CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER: 020994, 020375/S011

STATISTICAL REVIEW(S)

Statistical Review and Evaluation

MAR 3 1999

NDA 20-994
Drug name: Climara Transdermal (estradiol transdermal system)
Applicant: Berlex Labs
Indication: Prevention of Osteoporosis

Volumes reviewed: 1.1 1.20-34 dated 1 May 1998
(Received by HFD-510 on 5 May 1998)
Electronic Data

Reviewer: Z. Jonathan Ma, Ph.D., HFD-715
User fee date: 5 March 1999

Project manager: Randy Hedin (HFD-510)
Medical reviewer: Joanna Zawadzki, M.D. (HFD-510)

Introduction

Postmenopausal osteoporosis is primarily caused by the decrease in ovarian estrogen production which is associated with menopause. A reduction in bone mineral density of approximately 1% to 3% per year begins shortly after the onset of menopause and continues for up to 10 years. The incidence of fractures is greater when BMD is less than 1.0 gm/cm^2.

Climara Transdermal was submitted by 3M Pharmaceuticals on July 15, 1993 under NDA # 20-375 and then was resubmitted on September 29, 1993. This NDA was approved on December 22, 1994 for the indications of symptoms associated with menopause and other female conditions. The ownership of the NDA was transferred to Berlex on November 2, 1995.

The current NDA is for the indication of prevention of osteoporosis in postmenopausal, hysterectomized women. One phase III study (Study 308-03B) was conducted by the sponsor for this indication.

This statistical review will mainly focus on the efficacy assessments of this NDA. The safety issues have been reviewed by the medical officer.

General Design of Study 308-03B

This was a randomized, double-blinded, placebo-controlled, parallel-group, multicenter (10 centers in the U.S.), phase III study to evaluate the safety and efficacy of four dose levels of transdermal estradiol in the prevention of osteoporosis.

The primary objective of this study was to evaluate the efficacy of treatment utilizing the four different doses of estradiol given transdermally compared to placebo, for the prevention of osteoporosis of the lumbar spine in postmenopausal, hysterectomized women.

The four doses of transdermal patches that were used in this study are as follows:

6.5 cm^2 containing 2.04 mg E_2,
12.5 cm^2 containing 3.9 mg E_2,
15 cm^2 containing 4.68 mg E_2,
25 cm^2 containing 7.8 mg E_2.

Women were eligible to enter the study if they met all of the following criteria:

- Hysterectomy with or without oophorectomy;
- In women who had not had oophorectomy, evidence of ovarian failure (amenorrhea, vasomotor symptoms, etc) ≥ 1 year, but ≤ 5 years prior to enrollment; age ≥ 45 years;
- In women with oophorectomy, oophorectomy ≥ 4 week and ≤ 5 years prior to enrollment; age ≥ 40 years;
- Serum estradiol levels ≤ 20 ph/mL;
- Serum FSH level ≥ 50 unites/L;
- BMD ≥ 0.9 g/cm^2 or BMD ≥ 0.805 g/cm^2
- Fasting baseline serum cholesterol ≤ 300 mg/dL, triglycerides ≤ 300 mg/dL, and glucose ≤ 140 mg/dL;
- Signed an informed consent.

Any of the following criteria would disqualify a subject:

- Known or suspected bone disease (including osteoporosis); BMD below 0.9 g/cm^2; hypo- or hypercalcemia; vitamin D deficiencies;
- Fracture within six months prior to the start of the study;
- Immobilization for at least two of the six months prior to enrollment;
- Any disease or condition that compromised the function of the body systems and could have resulted in altered absorption, excessive accumulation, impaired metabolism, or altered excretion of the study medication;

- Severe systemic disease which could have interfered with the conduct of the study or the interpretation of the results;
- Abnormal baseline laboratory values which were considered clinically significant and which gave suspicion of a specific organ dysfunction;
- Hot flashes of a frequency or severity that necessitated hormone treatment;

etc.

Subjects were withdrawn from the study for any one of the following reasons:

- Lumbar spine BMD below 0.9 g/cm^2 (< 0.805 g/cm^2 with) or an annualized loss greater than 6% at the 12 month visit (Visit 5) or later (defined as treatment failure);
- Occurrence for the first time of migraine headaches or more frequent occurrence of unusually severe headaches, sudden perceptual disorders (e.g., disturbances of vision or hearing), first signs of thrombophlebitis or thromboembolic symptoms (e.g., unusual pains in or swelling of the legs, stabbing pains or breathing or coughing for no apparent reason), a feeling of pain and tightness in the chest, pending operations (six weeks beforehand), immobilization (i.e., following accidents), onset of jaundice, onset of hepatitis, itching of the whole body, epileptic seizures, or a significant rise in blood pressure (defined as adverse events);
- Any condition described in the exclusion criteria (defined as adverse events).

