	1	-1.0
Vagifem	12	4	53	5	53	4	3	4	1.7	0	1.8	-1.1	0	-1.0
Premarin	24	56	58	8	57	7	7	5	1.7	1	2.0	-1.2	1	-1.0
Vagifem	24	74	58	7	57	8	7	5	1.7	1	1.7	-1.2	1	-1.0

Figure 3: Mean Composite Symptom Scores By Treatment Group



Reviewer Comment

The large percentage of patients dropping out of the Premarin group at baseline could be due to the fact that the study treatment was not blinded. Differential discontinuation due to treatment assignment negates one of the purposes of randomization in a controlled clinical trial. In both the ITT and Modified ITT analysis, the change from baseline scores of the eight Premarin patients who discontinued the trial after 1 visit are equal to zero (using last value carried forward). Assuming some of the patients would have had at least some relief of symptoms (as the placebo response appears to be strong in the one placebo-controlled trial), the discontinuation of these eight patients (and the additional six that discontinued at week 2) introduced bias into the analysis that favored Vagifem.

The sponsor calculated only 54 and 72 patients in Premarin and Vagifem treatment group, respectively, at the end of the study. The electronic dataset had values for 56 and 74 patients at Week 24.

2.4.3 Primary Analysis

The sponsor defined two different populations, the intent-to-treat (ITT) and the modified ITT. The ITT patients included all randomized patients (n=160). The modified ITT included all randomized patients who received treatment and had both a baseline and post-baseline measurement (n=151). The primary analysis was pre-specified as the modified ITT analysis. The sponsor calculated 149 patients for the modified ITT population (as compared to the 151 patients the reviewer calculated from the electronic datasets.) The sponsor pooled Centers #3 and 6 in the analysis. The mean difference and 95% confidence interval was 0.06 units (-0.17, 0.30). The reviewer's results (using 151 patients), and including all centers as separate centers, were similar, see Table 6 below.

Table 5: Study 5/CAN Reviewer's Summary Statistics

	Vagifem 25 ug	Premarin Cream
ITT	N= 80	N=80
Week 0	1.68 ± 0.70	1.63 ± 0.73
Week 12	0.51 ± 0.62	0.63 ± 0.65
Change	-1.16 ± 0.79	-1.00 ± 0.80
Modified ITT	N= 79	N=72
Week 0	1.68 ± 0.70	1.67 ± 0.70
Week 12	0.50 ± 0.62	0.56 ± 0.56
Change	-1.18 ± 0.79	-1.11 ± 0.77

Table 6: Study 5/CAN Reviewer's Analyses*

	Vagifem 25 ug	Premarin	Treatment Difference		
ITT	N=80	N=80	Mean	95% CI	p-value
Model with Center & Trt	-1.15	-0.99	0.16	(-0.08, 0.40)	0.2026
Model with Center, Trt & Baseline	-1.17	-1.05	0.13	(-0.05, 0.31)	0.1660
Modified ITT	N=79	N=72			
Model with Center & Trt	-1.16	-1.09	0.06	(-0.18, 0.30)	0.5960
Model with Center, Trt & Baseline	-1.19	-1.15	0.05	(-0.13, 0.22)	0.5982

*None of the centers are pooled in these analyses. Results of analyses combining Centers #3 and 6 are similar.

The results did not demonstrate a difference between Vagifem and Premarin. The lower limit of the 95% confidence interval around the difference in change from baseline symptom score, adjusted for center, is at most -0.18 units (using the modified ITT population). The 95% confidence intervals calculated using both the ITT and the modified ITT analysis lie within the (-0.4, 0.4) limits that the sponsor proposed would indicate "equivalence" under the revised objective. Since the large and differential dropout rate biased the results of a LOCF analysis in favor of Vagifem, analyses were performed using the dropout patients with composite symptom scores equal to zero at Week 7.

Table 7: Study 5/CAN Reviewer's Summary Statistics (assuming patients who dropped out responded)

	Vagifem 25 ug	Premarin Cream
ITT	N=80	N=80
Week 0	1.68 ± 0.70	1.63 ± 0.73
Week 12	0.50 ± 0.61	0.50 ± 0.56
Change	-1.18 ± 0.78	-1.13 ± 0.78

Table 8: Study 5/CAN Reviewer's Analyses (assuming patients who dropped out responded)

	Vagifem 25 ug	Premarin	Treatment Difference		
ITT	N=80	N=80	Mean	95% CI	p-value
Model with Center & Trt	-1.15	-1.11	0.037	(-0.20, 0.27)	0.7585
Model with Center, Trt & Baseline	-1.17	-1.17	0.0085	(-0.16, 0.17)	0.9187

Reviewer Comment

The results of these analyses suggest no difference between Vagifem and Premarin Cream (see Table 8, above). However, these results are subject to bias (because the patients knew which treatment they were taking), the direction and magnitude of which is unknown and cannot be estimated.

Subsets of Patients Before and After Protocol Amendment

The protocol amendment, dated 7/21/93, changed the entrance criteria. The entrance criteria prior to this amendment did not include any restrictions regarding symptom scores. After the amendment, patients must have classified two or more symptoms (vaginal dryness, soreness, irritation, or dyspareunia) as moderate or severe in order to be eligible for the study. The mean changes from baseline for the two different groups (patients who entered study before amendment versus after) are presented for each treatment group in Table 9 and Table 10 below.

Table 9: Before Amendment

	Vagifem 25 ug	Premarin Cream
ITT	N=6	N=5
Week 0	1.28 ± 0.74	0.53 ± 0.56
Week 12	0.17 ± 0.18	0.13 ± 0.18
Change	-1.11 ± 0.62	-0.40 ± 0.55
Modified ITT	N=6	N=4
Week 0		0.67 ± 0.54
Week 12	(same as above)	0.17 ± 0.19
Change		-0.50 ± 0.58

Table 10: After Amendment

	Vagifem 25 ug	Premarin Cream
ITT	N=74	N=75
Week 0	1.71 ± 0.69	1.71 ± 0.68
Week 12	0.54 ± 0.63	0.67 ± 0.68
Change	-1.17 ± 0.81	-1.04 ± 0.80
Modified ITT	N=73	N=68
Week 0	1.71 ± 0.69	1.73 ± 0.66
Week 12	0.53 ± 0.63	0.58 ± 0.57
Change	-1.18 ± 0.80	-1.15 ± 0.77

Reviewer Comment
Since the number of patients who entered the study before the amendment was approved is small, it is difficult to make comparisons of treatment effect between groups.

2.4.4 Secondary Efficacy Variable

As stated above in Section 2.2, page 5, the objective endpoint, Maturation Value, assessed by a blinded laboratory technician, will be used as primary basis of comparability of efficacy of the two treatment groups, since the primary endpoint, symptom scores, was subjective, possibly biased, and not defined in a standardized manner to patients.

The percentages of parabasal, intermediate, and superficial cells were assessed at Visits 1 (-4 weeks), 3 (2 weeks), 4 (12 weeks) and 5 (24 weeks). Cytology was not assessed at Visit 2 (baseline), therefore, the vaginal cytology smear at Visit 1 (the screening visit) is used as the baseline value. A Maturation Value (MV) was also calculated based on the following equation:

MV=0 x parabasal cells % + 0.5 x intermediate cells % + 1.0 x superficial cells %.

Larger values of MV indicate greater estrogen effect. Comparison of the treatment effect on change from baseline in these percentages and MV were made using a protocol-specified analysis of variance model. The sponsor's electronic dataset had values for patients at Week 0. It is unclear whether these values were carried forward from the screening visit (Week –4). (Only symptom scores from Week 0-12 were included in the dataset). Therefore, this reviewer used the cytology data identified as Week 0 (in the electronic dataset) as the baseline value. The changes in MV, adjusting for baseline, were statistically significantly greater in the Premarin Cream treatment group at every treatment period visit (Weeks 2, 12 and 24), see Table 12.

Table 11: Summary Statistics of Maturation Value of Vaginal Cytology

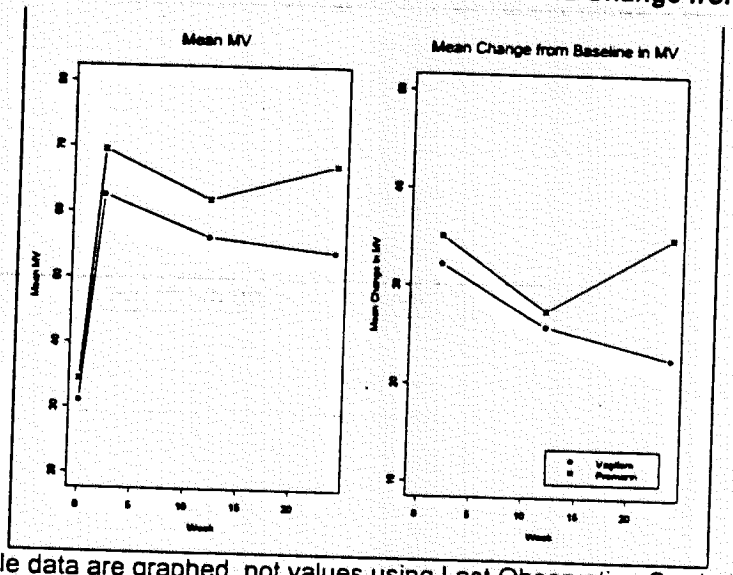
	Vagifem 25 ug	Premarin Cream
Week 0	N=74: 31.0 ± 23.9	N=60: 34.5 ± 21.9
Week 2	N=72: 62.6 ± 12.6	N=59: 69.6 ± 16.0
Week 12	N=64: 56.3 ± 9.2	N=49: 62.0 ± 15.0
Week 24	N=62: 54.1 ± 10.9	N=48: 67.4 ± 13.8

Table 12: Mean Change from Baseline in MV Cytology at Weeks 12 and 24

	Vagifem 25 ug	Premarin	Treatment Difference		
Change at Week 2	N=72	N=59	Mean	95% CI	p-value
Model with Center & Trt	38.0	42.1	4.1	(-4.9, 13.2)	0.3646
Model with Center, Trt & Baseline	33.1	40.7	7.6	(3.0, 12.2)	0.0014
Change at Week 12	N=64	N=49			
Model with Center & Trt	29.7	31.7	2.0	(-7.5, 11.5)	0.6772
Model with Center, Trt & Baseline	24.6	30.4	5.7	(1.0, 10.4)	0.0173
Change at Week 24	N= 62	N=48			
Model with Center & Trt	27.1	39.3	12.3	(2.4, 22.2)	0.0156
Model with Center, Trt & Baseline	22.5	35.9	13.4	(8.6, 18.2)	0.0001

The difference between treatment group is greatest at Week 24 (13.3 units, p=0.0001). Note that the differences in mean MV scores between treatment groups were about two to three times the differences between the Vagifem 25 and placebo treatment groups in the placebo controlled trial (see Table 31, page 21).

Figure 4: Study 5/CAN Graphs of Maturation Values and Change from Baseline*



*All available data are graphed, not values using Last Observation Carried Forward.

The sponsor also analyzed the mean changes in percent of superficial cells between the screening visit and Weeks 2, 12 and 24. The changes were statistically significantly greater in the Premarin Cream group (p<0.0001), see Table 13 below.

Table 13: Sponsor's Analysis of Mean Percent of Superficial Cells

	Vagifem			Premarin			Diff	p-value
	N	Mean	Change	N	Mean	Change		
Week –4	74	3.51	-	60	2.92	-		
Week 2	72	25.69	22.29	59	38.98	36.10	13.81	<0.001
Week 12	64	13.52	10.16	49	26.63	23.88	13.72	<0.001
Week 24	62	12.34	9.52	48	35.63	32.81	23.09	<0.001

Reviewer Comment

Due to the unreliable results of the primary endpoint (subjective symptom scores in an open-label trial), this reviewer utilized the results of this objective endpoint as the primary basis of comparable efficacy. However, bias is introduced into this analysis as well, due to the large amount of missing data. At Week 2, the amount of missing data was smallest (Vagifem: 10%; Premarin: 26%). The Week 2 data suggest a greater estrogenic effect of Premarin over Vagifem. The amount of missing data increases to 23% and 39% for the Vagifem and Premarin groups, respectively at Week 12. The direction and magnitude of this bias introduced by this is unknown. Although the results are not conclusive due to the large amount of missing data, the results appear to suggest a greater estrogenic effect of Premarin over Vagifem.

2.4.5 Adverse Events

Fifty-eight percent (46/80) of Vagifem patients reported at least one adverse event compared with 70% (55/79) of Premarin patients. The most frequently reported adverse events were reproductive system symptoms, such as bleeding, breast and perineal pain, reported by 27 (34%) Premarin patients and 7 (9%) Vagifem patients.

The endometrial biopsy results are presented in Table 14, below. Only one Vagifem patient had an endometrial biopsy result other than atrophic endometrium or insufficient tissue, i.e., proliferative endometrium. After treatment with

Premarin, 13 patients had endometrial biopsy results other than atrophic endometrium or insufficient tissue. Four patients had weakly proliferative results, seven had proliferative endometrium, one had simple hyperplasia without atypia, and one had complex hyperplasia without atypia.

Table 14: Study 5.CAN Endometrial Biopsy Results

	Vagifem 25 ug	Premarin 2 g
N Randomized	80	79
Patient with uterus (non-hysterectomized)	80	79
Total Biopsies	49	49
Insufficient Tissue	14 (28%)	21 (42%)
Atrophic Endometrium	34 (68%)	15 (30%)
Weakly Proliferative	0 (0%)	4 (8%)
Proliferative	1 (2%)	7 (14%)
Simple Hyperplasia	0 (0%)	1 (2%)
Complex Hyperplasia	0 (0%)	1 (2%)

Reviewer Comment

The rates of adverse events may have been subject to bias due to the fact that each patient knew 1) that she was taking a drug; and 2) which drug she was taking (approved or unapproved). Assuming the Premarin patients had access to the Premarin label, the adverse events listed in the label may have influenced their perception of pain and itching (among other things). It is possible that the differential effects between the drugs are due, in part, to bias introduced by patients knowing which drug they were taking.

The endometrial biopsy results indicate that the risk for hyperplasia is greater on Premarin 2 g than on Vagifem 25 ug.

2.5 Conclusions

Interpretation of the efficacy results of Study 5/CAN is problematic. Ideally, to use Study 5/CAN as evidence of efficacy, Premarin Cream, the active comparator, would have to have been compared to placebo (in a previous study). The efficacy of Premarin Cream for the indication of symptom scores of Vaginal Atrophy has never been established in a placebo-controlled trial, to this reviewer's knowledge.

Further, the results of the trial do not provide conclusive statistical evidence that the two treatments provide an equivalent effect due to the bias introduced by the open-label nature of the study and the differential percentages of missing data. The one objective measure, vaginal cytology smear, although subject to bias due to missing data, appeared to indicate that Premarin had a greater estrogenic effect than Vagifem as early as Week 2 which was maintained through Week 24. However, the clinical relevance of this endpoint is in question.

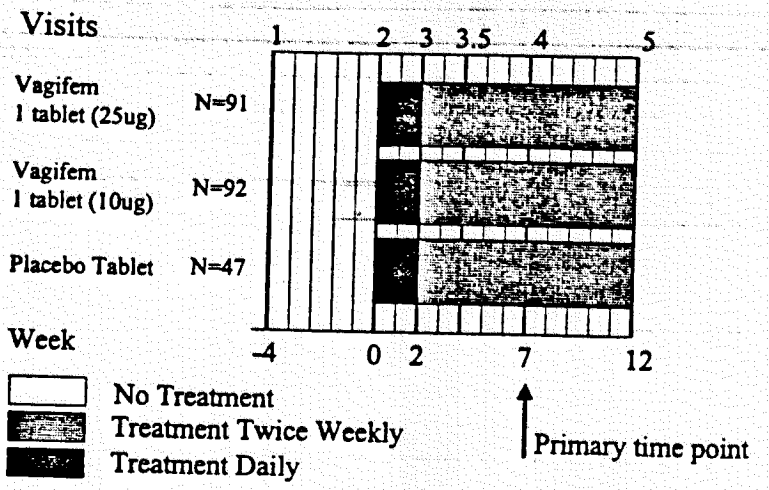
The safety results of Study 5/CAN indicate a lower risk of hyperplasia associated with Vagifem than Premarin.

3 Study 9/USA

3.1 Design and Objectives

Study 9/USA, a large (n=230) randomized double-blind, placebo controlled parallel study comparing Vagifem 25 ug and 10 ug to Placebo in post-menopausal women with symptoms of vaginal atrophy (soreness, dryness and irritation). Randomization was 2:2:1 (Vagifem 25 ug, E2: 10 ug, and placebo). The primary endpoint was the change from baseline in vaginal symptoms at seven weeks.

Figure 5: Study Design 9/USA



3.2 *Primary Endpoint*

The primary endpoint was the change from baseline in average score of vaginal symptoms (dryness, soreness, and irritation) at 7 weeks. Vaginal symptoms were rated from 0 to 3 (none=0, mild=1, moderate=2, severe=3). Patients were given detailed definitions of both the symptoms and of the severity of the different categories (see Table 1).

Reviewer Comment

The protocol originally stated that the primary endpoint was the difference in proportion of patients with "none" or "mild" symptoms (dryness, soreness, irritation, dyspareunia and vaginal discharge) at Week 12. As a result of three meetings between FDA and the sponsor (3/6/96, 4/9/96 and pre-NDA 4/29/96), the primary variable was changed to a composite score (average) of three of the original five vaginal symptoms: dryness, soreness and irritation.[3] The primary time point was also changed from Week 12 to Week 7. The reason for this change was to demonstrate early relief of vaginal symptoms. When the change in primary efficacy variable was made, the primary analysis was changed from an Mantel-Haenszel ordered categories chi-square test on the percent of patients with mild or no symptoms in each treatment group to an analysis of covariance of the differences in the change from baseline of the composite score, including center, treatment and center-by-treatment interaction effects. Based on the given sample sizes (91 Vagifem 25 ug, 47 placebo) and the standard deviation of 0.85, there was 50% power to detect a difference of 0.3 at the 0.05 two-tailed level. There was 74% and 90% power to detect differences of 0.4 and 0.5 units, respectively.

3.3 *Demographics*

The demographics were similar across treatment groups. Note that about half the patients had had hysterectomies in this study. This is in contrast to the open-label comparator study in which all patients had intact uteri. For the patients with intact uteri, and those who had had hysterectomies after menopause, the number of years since last menses was greater in this study than in the open-label comparator study.