Subjects were instructed to apply the patches to a clean, dry area of the anterior trunk which were not oily, damaged or irritated. Patches were instructed to be worn for one week before being changed to new ones.

Because the 4 active dose levels were contained in visibly distinguishable patches, blinding and randomization were accomplished by

- applying two patches of different sizes at the same time;
- randomizing each block of 11 subjects to the 5 treatment groups in a ratio of 2:2:2:2:3.

The following table illustrates the treatment assignments for every block of 11 subjects.

Table 1. The Block Randomization Scheme

Subject	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
Patch A	6.5P	6.5A	6.5P	12.5P	12.5A	12.5P	15P	15A	15P	6.5P	6.5A
Patch B	12.5P	12.5P	12.5A	15P	15P	15A	25P	25P	25A	25A	25P
Dose	0	6.5	12.5	0	12.5	15	0	15	25	25	6.5

In Table 1, "P" stands for a placebo patch an "A" stands for an active patch. The block randomization was carried out by center.

The primary efficacy variable was the percent change from baseline in the bone mineral density (BMD) of the spine (AP view, L2 - L4).

The secondary efficacy variables included: the percent changes from the baseline in the

- BMD of nondominant radius (midshaft),
- BMD of femoral neck (same side as radius),
- BMD of the total hip,
- Serum osteocalcin,

and the change and percent change from baseline in the

- urinary deoxypyridinoline/creatinine ratio, and
- urinary pyridinoline/creatinine ratio.

Bone densitometry was performed at Baseline and Visits 3 (6 months), 5 (12 months), 7 (18 months) and 9 (24 months) or the Final Visit (for subjects withdrawing) using dual-energy x-ray absorptiometry (DEXA).

The study enrolled a total of 175 postmenopausal, hysterectomized women at 10 US sites. Among them, 32 subjects were randomized into the 6.5 cm^2 groups, 31 into 12.5 cm^2 group, 31 into 15 cm^2 group, 35 into 25 cm^2 group, and 46 into placebo group.

The following table shows the demographic and baseline characteristic parameters broken down by treatment groups. Vital signs such as blood pressure and heart rate, and some bone metabolism parameters were also measured at baseline. No notable differences in these parameters between the treatment groups were found.

Table 2. Demographic and Baseline Characteristic Parameter

Characteristic Parameter	6.5 cm^2 (N=32)	12.5 cm^2 (N=31)	15 cm^2 (N=31)	25 cm^2 (N=35)	Placebo (N=46)	Total (N=175)
Age (yrs)						
Mean (S.D.)	52 (3.9)	51 (5.4)	50 (4.9)	51 (5.8)	52 (4.3)	51 (4.9)
Range	44-59	42-62	42-60	40-71	40-60	40-71
Race						
White	26 (81%)	24 (77%)	26 (84%)	32 (91%)	38 (83%)	146 (83%)
Non-white	6 (19%)	7 (23%)	5 (26%)	3 (9%)	8 (17%)	29 (17%)
Weight (lbs)						
Mean (S.D.)	162 (32)	167 (38)	158 (25)	165 (34)	163 (40)	163 (34)
Height (ins)						
Mean (S.D.)	65 (1.9)	64 (1.9)	64 (2.9)	64 (2.9)	64 (2.4)	64 (2.4)
Baseline BMD (g/ cm^2)						
Lumber Spine / N	31	30	31	34	46	172
Mean (S.D.)	1.12 (0.2)	1.10 (0.2)	1.09 (0.1)	1.12 (0.2)	1.13 (0.2)	1.11 (0.2)
Radius / N	32	31	30	35	45	173
Mean (S.D.)	0.83 (0.1)	0.79 (0.1)	0.79 (0.1)	0.81 (0.1)	0.81 (0.1)	0.81 (0.1)
Femoral Neck / N	32	31	30	35	45	173
Mean (S.D.)	0.90 (0.2)	0.86 (0.1)	0.87 (0.1)	0.86 (0.2)	0.89 (0.2)	0.88 (0.2)
Total Hip / N	28	31	29	32	44	164
Mean (S.D.)	0.98 (0.2)	0.93 (0.1)	0.91 (0.1)	0.94 (0.1)	0.95 (0.2)	0.94 (0.2)

Source: Table TT11, Vol. 20

Primary Efficacy Analyses

Sponsor's Analyses

The sponsor planned to perform the following three sets of analyses:

- Intent-to-treat (ITT) analysis: all subjects randomized to study who took at least one dose of drug. However, as will be discussed later in this section, sponsor's ITT analyses excluded a large portion of the subjects and two additional ITT analyses were jointly performed by this reviewer and the sponsor and the results will be discussed later in this review.