Table 15: Study 9/USA Demographics

	Placebo	E2: 10 ug	Vagifem (25 ug)
Total n randomized	47	92	91
Age (yrs)			
Mean (SD)	57.6 (4.8)	57.7 (6.5)	58.3 (7.4)
Min-Max	50-70	46-79	46-78
N (%) Caucasian	41 (87.2)	83 (90.2)	88 (96.7)
Time since last menses (yrs)			
Mean (SD)	13.6 (8.1)	13.5 (7.8)	14.8 (9.6)
Min-Max	1-33	1-34	1-40
N (%) Had Hysterectomy	23 (48.9)	44 (47.8)	42 (46.2)

[3] The sponsor stated in the NDA that discharge was reported as either mild or not present for >90% of the patients. Dyspareunia was not evaluable for all patients, (Volume 68, page 24). Therefore, these symptoms are not summarized in this review.

3.4 Results

3.4.1 Compliance

A record of all study medication dispensed and returned was kept by the investigator. Compliance was measured using the "Accountability of Unused Study Drug Record". Each investigator sent drug accountability records to the sponsor. The sponsor did not include results of compliance records in the NDA.

3.4.2 Patient Discontinuations

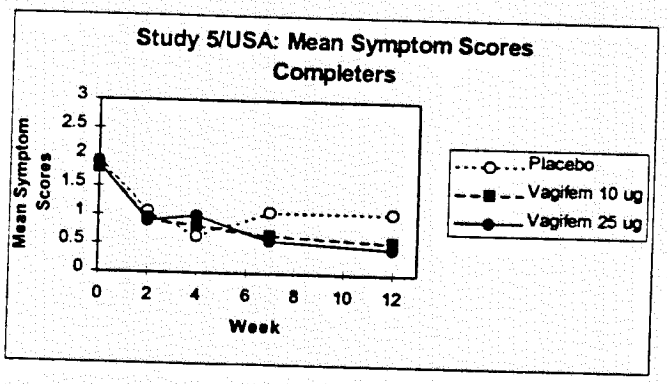
About 8% of patients (n=19) discontinued before Week 7, the primary time point. The differences between treatment groups were small (Placebo: 4%; 10 ug estradiol: 11%; Vagifem 25 ug: 8%).

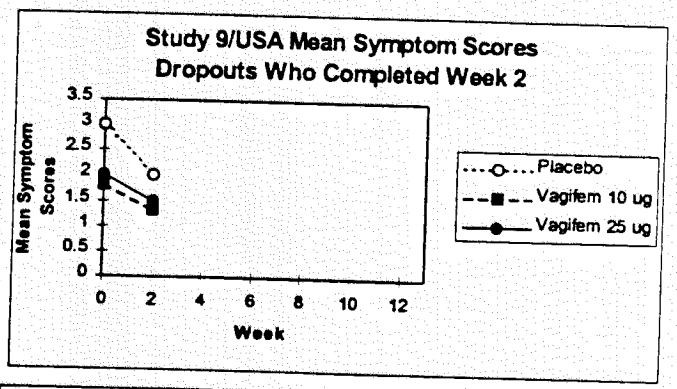
The demographics, baseline characteristics, response and patterns of symptom score over time were fairly similar between the dropouts and the completers. Last observation carried forward was used in the graphs below and in the sponsor's and reviewer's analyses.

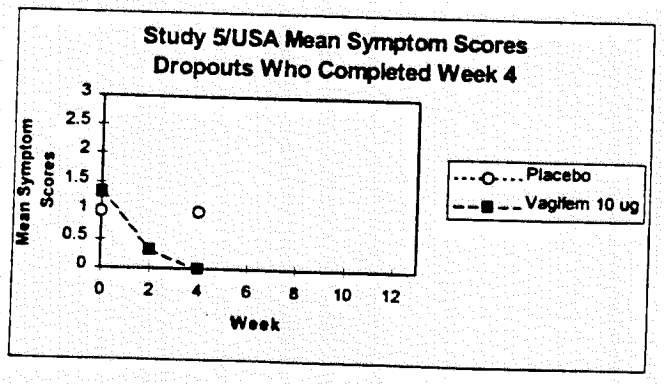
Table 16: Study 9/USA Patient Discontinuations

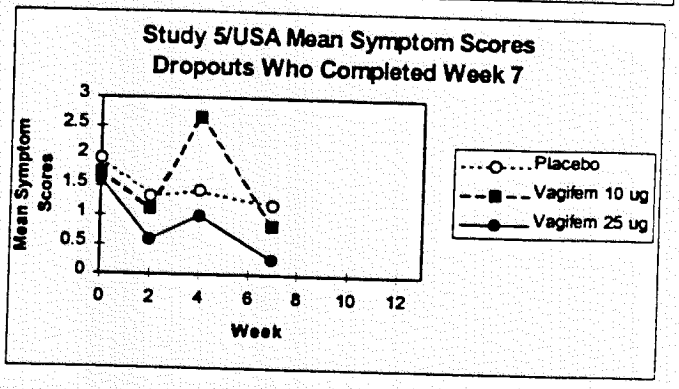
TRT	Last Visit	N	Age			# of Years Since Amenorrhea			Baseline Avge Composite Sx Score			Change From Baseline		
			Mean	Std Dev	Median	Mean	Std Dev	Median	Mean	Std Dev	Median	Mean	Std Dev	Median
Vagifem 10 ug	0	5	61	7	60	10	9	9	1.9	1	2	0.0	0	0
Vagifem 25 ug	0	1	59	-	59	15	-	15	1.7	-	2	0.0	-	0
Placebo	2	1	62	-	62	23	-	23	3.0	-	3	-1.0	-	-1
Vagifem 10 ug	2	4	58	10	58	13	7	11	1.8	0	2	-0.5	1	-1
Vagifem 25 ug	2	6	50	2	51	9	4	11	2.0	1	2	-0.5	1	0
Placebo	4	1	50	-	50	17	-	17	1.0	-	1	0.0	-	0
Vagifem 10 ug	4	1	51	-	51	28	-	28	1.3	-	1	-1.3	-	-1
Placebo	7	7	58	4	58	16	10	16	2.0	0	2	-0.8	1	-1
Vagifem 10 ug	7	10	60	7	61	15	6	14	1.7	0	2	-0.9	1	-1
Vagifem 25 ug	7	5	58	5	59	14	9	12	1.6	0	2	-1.3	1	-1
Placebo	12	38	58	5	56	13	8	12	1.9	1	2	-0.9	1	-1
Vagifem 10 ug	12	72	57	6	56	13	8	13	1.8	1	2	-1.1	1	-1
Vagifem 25 ug	12	79	59	7	58	15	10	13	1.9	1	2	-1.3	1	-1

Figure 6: Mean Composite Symptom Scores By Treatment Group









3.4.3 Primary Analysis

Summary statistics of the primary variable and the individual components (irritation, dryness and soreness) are graphed in the Appendix A, pages 23-24.

Due to the outcomes of several different meetings with the FDA (3/6/96, 4/9/96 and pre-NDA 4/29/96), the sponsor performed two different models for the two populations, ITT and modified ITT.

ITT: Placebo and Vagifem 25 ug only; ANOVA with center, treatment and center-by-treatment interaction

Modified ITT: All 3 treatment groups; ANOVA with center and treatment

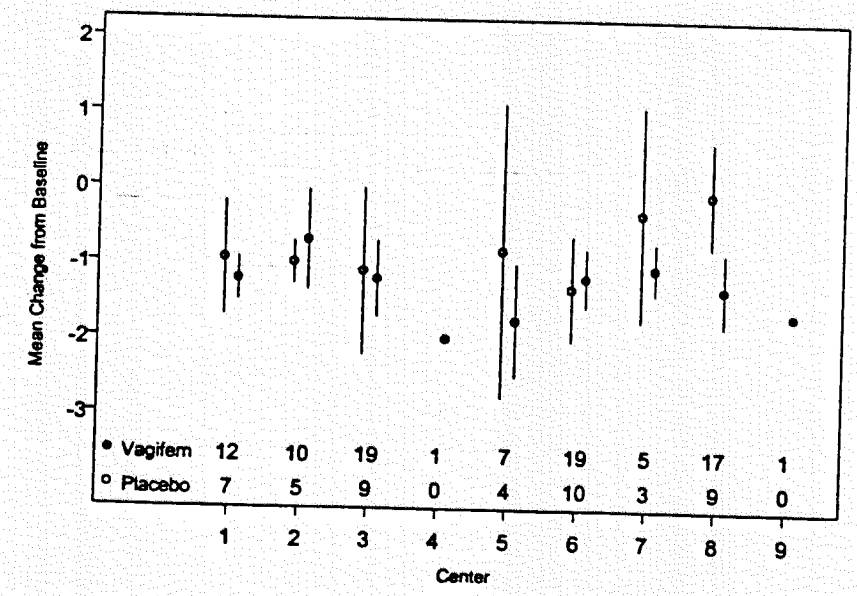
Only 1 patient from Vagifem 25 ug was excluded from the modified ITT analyses (patient had only 4 days on treatment) and 5 patients from the Vagifem 10 ug group.

This reviewer performed the same analyses for the two populations. First, analyses with the treatment-by-center interaction effect were performed in order to estimate the interaction effect. Analyses without treatment-by-center interaction were performed to estimate the overall treatment effect without the interaction effect. Only the placebo and Vagifem 25 ug treatment groups were included in the models. (Results from models with Vagifem 10 ug and placebo are presented below as well.)

Since the populations differed by only 1 patient, the results of the model with the interaction term are presented for one population only (ITT). The results were almost identical for the modified ITT population.

There was a center-by-treatment interaction (p=0.0560) in this study. The sponsor attributed it to two centers with superior placebo groups (Centers #2 and 6), see Figure 7 below. Center #2 had a placebo patient (ID #45) who discontinued early (Day 14) whom the sponsor considered an "outlier" (symptoms increased from 1.33 to 2.67 units). The sponsor attributed the superior placebo group in Center #6 to a baseline imbalance in mean composite symptom scores (Placebo =1.23 and Vagifem=1.82). After controlling for baseline, the mean difference in Center #6 does favor Vagifem, however, even when deleting patient #45 and accounting for baseline in the model, the center-by-treatment interaction is still evident (p=0.0887).

Figure 7: Study 9/USA Mean Changes from Baseline and 95% Confidence Intervals by Treatment Group (Vagifem 25 ug and Placebo) and Center



It appears as though Center #8 contributed most to the center-by-treatment interaction. Center #8, one of the largest centers (n=26), had an unusually small placebo response, relative to the other centers. There were no placebo

patients from Center #8 that "stood out" from the rest of the placebo patients in this center. It appeared as though the entire center had placebo patients that "stood out" from the rest of the placebo patients in the study, see Table 36 in Appendix A, page 23. Almost all the placebo patients in Centers 1-7 responded (31/38, 82%) whereas only 3/9 patients from Center #8 responded (33%). (Response rates in Centers 1-7 for Vagifem 10 ug = 92%; 25 ug = 89%; Center #8: Vagifem 10 ug = 76%; Vagifem 25 ug = 82%).

Further examination of the individual symptom score results in Center #8 revealed differences in irritation scores at baseline and subsequent visits, and differences in relationships between pairs of symptoms, see Appendix B, pages 26-27. The sponsor could not explain these differences.[4] Another unusual finding of the individual symptom scores was that all the patients in Center #7 recorded "none" for irritation at every visit, including baseline. This phenomenon was not seen in the other centers. Center #7 had the second largest treatment effect, after Center #8, see Table 17 below.

Table 17: Study 9/USA Results of Analyses with Center-by-Treatment Interaction Term

	ITT		ITT minus Center #8	
	SS III	p-value	SS III	p-value
Baseline	41.6	0.0001	33.8	0.0001
Treatment	4.6	0.0030	1.8	0.0609
Center	14.4	0.0009	2.2	0.7216
Trt*Center	6.4	0.0560	2.0	0.5494
	Trt Effect Estimate	p-value	Trt Effect Estimate	p-value
Center 1	0.29*	0.3934	0.29*	0.3877
Center 2	-0.37	0.3429	-0.37	0.3309
Center 3	0.30	0.3017	0.29	0.2953
Center 5	0.54	0.2256	0.55	0.2053
Center 6	0.27	0.3306	0.26	0.3347
Center 7	0.73	0.1597	0.73	0.1480
Center 8	1.22	0.0001		

*In the analysis of the modified ITT (excludes Vagifem 25 ug patient #31, change=0, number of days on study=4), the estimate of the treatment effect in Center #1 is 0.40.

Reviewer Comment

It is unlikely that the "none" irritation scores at Center #7 had any effect on the direction of treatment effect. However, the fact that all patients had the same irritation scores at every visit potentially lowered the variability of symptom scores and change from baseline scores at this center.

In order to estimate an overall treatment effect (not treatment effects at each individual center), a model without the center-by-treatment interaction effect must be performed. Since baseline symptom score explained most of the variability in change from baseline symptom score (and it was linearly related to change), models including baseline were performed in addition to the sponsor's models. The models (including Center #8) demonstrated a statistically significant treatment effect of about 0.39-0.44 units, depending on the model (with or without baseline) and the patient population (ITT or modified ITT). These overall treatment effects are not necessarily relevant because there is a statistically significant center-by-treatment effect in the population. That means that the treatment effect is statistically significantly different in one or more of the centers than in the other centers.

[4] The sponsor stated that, "Our overall interpretation is that the subject management and instruction at this site was consistent and careful; the investigator and staff are among the best in our experience." (Facsimile, March 17, 1999).

Table 18: Study 9/USA Summary Statistics of Primary Variable (Includes Center #8)

	Vagifem 25 ug	Vagifem 10 ug	Placebo
ITT	N=91	N=92	N=47
Week 0	1.85 ± 0.60	1.82 ± 0.61	1.93 ± 0.66
Week 7	0.63 ± 0.69	0.79 ± 0.83	1.08 ± 0.91
Change	-1.22 ± 0.86	-1.03 ± 0.85	-0.85 ± 1.05
Modified ITT	N=90	N=87	N=47
Week 0	1.86 ± 0.60	1.81 ± 0.61	
Week 7	0.62 ± 0.68	0.72 ± 0.79	(same as above)
Change	-1.24 ± 0.86	-1.10 ± 0.84	

Table 19: Mean Change from Baseline in Composite Symptom Score: Vagifem 25 ug*

	Vagifem 25 ug	Placebo	Treatment Difference		
ITT	N=91	N=47	Mean	95% CI	p-value
Model with Center & Trt	-1.22	-0.85	0.38	(0.04, 0.71)	0.0265
Model with Center, Trt & Baseline	-1.32	-0.87	0.45	(0.19, 0.71)	0.0007
Modified ITT	N=90	N=47			
Model with Center & Trt	-1.24	-0.85	0.39	(0.06, 0.72)	0.0207
Model with Center, Trt & Baseline	-1.34	-0.87	0.47	(0.22, 0.73)	0.0004

** Includes patients from Center #8.*

Table 20: Mean Change from Baseline in Composite Symptom Score: Vagifem 10 ug*

	Vagifem 10 ug	Placebo	Treatment Difference		
ITT	N=92	N=47	Mean	95% CI	p-value
Model with Center & Trt	-1.03	-0.84	0.19	(-0.14, 0.51)	0.2601
Model with Center, Trt & Baseline	-1.11	-0.84	0.27	(-0.01, 0.55)	0.0578
Modified ITT	N=87	N=47			
Model with Center & Trt	-1.08	-0.84	0.25	(-0.8, 0.58)	0.1390
Model with Center, Trt & Baseline	-1.17	-0.83	0.34	(0.06, 0.61)	0.0169

** Includes patients from Center #8.*

The analyses excluding the patients from Center #8 demonstrated a smaller treatment effect, 0.15-0.28 units, depending on the model and the patient population (see Table 22, below). A difference of 0.28 units (using the modified ITT population and controlling for baseline differences) was statistically significant at the 0.05 level (p=0.0482). The ITT analysis, which included a Vagifem patient who received treatment but was only in the study 4 days and did not have any post-baseline measurements (patient #31), yielded a treatment difference of 0.26 units which was not statistically significant (0.0710).

Table 21: Study 9/USA Summary Statistics of Primary Variable (Excludes Center #8)

	Vagifem 25 ug	Vagifem 10 ug	Placebo
ITT	N=74	N=75	N=38
Week 0	1.76 ± 0.57	1.72	1.86 ± 0.69
Week 7	0.56 ± 0.66	0.67	0.82 ± 0.75
Change	-1.20 ± 0.84	-1.04	-1.03 ± 1.00
Modified ITT	N=73	N=71	N=38
Week 0	1.76 ± 0.58	1.70	1.86 ± 0.69
Week 7	0.55 ± 0.65	0.60	0.82 ± 0.75
Change	-1.21 ± 0.84	-1.10	-1.03 ± 1.00

Table 22: Mean Change from Baseline in Composite Symptom Score: *Vagifem 25 ug

	Vagifem 25 ug	Placebo	Treatment Difference		
ITT	N=74	N=38	Mean	95% CI	p-value
Model with Center & Trt	-1.32	-1.17	0.15	(-0.22, 0.51)	0.4249
Model with Center, Trt & Baseline	-1.39	-1.13	0.26	(-0.02, 0.53)	0.0710
Modified ITT	N=73	N=38			
Model with Center & Trt	-1.33	-1.17	0.16	(-0.20, 0.52)	0.3683
Model with Center, Trt & Baseline	-1.41	-1.13	0.28	(0.002, 0.552)	0.0482

Excludes patients from Center #8.

Table 23: Mean Change from Baseline in Composite Symptom Score: *Vagifem 10 ug

	Vagifem 10 ug	Placebo	Treatment Difference		
ITT	N=75	N=38	Mean	95% CI	p-value
Model with Center & Trt	-1.01	-1.00	0.015	(-0.34, 0.37)	0.9316
Model with Center, Trt & Baseline	-1.14	-1.01	0.13	(-0.16, 0.43)	0.3777
Modified ITT	N=71	N=38			
Model with Center & Trt	-0.98	-1.06	0.076	(-0.28, 0.43)	0.6756
Model with Center, Trt & Baseline	-1.19	-0.98	0.21	(-0.77, 0.50)	0.1489

Excludes patients from Center #8.

Reviewer Comment

The sponsor's pre-specified model (controlling for center and treatment, not baseline) did not adequately explain the variability in change score. Baseline explained most of the variability (and was linearly related to change), therefore, it was included in the reviewer's post-hoc models. (In viewing post-hoc analyses, p-values should not be regarded as confirmatory.) The treatment effects were small in Centers #1-7, (< 0.30 units) and the lower limit of the 95% confidence intervals barely excluded zero (95% lower limit: 0.002). Therefore, the overall treatment effect at these seven centers could be as small as 0 and as large as 0.5 units on a four-point scale. Vagifem may improve symptoms of vaginal atrophy by at best, one half a unit. This translates into moving from the high end of one category to the middle of the same category (for example the high end of moderate: 2.9, to the middle of moderate: 2.4). At worst, Vagifem is no better than placebo at relieving symptoms. Results of analyses of the lower dose (10 ug estradiol) are suggestive of a slight dose response (10 ug: 0.21 difference; 25 ug: 0.28 difference). However, the increase of 0.07 units on a four-point scale is small.