- Evaluable Subjects Analysis: all subjects randomized to study, who took no prohibited medications, had a 75% compliance or higher, completed at least 13 cycles, and had no major violations of the inclusion/exclusion criteria.

- End Point Analysis: Data from a subject's last visit on study medication, at which time the measurement was made, were to be used in this analysis. This analysis excluded subjects who dropped out before 6 months and included all subjects who had any post-baseline measurements.

As stated previously, the primary efficacy parameter was the mean percent change from the baseline in the bone mineral density of spine A-P view (L2-L4). The results in the following table were generated by ANOVA models with terms Treatment and Site. Pairwise comparisons were obtained using the CONTRAST statement in the SAS Procedure GLM. The p-values shown in the following table are unadjusted p-values for these pairwise comparisons.

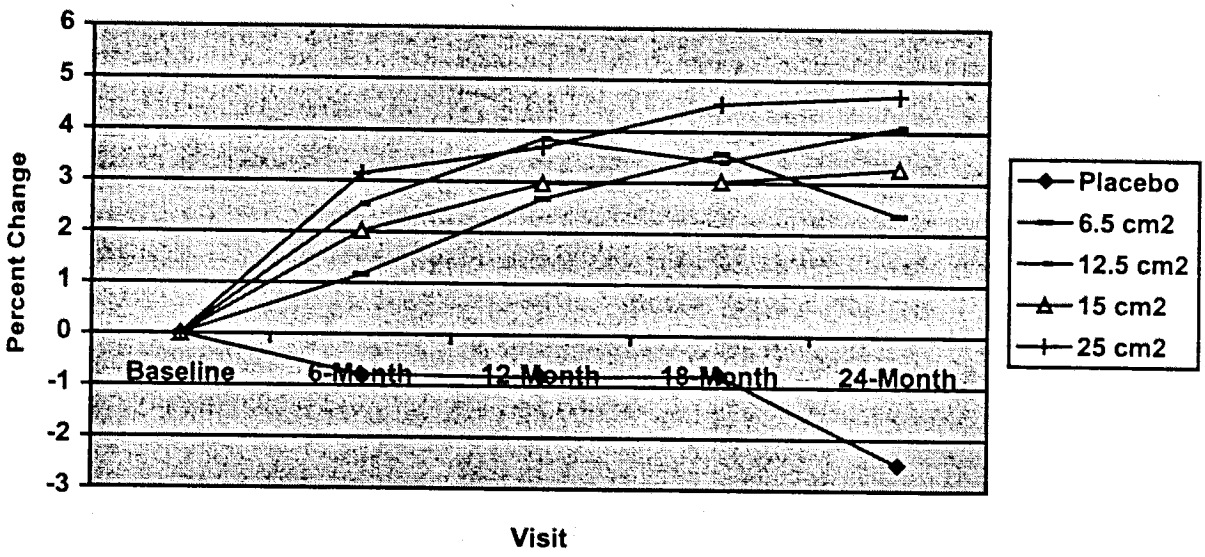
Table 3. Sponsor's Primary Efficacy Analyses

Mean Percent Change From Baseline in BMD (g/ cm^2) of Spine A-P View (L2-L4)
By Treatment and Visit

Treatment Group		6 months	12 months	18 months	24 months	Endpoint
Placebo	N/Mean	34/-0.78	26/-0.82	22/-0.78	21/-2.49	34/-2.33
6.5 cm^2	N/Mean	25/1.16	20/2.67	17/3.57	16/2.37	25/2.32
	p-Value	0.009	0.003	0.0009	0.0008	<0.0001
12.5 cm^2	N/Mean	23/2.54	21/3.84	18/3.41	18/4.09	23/3.74
	p-Value	<0.0001	<0.0001	0.0001	<0.0001	<0.0001
15 cm^2	N/Mean	24/2.02	24/2.97	22/3.02	20/3.28	25/3.45
	p-Value	0.0003	0.001	0.002	<0.0001	<0.0001
25 cm^2	N/Mean	27/3.14	24/3.68	23/4.53	21/4.70	27/5.20
	p-Value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Source: TT14, Vol. 20

Figure 1. Sponsor's Primary Efficacy Analyses



Since there were four pairwise comparisons between the active groups and the placebo group, the multiple testing p-values were adjusted by the Hochberg (1988) step-up procedure. The following gives a brief description to this procedure.