With a small treatment effect seen in seven centers (0.26-0.28) using a post-hoc model accounting for baseline, and a large treatment effect (1.22 units) in a center with characteristics of symptom scores that were different from the other centers, a second study is desirable to validate the findings of efficacy and necessary to quantify the overall treatment effect. In the absence of a second placebo-controlled trial to validate the findings of this study, subset analyses were performed to determine if the treatment effect (within Centers #1-7) was internally consistent. The results of these analyses (presented below) demonstrated that the treatment effect was inconsistent across the various subgroups.

3.4.4 Subgroup Analyses

Hysterectomized vs. Non-hysterectomized

One-hundred nine, or 47% of the patients in this study had had hysterectomies. The baseline symptom scores of the hysterectomized patients appeared to be similar to those of the non-hysterectomized patients, with the placebo patients in both groups having higher baseline symptom scores than the Vagifem 25 ug patients, see Table 24 below. The treatment effect of the seven poolable centers appeared to be more pronounced in the patients who had had hysterectomies, see Table 25 below.

Table 24: Study 9/USA Summary Statistics of Primary Variable (Excludes Center #8)

	Vagifem 25 ug	Placebo
Hysterectomized	N=35	N=18
Week 0	1.78 ± 0.54	1.87 ± 0.65
Week 7	0.66 ± 0.64	0.98 ± 0.86
Change	-1.16 ± 0.85	-0.89 ± 1.10
Non-hysterectomized	N=39	N=19
Week 0	1.74 ± 0.61	1.88 ± 0.76
Week 7	0.48 ± 0.67	0.70 ± 0.64
Change	-1.26 ± 0.83	-1.18 ± 0.95

** Excludes patients in Center #8.*

Table 25: Mean Change from Baseline in Composite Symptom Score: *Vagifem 25 ug**

	Vagifem 25 ug	Placebo	Treatment Difference		
Hysterectomized ITT	N=35	N=18	Mean	95% CI	p-value
Model with Center & Trt	-1.14	-0.93	0.21	(-0.37, 0.80)	0.4623
Model with Center, Trt & Baseline	-1.25	-0.83	0.42	(-0.03, 0.88)	0.1437
Non-hysterectomized ITT	N=39	N=19			
Model with Center & Trt	-1.23	-1.16	0.074	(-0.44, 0.59)	0.7746
Model with Center, Trt & Baseline	-1.30	-1.14	0.16	(-0.22, 0.54)	0.3998

Number of Years of amenorrhea

The sponsor subset the patients into two groups: <10 years post-menopause and ≥10 years post-menopause. Although the overall average of baseline scores were similar between the under and over 10 years post-menopausal groups, the baseline values were lower among the Vagifem patients in the <10 years post-menopause group, see Table 26 below. The symptom scores of the placebo patients appeared to decrease more in this subset of patients, however, after adjusting for the baseline differences, the beneficial effect of placebo narrowed to a negligible difference of –0.06 units between treatment groups. The baseline values were similar across treatment groups for the ≥10 years post-menopause subset. The treatment effect of the seven poolable centers appeared to be restricted to the ≥10 years post-menopausal group, see Table 27 below. However, there are only 12 patients in the placebo group contributing to the treatment effect estimate in the patients <10 years post-menopausal.

Table 26: Study 9/USA Summary Statistics of Primary Variable (Excludes Center #8)

	Vagifem 25 ug	Placebo
<10 yrs post-menopause ITT	N=25	N=38
Week 0	1.68 ± 0.58	1.92 ± 0.84
Week 7	0.64 ± 0.85	0.75 ± 0.62
Change	-1.04 ± 0.97	-1.17 ± 1.12
≥10 yrs post-menopause ITT	N=49	N=26
Week 0	1.80 ± 0.57	1.85 ± 0.63
Week 7	0.52 ± 0.54	0.88 ± 0.82
Change	-1.31 ± 0.75	-0.97 ± 0.99

Table 27: Mean Change from Baseline in Composite Symptom Score: *Vagifem 25 ug**

	Vagifem 25 ug	Placebo	Treatment Difference		
<10 yrs post-menopause ITT	N=25	N=12	Mean	95% CI	p-value
Model with Center & Trt	-0.98	-1.15	-0.17	(-0.98, 0.63)	0.6628
Model with Center, Trt & Baseline	-1.08	-1.14	-0.057	(-0.67, 0.56)	0.8516
≥10 yrs post-menopause ITT	N=49	N=26			
Model with Center & Trt	-1.31	-0.98	0.32	(-0.11, 0.76)	0.1445
Model with Center, Trt & Baseline	-1.30	-1.00	0.30	(0.092, 0.747)	0.0129

** Excludes patients from Center #8.*

Baseline Values

This reviewer grouped the patients based on the median baseline value (1.67 units). The patients in the low baseline subset had similar baseline scores across treatment groups. The placebo patients in the high baseline subset had greater baseline scores than the Vagifem 25 ug group. The treatment effect of the seven poolable centers appeared to be more pronounced among the patients with the high baseline values.

Table 28: Study 9/USA Summary Statistics of Primary Variable (Excludes Center #8)

	Vagifem 25 ug	Placebo
Baseline <1.67 units	N=26	N=13
Week 0	1.22 ± 0.60	1.20 ± 0.56
Week 7	0.51 ± 0.71	0.72 ± 0.63
Change	-0.70 ± 0.83	-0.49 ± 0.70
Baseline ≥ 1.67 units	N=48	N=24
Week 0	2.06 ± 0.37	2.24 ± 0.54
Week 7	0.59 ± 0.85	0.90 ± 0.71
Change	-1.50 ± 0.93	-1.33 ± 0.96

Table 29: Mean Change from Baseline in Composite Symptom Score: *Vagifem 25 ug

	Vagifem 25 ug	Placebo	Treatment Difference		
Low Baseline	N=26	N=13	Mean	95% CI	p-value
Model with Center & Trt	-0.55	-0.53	0.02	(-0.55, 0.59)	0.9400
Model with Center, Trt & Baseline	-0.81	-0.65	0.16	(-0.39, 0.72)	0.5576
High Baseline	N=48	N=24			
Model with Center & Trt	-1.49	-1.37	0.12	(-0.28, 0.52)	0.5479
Model with Center, Trt & Baseline	-1.64	-1.32	0.32	(-0.03, 0.67)	0.0698

** Excludes patients from Center #8.*

Reviewer Comment

The treatment effect was not robust across subgroups of patients. The clinical relevance of this is unknown because there is no physiological reason the subgroups of hysterectomized vs. non-hysterectomized patients, or <10 years post-menopausal vs. ≥10 years post-menopausal patients should have different responses. Since these differences cannot be explained by physiology (or by baseline differences), the lack of a consistent treatment effect suggests a need for a second placebo-controlled trial to validate the overall results.

3.4.5 Secondary Analyses

The investigator examined and graded "vaginal health" according to definitions of the following symptoms: vaginal secretions, vaginal epithelial integrity, vaginal epithelial surface thickness, vaginal color and vaginal pH. These items were recorded on a 4-point scale (0-3) as symptom scores were and were to be analyzed similarly. The reviewing medical officer does not regard the examination of "vaginal health" as clinically relevant due, in part, to the difficulties in determining severity for some of the symptoms and the fact that pH may not necessarily be related to vaginal atrophy (due to infections). The difficulties determining severity for some of the symptoms include, for example, for vaginal secretions, mild = "superficial coating of secretions, difficulty with speculum insertion" and moderate = "scant and not covering entire vaginal vault, may need lubrication with speculum insertion to prevent pain". If the speculum needs lubrication to insert to prevent pain, then the investigator would undoubtedly have "difficulty with speculum insertion". Nevertheless, the vaginal health scores were analyzed to determine if the treatment-by-center interaction was evident for vaginal health. It was evident (p=0.0753), and exclusion of Center #8 eliminated the interaction (p=0.5668). This is another indication that Center #8 was different from the other centers and that an overall treatment effect cannot be reliably estimated from this study.

The percentages of parabasal, intermediate, and superficial cells were assessed at all visits. The value at Week 0 was used as the baseline value. A Maturation Value (MV) was calculated based on the following equation:

MV=0 x parabasal cells % + 0.5 x intermediate cells % + 1.0 x superficial cells %.

Larger values of MV indicate greater estrogen effect. Comparison of the treatment effect on change from baseline in these percentages and MV were made using a protocol-specified analysis of variance model. The changes in MV, adjusting for baseline, were statistically significantly greater in the Vagifem treatment groups at Weeks 2 and 7. The Week 4 visit was added by an amendment, therefore the results were not compared across treatment groups at Week 4. The results at Week 12 were not statistically significantly different between treatment groups. The results of Week 7 for are presented in Table 31 (Vagifem 25 ug) and Table 32 (Vagifem 10 ug) below.

There was no treatment-by-center interaction in the analyses of maturation value, therefore, Center #8 patients are included in the analyses of maturation value.

Table 30: Summary Statistics of Maturation Value of Vaginal Cytology

	Vagifem 25 ug	Vagifem 10 ug	Placebo
ITT			
Week 0	N=87: 47.4 ± 9.5	N=82: 47.6 ± 9.2	N=44: 46.1 ± 12.0
Week 7	N=87: 63.5 ± 12.7	N=82: 59.2 ± 9.6	N=44: 55.2 ± 10.3
Week 12	N=87: 59.7 ± 9.9	N=84: 58.7 ± 9.5	N=45: 53.8 ± 10.4
Modified ITT			
Week 0	N=86: 48.6 ± 9.3	N=77: 47.3 ± 9.3	
Week 7	N=86: 63.9 ± 12.2	N=77: 59.7 ± 9.6	(same as above)
Week 12	N=86: 60.1 ± 9.3	N=79: 59.2 ± 9.5	

Table 31: Reviewer's Analyses of Maturation Values of Vaginal Cytology: *Vagifem 25 ug**

Change at Week 7	Vagifem 25 ug	Placebo	Treatment Difference		
ITT	N=87	N=44	Mean	95% CI	p-value
Model with Center & Trt	15.7	8.7	-7.0	(-12.5, -1.6)	0.0119
Model with Center, Trt & Baseline	16.4	8.3	-8.1	(-12.4, -3.8)	0.0003
Modified ITT	N=86	N=44			
Model with Center & Trt	15.9	8.7	-7.2	(-12.7, -1.7)	0.0105
Model with Center, Trt & Baseline	16.6	8.1	-8.52	(-12.7, -4.3)	0.0001

** Includes patients in Center #8.*

Table 32: Reviewer's Analyses of Maturation Values of Vaginal Cytology: *Vagifem 10 ug**

Change at Week 7	Vagifem 10 ug	Placebo	Treatment Difference		
ITT	N=84	N=44	Mean	95% CI	p-value
Model with Center & Trt	11.9	8.8	-3.2	(-8.4, 2.1)	0.2335
Model with Center, Trt & Baseline	12.9	8.5	-4.4	-8.1, -0.76)	0.0184
Modified ITT	N=79	N=44			
Model with Center & Trt	12.5	8.7	-3.9	(-9.1, 1.4)	0.1518
Model with Center, Trt & Baseline	13.5	8.7	-4.9	(-8.6, -1.2)	0.0106

** Includes patients in Center #8.*

Reviewer Comment

The lack of treatment-by-center interaction in the analysis of vaginal cytology may be due to the fact that the results were read by a blinded laboratory reader. This would suggest that the interaction was mainly due to the Center #8 investigator (not the patients in Center #8) and that the symptom score results at Center #8 are unreliable.

3.4.6 Reviewer's Analyses of Percentages of Responders

In addition to looking at averages across treatment groups, it is useful to describe the results of the trial in terms of percentages of patients who "responded". Additional analyses were performed for this review, in order to determine the difference in the number and percent of patients who responded between treatment groups. Tables 33 and 34 below present the results of these analyses.

Table 33: Study 9/USA Percent of Responders*

	Placebo	Vagifem 25 ug	Difference in %	Chi-square p-value
# (%) of patients who had some decrease (> 0)	30/37 (81)	66/74 (89)	8%	0.239
# (%) of patients with ≥ 0.25 unit decrease	30/37 (81)	66/74 (89)	8%	0.239
# (%) of patients with ≥ 0.50 unit decrease	27/37 (73)	63/74 (85)	12%	0.123
# (%) of patients with ≥ 1 unit decrease	24/37 (65)	59/74 (80)	15%	0.089
# (%) of patients with ≥ 1.5 unit decrease	10/37 (27)	25/74 (34)	7%	0.470
# (%) of patients with ≥ 2 unit decrease	7/37 (19)	15/74 (20)	1%	0.866

* Excludes patients from Center #8.

The results in the table above are difficult to interpret due to the large baseline imbalance and the relationship between baseline and change from baseline. The same analyses are presented below stratified by baseline severity.

Table 34: Study 9/USA Percent of Responders Stratified by Baseline Severity*

	Low Baseline (< median 1.67)			High Baseline (≥ median 1.67)			M-H** p-value
	Placebo	Vagifem 25 ug	Diff in %	Placebo	Vagifem 25 ug	Diff in %	
# (%) of patients who had some decrease	8/13 (62)	21/26 (81)	19	22/24 (92)	45/48 (94)	2	0.2273
# (%) of patients with ≥ 0.25 unit decrease	8/13 (62)	21/26 (81)	19	22/24 (92)	45/48 (94)	2	0.2273
# (%) of patients with ≥ 0.50 unit decrease	7/13 (54)	19/26 (73)	19	20/24 (83)	44/48 (92)	9	0.1140
# (%) of patients with ≥ 1 unit decrease	7/13 (54)	17/26 (65)	11	17/24 (71)	42/48 (88)	17	0.0870
# (%) of patients with ≥ 1.5 unit decrease	0/13 (0)	0/26 (0)	0	10/24 (42)	25/48 (52)	10	0.4056
# (%) of patients with ≥ 2 unit decrease	0/13 (0)	0/26 (0)	0	7/24 (29)	15/48 (31)	2	0.8565

* Excludes patients from Center #8.
** M-H stands for Mantel-Haenszel.

Reviewer Comment

The results of each analysis above are correlated because as the unit decrease increases, the number of responders in each analysis is a subset of those from the previous analysis. Perhaps the strongest and most interpretable result of these analyses is the following: 88% of the Vagifem patients who were more severe at baseline changed at least 1 category of severity by Week 7 as compared to 71% of placebo patients. This is a 17 unit difference between treatment groups; however, this difference was not statistically significant. The effect was not as great among the milder subset of patients (11 unit difference). Further, there did not appear to be as much of a benefit above a 1 category decrease (differences in percent of patients experiencing a 2 category decrease was 0 in the mild patients and 3 in the more severe patients).

3.4.7 Adverse Events

One-hundred forty-two of 230 patients (62%) treated with study drug reported at least one adverse event: Placebo 55%, Estradiol 10 ug 61%, and Vagifem 25 ug 66%. Endometrial biopsy for one Vagifem patient (#179) showed simple hyperplasia without atypia at the end of the study (see Table 35, below). The diagnosis was confirmed by two pathologists. The event was resolved by a follow-up biopsy >2.5 months after the end of study visit which showed proliferative endometrium. Two Vagifem patients were diagnosed with carcinoma within a short period after the study had ended (87 days: basal cell carcinoma; 46 days: clear-cell carcinoma, right renal mass). One placebo patient died from an accidental drowning 19 days after discontinuing study drug, and another placebo patient was diagnosed with an enlarged right ovarian cyst 84 days after discontinuing study drug.

Table 35: Study 9/USA Endometrial Biopsy Results

	Vagifem 25 ug	Placebo
N Randomized	91	47
Patient with uterus (non-hysterectomized)	48	24
Total Biopsies	32	21
Insufficient Tissue	3 (9%)	3 (14%)
Atrophic Endometrium	27 (84%)	18 (86%)
Weakly Proliferative	0 (0%)	0 (0%)
Proliferative	1 (3%)	0 (0%)
Simple Hyperplasia	1 (3%)	0 (0%)
Complex Hyperplasia	0 (0%)	0 (0%)

3.5 Conclusions

The US placebo-controlled trial provides weak statistical evidence of the efficacy of Vagifem 25 ug. The treatment-by-center interaction in this study compromises the overall inferences of the trial. The average treatment effect of the seven poolable centers demonstrated a small (0.26-0.28 units) but inconsistent treatment effect. This treatment effect was more pronounced in patients with hysterectomies, patents who were at least 10 years post-menopausal, and patients with severe symptoms at baseline.

The benefit-to-risk ratio may be small, due to the apparent small beneficial effect and the recognized risk of hyperplasia for all estrogen products. However, the estimate of the treatment effect was not able to be conclusively estimated in Study 9/USA due to the statistically significantly different results across centers, unusual results within two of the centers, and the lack of robustness of the results of the primary efficacy variable.

4 Overall Conclusions

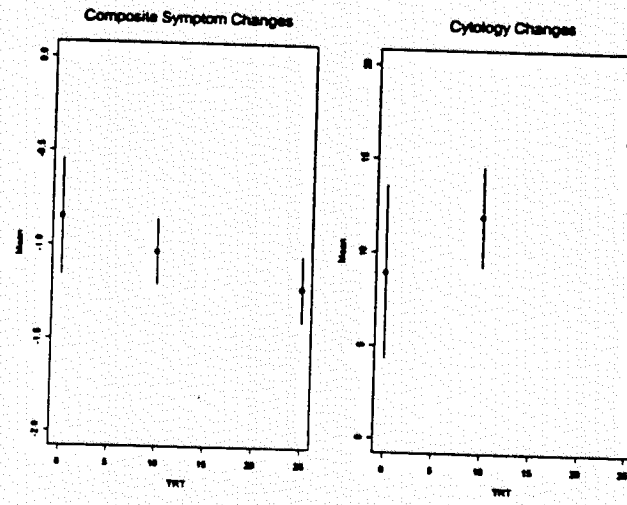
The sponsor submitted one double-blind placebo-controlled trial (Study 9/USA) and one open-label active comparator (Premarin Cream) trial (Study 5/CAN) as primary evidence of efficacy of Vagifem 25 ug. The placebo-controlled trial demonstrated weak evidence of the efficacy of Vagifem. The treatment effect was unable to be reliably quantified (due to the statistically significantly different results across centers, unusual results within two of the centers, and the lack of robustness of the results of the primary efficacy variable). The results of the open-label active-comparator trial, taken at face value, suggest no difference between Vagifem and Premarin Cream. Interpretation of the open-label trial is problematic due to the fact that the trial was open-label and thus subject to bias, and the fact that Premarin has never been determined to be effective in a placebo-controlled trial, to this reviewer's knowledge. This reviewer feels that a second placebo-controlled trial should be performed to validate the findings of efficacy in Study 9/USA and to provide a reliable estimate of the treatment effect.