Suppose $p_{(1)}, \ldots, p_{(m)}$ are the ordered p-values (from the smallest to the largest) for the m multiple comparisons and suppose α is the overall type I error. The Hochberg procedure proceeds from the largest p-value $p_{(m)}$ to the smallest one $p_{(1)}$ comparing $p_{(i)}$ with $\alpha/(m-i+1)$, $i=m, m-1,\ldots,1$. It stops at the first i where $p_{(i)} \leq \alpha/(m-i+1)$. The null hypothesis$_{(i)}$ is then rejected and so are all of the null hypotheses with lower or equal p-values.

Since the sponsor planned an interim analysis, the overall significance level at the final testing was obtained as 0.0412 in sponsor's final report. The detailed calculation can be found in the Appendix. It is also noted in the Appendix that if the sponsor had not made changes to their plan for the interim analysis after the trial had started, the overall significance level at the final testing would be 0.036.

It is easy to see that all of the p-values in the table remain significant after adjusting for multiple comparisons using the Hochberg method even using the overall level of 0.036. Therefore, it was concluded that the difference in the primary efficacy parameter, the mean percent change from the baseline in the BMD of spine A-P view (L2-L4), between every active group and the placebo was statistically significant at all time points: 6 months, 12 months, 18 months and 24 months.

Sponsor's endpoint analysis also remained significant p-values for the same testings after adjustments for multiple testing using the Hochberg method.

What should be pointed out here, however, is that sponsor did not make any attempts to include in their analyses those subjects who dropped out between baseline and the 6 months. For example, 12 subjects in the placebo group were excluded from sponsor's primary efficacy analyses because they did not have any post-baseline measurements for the primary efficacy endpoint. Thus, the dropout rate was 26% (12 out of 46) for the placebo group. Similarly, in the same analyses, the dropout rate was 22% (7 out of 32) for the 6.5 cm^2 group, 26% (8 out of 31) for the 12.5 cm^2 group, 19% (6 out of 31) for the 15 cm^2 group, and 23% (8 out of 35) for the 25 cm^2 group. In other words, the sponsor's primary efficacy analyses did not include all subjects randomized at the baseline. The dropout rates were between 19-26% across all groups.

The Agency's Guidance (ICH E9) states that the intent-to-treat (ITT) analyses should include all randomized subjects. That is, the primary efficacy analyses should be based on all randomized subjects instead of protocol compliant patients only (i.e., excluding non-compliant patients). This analysis preserves the comparability among the treatment groups with respect to all potential risk factors and avoids inflation of type I error.

Although excluding subjects without post-randomization data could, in many circumstances, be considered a reasonable thing to do, the sponsor's primary efficacy analysis described above excluded a large portion of the randomized subjects (overall 23%, 41 out of 175) and therefore could introduce considerable bias.

With the help from the sponsor, this reviewer performed the following two additional ITT analyses on the primary efficacy parameter using conservative imputation methods. The imputation methods were intended to examine the robustness of the sponsor's ITT results. Under this reviewer's request and instructions, the sponsor also performed the same set of analyses that gave almost identical results. The slight discrepancies might have been caused by rounding errors and/or special handling of some particular subject(s).

Additional Intent-To-Treat (ITT) Analyses

- Approach I (ITT1): Last Observation Carried Forward (LOCF) method was used to impute the dropouts including those who dropped out before 6 months.

For subjects who dropped out before 6 months, baseline values were carried forward to impute the post-baseline measurements. This imputation procedure was suggested by the Medical Division. That is, the percent change from baseline in BMD of the spine was assumed to be zero for these subjects. There were three subjects (pt# 301013, 307014 and 309006) who did not have baseline measurements and their percent changes were also assumed to be zero. One subject (pt# 304004) had only baseline and 18-month measurements. His/her values of the BMD of the spine at 6 and 12 months were

imputed by linear interpolation and the 18-month measurement was carried forward to the 24-month measurement.