Vagifem appeared to have a low estrogenic effect and hence could be safer than Premarin and other estrogen products. It is a clinical decision whether the degree of benefit of Vagifem (unknown at this time) warrants exposing patients to the potential risk of hyperplasia. In making this decision, it is this reviewer's opinion that an estimate of the benefit of this and other estrogen products would be useful to consumers and physicians in weighing the risks and benefits. The proposed Vagifem label does not state any estimate of benefit at this time. A second placebo-controlled trial would be needed in order to identify a reliable estimate of the benefit of Vagifem for the label.

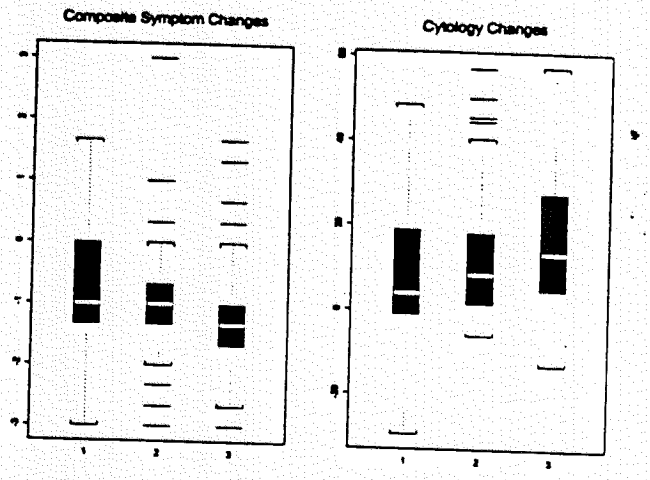
Appendix A: Summary Graphs

All Patients

Change from Baseline Mean & 95% CI



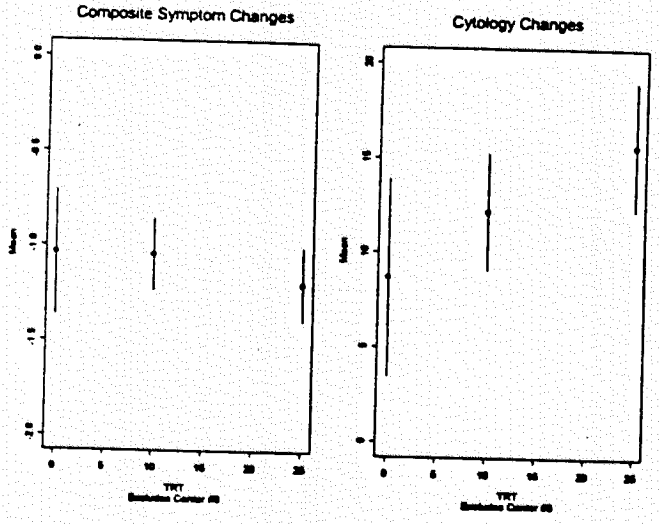
Boxplots



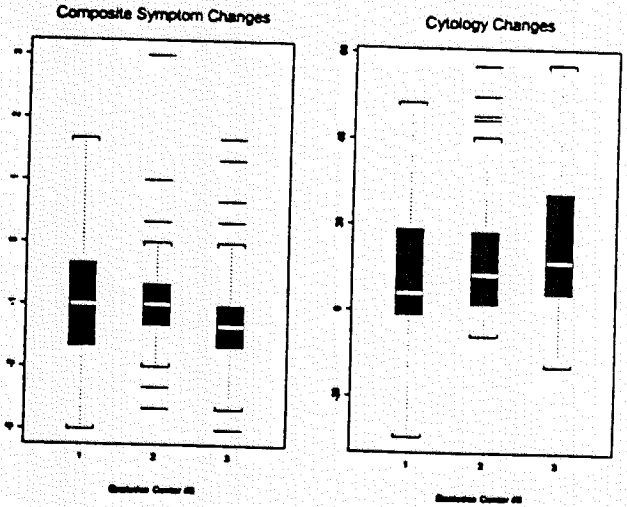
(1=Placebo; 2=Estradiol 10 ug; 3=Vagifem 25 ug)

Excludes Center #8

Change from Baseline Mean & 95% CI



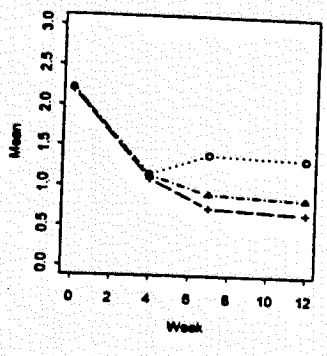
Boxplots



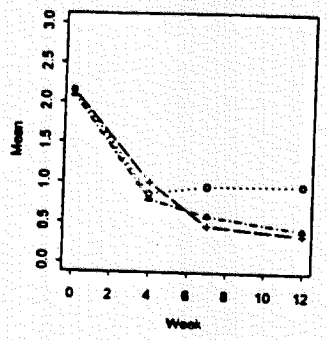
(1=Placebo; 2=Estradiol 10 ug; 3=Vagifem 25 ug)

All Patients

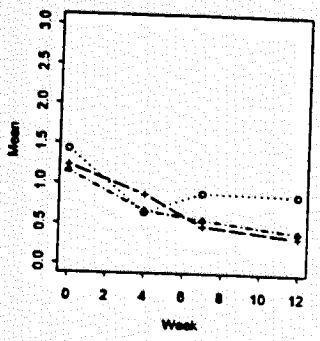
DRYNESS



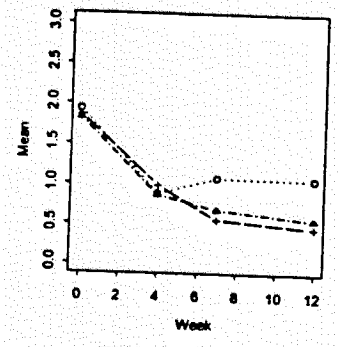
SORENESS



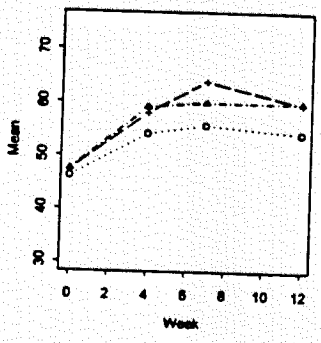
IRRITATION

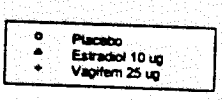


Composite Symptoms



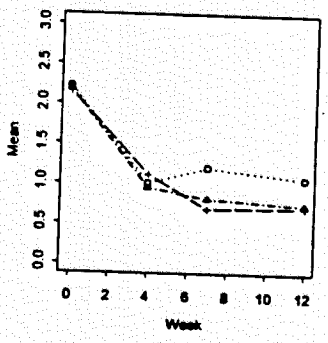
Cytology (MV Scores)



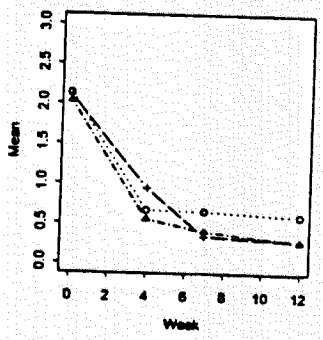


Excludes Center #8

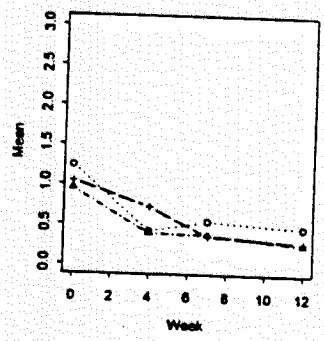
DRYNESS



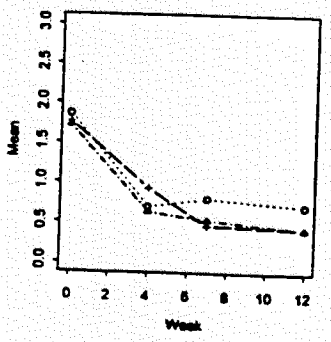
SORENESS



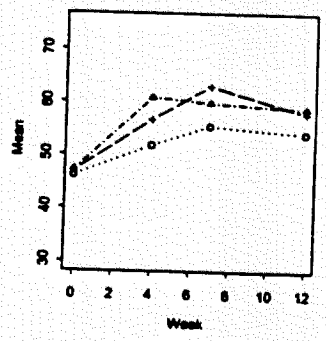
IRRITATION

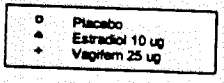


Composite Symptoms



Cytology (MV Scores)





Appendix B: Center Differences

Table 36: Placebo Patients Composite Symptom Scores (sorted by change from baseline)