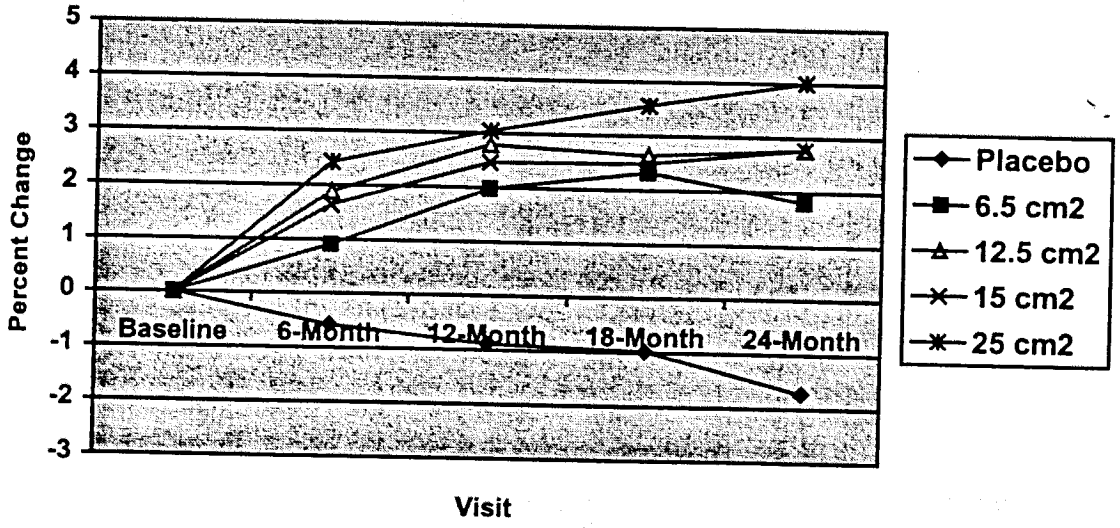
The results in Table 4 were obtained using ANOVA models with terms Treatment and Site. Pairwise comparisons were obtained using the CONTRAST statement in the SAS Procedure GLM. The p-values shown in Table 4 are unadjusted p-values for these pairwise comparisons.

Table 4. ITT1 Analyses, Primary Efficacy Endpoint

Mean Percent Change From Baseline in BMD (g/ cm^2) of Spine A-P View (L2-L4) by Treatment and Visit

Treatment Group		6 months	12 months	18 months	24 months
Placebo	N/Mean	46/-0.58	46/-0.89	46/-0.98	46/-1.72
6.5 cm^2	N/Mean	32/0.91	32/1.98	32/2.33	32/1.81
	p-Value	0.020	0.0005	0.0002	<0.0001
12.5 cm^2	N/Mean	31/1.88	31/2.79	31/2.64	31/2.77
	p-Value	<0.0001	<0.0001	<0.0001	<0.0001
15 cm^2	N/Mean	31/1.64	31/2.46	31/2.49	31/2.78
	p-Value	0.0003	<0.0001	<0.0001	<0.0001
25 cm^2	N/Mean	35/2.42	35/3.04	35/3.56	35/4.01
	p-Value	<0.0001	<0.0001	<0.0001	<0.0001

Figure 2. ITT1 Analyses, Primary Efficacy Endpoint



Again, the p-values were for the comparisons of each dose against placebo. All of the p-values remained significant after adjusting for multiple comparisons using the Hochberg method. These results were consistent with sponsor's original analyses.

- Approach II (ITT2): dropouts, including those who dropped out before 6 months, were assumed to follow the average placebo response.

To be more specific, this approach assumed that once a subject was off the patch, his/her response followed the average response of the placebo group. The average placebo response was calculated based on the observed post-baseline measurements in the placebo group. The mean percent changes in the placebo group were -0.78% at 6 months relative to baseline, -0.039% at 12 months relative to 6 months, 0.035% at 18 months relative to 12 months, and -1.72% at 24 months relative to 18 months. The three subjects (pt# 301013, 307014 and 309006) who did not have baseline measurements were also assumed to follow the average placebo response. For the subject (pt# 304004) who had only baseline and 18-month measurements, his/her values at 6 and 12 months were imputed by linear interpolation.

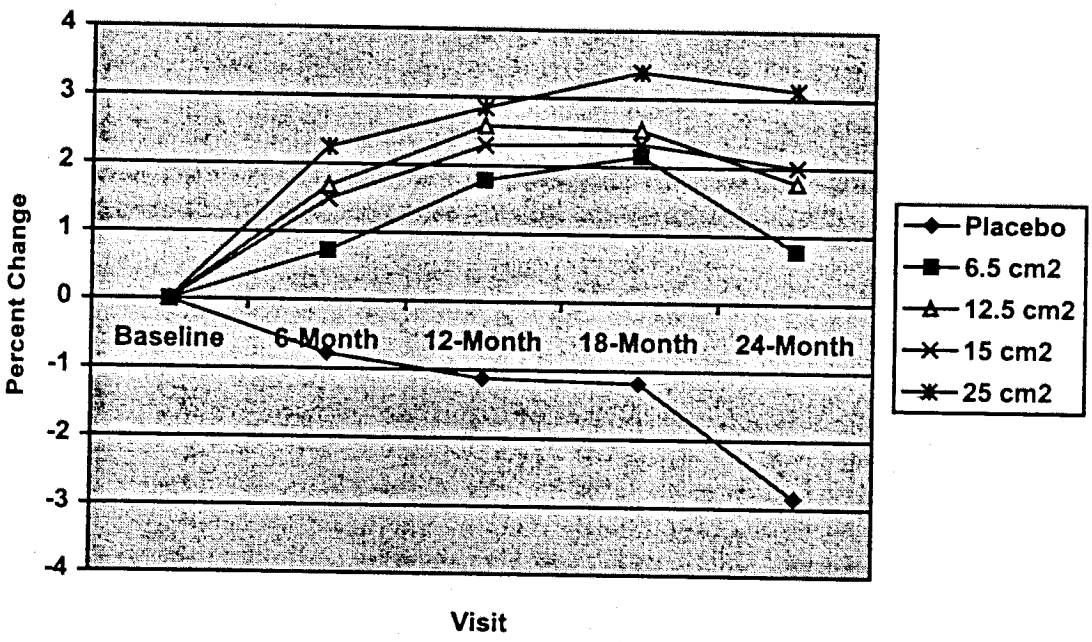
The results in Table 5 were obtained using ANOVA models with terms Treatment and Site. Pairwise comparisons were obtained using the CONTRAST statement in the SAS Procedure GLM. The p-values shown in Table r are unadjusted p-values for these pairwise comparisons.

Table 5. ITT2 Analysis, Primary Efficacy Endpoint

Mean Percent Change From Baseline in BMD (g/ cm^2) of Spine A-P View (L2-L4) by Treatment and Visit

Treatment Group		6 months	12 months	18 months	24 months
Placebo	N/Mean	46/-0.78	46/-1.12	46/-1.18	46/-2.85
6.5 cm^2	N/Mean	32/0.73	32/1.79	32/2.16	32/0.78
	p-Value	0.02	<0.0001	0.0002	0.0002
12.5 cm^2	N/Mean	31/1.68	31/2.58	31/2.54	31/1.78
	p-Value	<0.0001	<0.0001	0.0001	<0.0001
15 cm^2	N/Mean	31/1.49	31/2.30	31/2.34	31/2.01
	p-Value	0.0004	0.0006	<0.0001	<0.0001
25 cm^2	N/Mean	35/2.25	35/2.85	35/3.39	35/3.13
	p-Value	<0.0001	<0.0001	<0.0001	<0.0001

Figure 3. ITT2 Analyses, Primary Efficacy Endpoint



Again, the p-values were for comparisons of each dose against placebo. All of the p-values remained significant after adjusting for multiple comparisons using the Hochberg method. These results were consistent with sponsor's original analyses as well as the ITT1 analysis.

Pairwise Comparisons between Active Treatments

To investigate the differences among the doses, this reviewer performed pairwise comparisons between all four active treatments. The results are shown in the following table. Only the 6.5 vs 25 cm^2 comparisons at 6 and 24 months had nominally significant p-values. However, they were not significant according to the Hochberg method for adjusting for multiple comparisons. Again, the ITT1 and ITT2 analyses yielded similar results.

Nevertheless, this did not necessarily mean that the four doses performed equally. The trial was designed only to detect the possible differences between the active doses and the placebo, and did not have the power to detect any differences among the active doses. In fact, as will be shown, a linear dose-response relationship was detected from the data.

Table 6. P-Values for the Pairwise Comparisons between Active Treatments

		6.5 vs12.5	6.5 vs 15	6.5 vs 25	12.5 vs 15	12.5 vs 25	15 vs 25
ITT1	6 Months	0.11	0.22	0.014	0.71	0.42	0.23
	12 Months	0.35	0.55	0.20	0.74	0.76	0.51
	18 Months	0.59	0.79	0.13	0.78	0.35	0.22
	24 Months	0.28	0.28	0.015	0.99	0.19	0.19
ITT2	6 Months	0.14	0.23	0.019	0.78	0.41	0.26
	12 Months	0.39	0.55	0.22	0.79	0.74	0.54
	18 Months	0.64	0.79	0.15	0.83	0.34	0.24
	24 Months	0.31	0.23	0.018	0.85	0.18	0.25

Dose-response Relationship

The dose-response relationship was examined for the primary efficacy endpoint, percent change in the BMD of the spine, using the orthogonal polynomial regression analysis.

As shown in Table 7, while linear trends were significant at all four time points for the BMD of the spine (all p-values < 0.0001), quadratic trends were either not significant or only marginally significant (p-values were 0.2-0.5). Again, ITT1 and ITT2 analyses showed fairly consistent results. It should be noted that the doses used in the analysis were the actual dosage levels: (0,2.04, 3.90, 4.68, 7.8) instead of the patch sizes. These results were also consistent with those of the sponsor's original analyses which used only 6-month compliers.