Patient	Center	Age	Baseline	Week 7	Change
176	3	56	3	0	-3
7	6	64	3	0	-3
256	3	57	3	0.33	-2.67
78	5	68	3	0.67	-2.33
148	6	55	2.67	0.33	-2.33
251	3	55	2.67	0.67	-2
142	6	51	3	1	-2
214	8	54	3	1	-2
23	1	58	1.67	0	-1.67
52	3	56	2.33	0.67	-1.67
199	6	56	2	0.33	-1.67
17	1	57	1.33	0	-1.33
34	1	50	1.67	0.33	-1.33
228	1	63	1.33	0	-1.33
248	1	60	1.67	0.33	-1.33
135	2	53	1.33	0	-1.33
192	6	53	1.33	0	-1.33
201	6	56	1.67	0.33	-1.33
43	2	64	1.33	0.33	-1
50	2	51	2	1	-1
130	2	54	1.33	0.33	-1
182	3	62	3	2	-1
188	3	61	1.33	0.33	-1
74	5	60	2	1	-1
262	6	58	2.33	1.33	-1
101	7	59	1.33	0.33	-1
123	8	61	2	1	-1
37	2	56	1.67	1	-0.67
241	5	69	1.67	1	-0.67
15	6	63	2	1.33	-0.67
27	1	57	1.67	1.33	-0.33
172	3	62	1.33	1	-0.33
1	6	60	2.33	2	-0.33
217	8	55	2.33	2	-0.33
58	3	50	1	1	0
96	7	54	1.33	1.33	0
107	7	54	1.33	1.33	0
163	8	55	2	2	0
151	6	56	2	2.33	0.33
113	8	70	2.67	3	0.33
156	8	56	2	2.33	0.33
170	8	63	2	2.33	0.33
267	1	52	1.67	2.33	0.67
70	5	53	0	0.67	0.67
119	8	55	2	2.67	0.67
206	8	55	2	3	1
64	3	58	1.33	3	1.67

The following tables are counts of symptom scores in each of the categories of (0, 1, 2, 3) for each of the symptoms (Dryness, Soreness and Irritation). The center number is in the upper-left-hand corner of the table. Each patient is represented in these tables more than once (the number of symptom score recordings they had determines the number of times they are represented). Note the large correlation between symptom pairs in Center #8.

Dryness and Soreness

1

Soreness \ Dryness	0	1	2	3
0	36	34	2	0
1	5	14	2	0
2	0	2	36	1
3	0	1	0	0

Total: 133

2

Soreness \ Dryness	0	1	2	3
0	22	21	3	1
1	1	12	8	0
2	0	1	25	2
3	0	0	0	1

Total: 97

3

Soreness \ Dryness	0	1	2	3
0	30	25	12	0
1	4	32	9	0
2	0	3	31	3
3	0	1	7	23

Total: 180

5

Soreness \ Dryness	0	1	2	3
0	21	11	2	1
1	3	12	5	0
2	0	1	9	3
3	0	0	0	6

Total: 74

6

Soreness \ Dryness	0	1	2	3
0	41	36	11	2
1	1	13	10	0
2	1	3	36	8
3	0	0	2	16

Total: 180

7

Soreness \ Dryness	0	1	2	3
0	5	14	0	0
1	0	7	1	0
2	0	0	23	0
3	0	0	0	0

Total: 50

8

Soreness \ Dryness	0	1	2	3
0	30	4	1	0
1	10	26	0	0
2	1	5	60	8
3	0	0	5	27

Total: 177

Dryness and Irritation

1

Irritation \ Dryness	0	1	2	3
0	41	42	19	0
1	0	9	20	0
2	0	0	1	1
3	0	0	0	0

Total: 133

2

Irritation \ Dryness	0	1	2	3
0	20	19	12	1
1	3	14	13	1
2	0	1	11	2
3	0	0	0	0

Total: 97

3

Irritation \ Dryness	0	1	2	3
0	27	32	13	1
1	5	25	25	0
2	2	4	19	14
3	0	0	2	11

Total: 180

5

Irritation \ Dryness	0	1	2	3
0	20	9	6	2
1	4	14	4	0
2	0	1	6	2
3	0	0	0	6

Total: 74

6

Irritation \ Dryness	0	1	2	3
0	39	37	27	5
1	3	13	18	8
2	0	2	11	6
3	1	0	3	7

Total: 180

7

Irritation \ Dryness	0	1	2	3
0	5	21	24	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0

Total: 50

8

Irritation \ Dryness	0	1	2	3
0	31	5	5	0
1	9	27	3	0
2	1	2	52	9
3	0	1	6	26

Total: 177

Soreness and Irritation

1

Soreness \ Irritation	0	1	2	3
0	68	4	0	0
1	17	4	0	0
2	16	21	2	0
3	1	0	0	0

Total: 133

2

Soreness \ Irritation	0	1	2	3
0	39	5	3	0
1	2	15	4	0
2	11	11	6	0
3	0	0	1	0

Total: 97

3

Soreness \ Irritation	0	1	2	3
0	54	10	3	0
1	14	28	2	1
2	4	13	20	0
3	1	4	14	12

Total: 180

5

Soreness \ Irritation	0	1	2	3
0	30	5	0	0
1	5	14	1	0
2	1	3	8	1
3	1	0	0	5

Total: 74

6

Soreness \ Irritation	0	1	2	3
0	84	6	0	0
1	4	16	2	2
2	17	12	15	4
3	3	8	2	5

Total: 180

7

Soreness \ Irritation	0	1	2	3
0	19	0	0	0
1	8	0	0	0
2	23	0	0	0
3	0	0	0	0

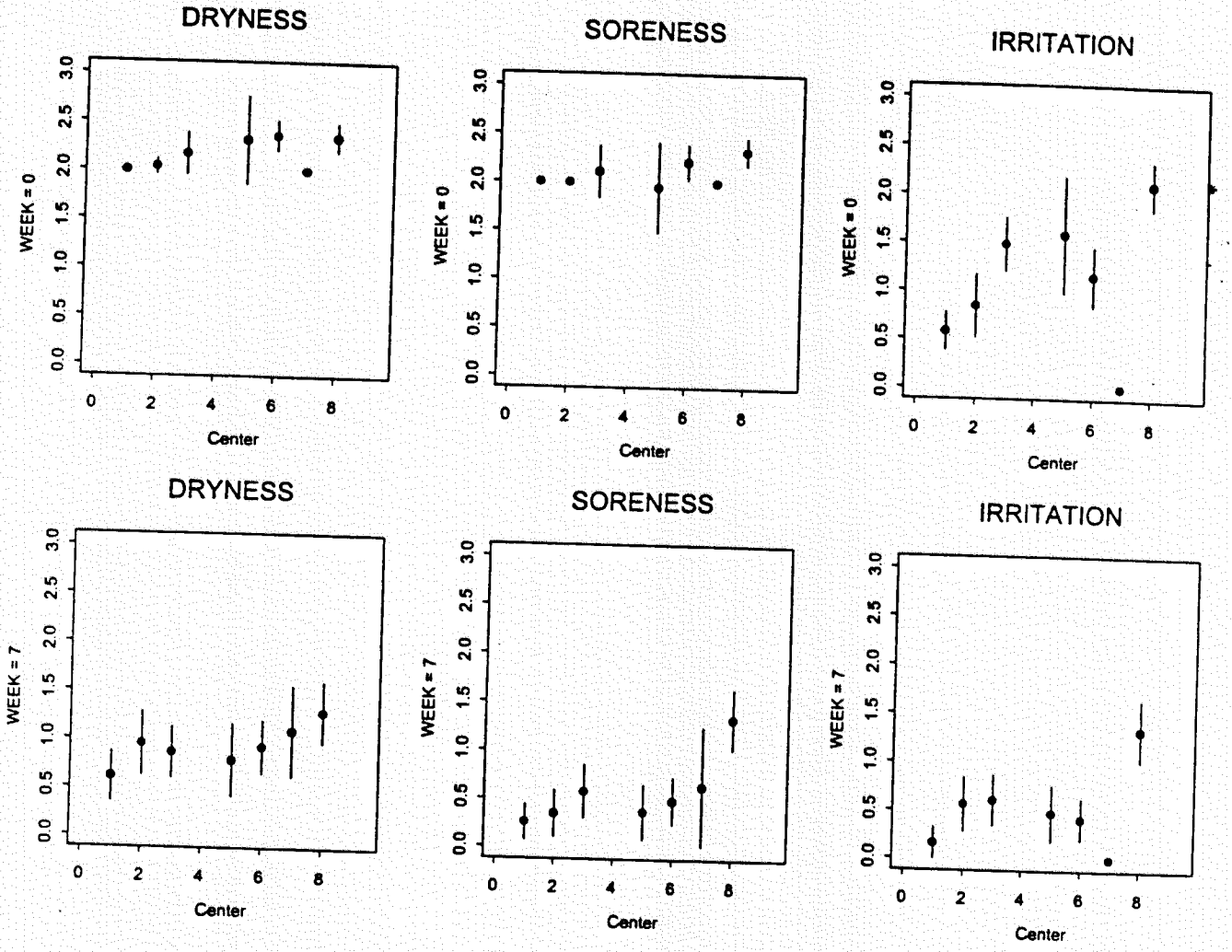
Total: 50

8

Soreness \ Irritation	0	1	2	3
0	33	2	0	0
1	3	33	0	0
2	5	4	63	2
3	0	0	1	31

Total: 177

Means and 95% Confidence Intervals At Week 0 and Week 7 By Center



Note:

a) the irritation scores at baseline and Week 7 are considerably higher in Center #8 than the other centers; and
b) the irritation scores in Center #7 = 0 at Weeks 0 and 7 (and all other visits) for all patients.

/S/

Barbara Elashoff
Mathematical Statistician

Concur: Dr. Nevius BEN 3/25/99

cc:
Orig. NDA 20-908
HFD-580 / Division File
HFD-580 / Rarick / Mann / Slaughter / Allen / Bennett
HFD-715 / Chron, division file
HFD-715 / BElashoff / ENevius