Table 7. P-Values from the Orthogonal Polynomial Regression Analysis Examining Dose-Response Relationship

Visit	6 Months	12 Months	18 Months	24 Months
ITT1				
Linear	<0.0001	<0.0001	<0.0001	<0.0001
Quadratic	0.24	0.03	0.07	0.047
Cubic	0.87	0.38	0.23	0.49
Quartic	0.45	0.57	0.85	0.79
ITT2				
Linear	<0.0001	<0.0001	<0.0001	<0.0001
Quadratic	0.20	0.02	0.06	0.044
Cubic	0.95	0.63	0.35	0.80
Quartic	0.64	0.85	0.98	0.86

Figure 4a. ITT1: Means of Percent Change of BMD of Spine by Doses

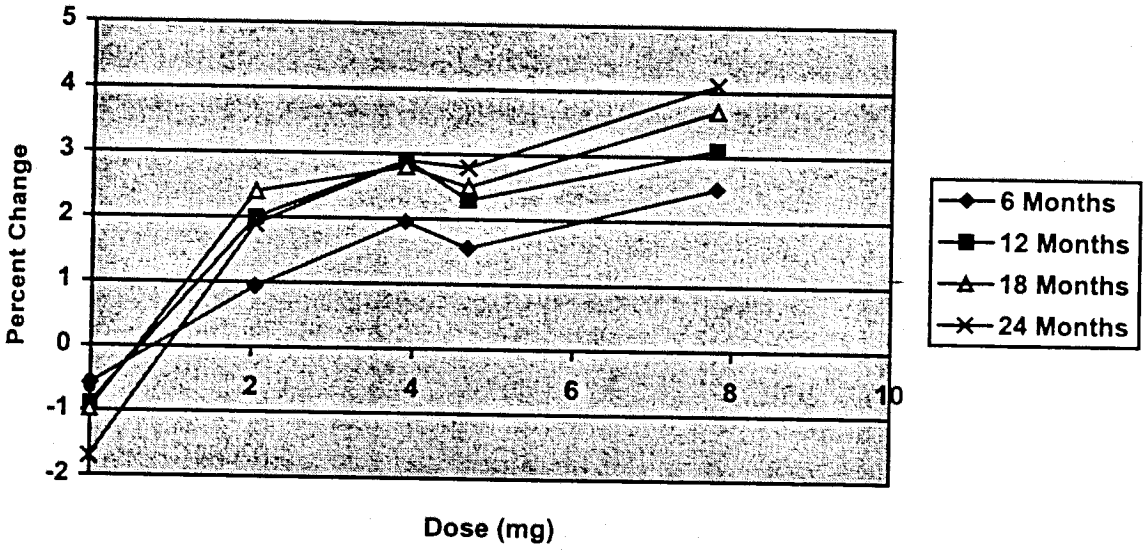
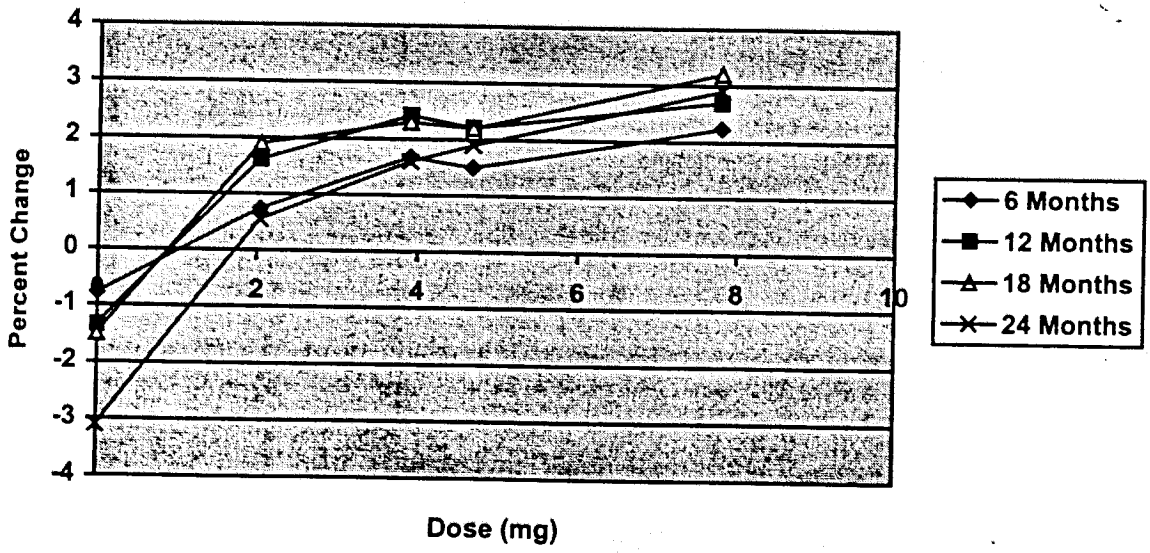


Figure 4b. ITT2: Means of Percent Change of BMD of Spine by Doses



In conclusion, it appears that dose-response followed approximately a linear trend.

Secondary Analyses

The sponsor's original analyses on the secondary efficacy endpoints were exactly the same as those on the primary efficacy endpoint. The sponsor later redid the analyses on the BMD of the total hip following the FDA's ITT analyses for the primary efficacy endpoint. The results, again, were consistent with those of their original analyses. The sponsor did not re-do the analyses on the rest of the secondary endpoints.

- BMD of nondominant radius (midshaft)

Only the 25 cm^2 group showed marginal significance at 18 and 24 months.

- BMD of femoral neck (same side as radius)

No group showed significance at any time points.

- BMD of the total hip

Sponsor's Original Analyses

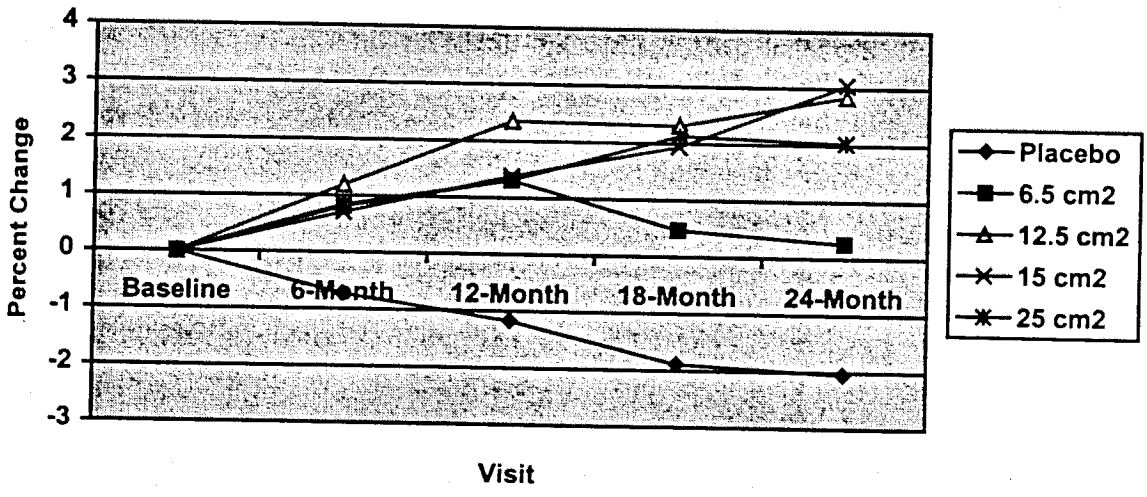
Table 8. Sponsor's Analyses on Percent Change in BMD of Total Hip

Mean Percent Change From Baseline in BMD (g/ cm^2) of Total Hip by Treatment and Visit (Sponsor's ITT)

Treatment Group		6 months	12 months	18 months	24 months	Endpoint
Placebo	N/Mean	34/-0.73	26/-1.17	22/-1.89	21/-2.04	34/-1.66
6.5 cm^2	N/Mean	23/0.81	18/1.31	16/0.47	14/0.26	23/0.65
	p-Value	0.014	0.001	0.011	0.020	0.004
12.5 cm^2	N/Mean	24/1.19	22/2.35	18/2.31	18/2.85	24/2.41
	p-Value	0.002	<0.0001	<0.0001	<0.0001	<0.0001
15 cm^2	N/Mean	23/0.71	22/1.38	21/1.94	20/3.05	23/2.61
	p-Value	0.018	0.0003	<0.0001	<0.0001	<0.0001
25 cm^2	N/Mean	24/0.84	22/1.31	22/2.13	21/2.03	25/1.98
	p-Value	0.013	0.0003	<0.0001	<0.0001	<0.0001

P-values are for comparisons of each dose against placebo. All of the p-values remained significant after adjusting for multiple comparisons using the Hochberg method.

Figure 5. Sponsor's Analyses on Percent Change in BMD of Total Hip



Additional Intent-to-treat Analysis

The following two ITT analyses were performed by the sponsor under the request and instructions of this reviewer.

- Approach I (ITT1): Last Observation Carried Forward (LOCF) method was used to impute the dropouts including those who dropped out before 6 months.

For subjects who dropped out before 6 months, baseline values were carried forward to impute the post-baseline measurements. That is, the percent change from baseline in BMD of the total hip was assumed zero for these subjects. The results are shown in the following table.