DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF BIOSTATISTICS

Statistical Review and Evaluation

CLINICAL STUDIES

NDA:	20-560 / (b) (4) 038
Name of drug:	Fosamax (alendronate)
Applicant:	Merck
Indication:	(b) (4)
Location of electronic datasets	\\Cdsesub1\n20560\S_038\2003-01-31\crt\datasets\stat\MK0217\OI_year_1\Prot135\Data_analysis
Documents reviewed	\\Cdsesub1\n20560\S_038\2003-01-31\clinstat\studies Technical section for statistics volumes 1-4 (hardcopy)
Project manager:	Randy Hedin, B.S. (HFD-510).
Clinical reviewer:	Bruce Schneider, M.D. (HFD-510)
Dates:	Received 1/31/03; user fee (6 months) 7/31/03
Statistics team leader and statistical reviewer:	J. Todd Sahlroot, Ph.D. (HFD-715)
Biometrics division director and secondary reviewer:	S. Edward Nevius, Ph.D. (HFD-715)

Keywords: NDA review, clinical study, pediatric exclusivity, one study application

Table of Contents

1 Summary and conclusions

The sponsor submitted 12-month data for an ongoing, randomized, double-blind, placebo-controlled, multicenter trial of Fosamax (alendronate) in 139 pediatric patients with osteogenesis imperfecta (OI). The primary endpoint was change from baseline in lumbar spine BMD z-score at 12 months. Randomization to Fosamax or placebo (3:1 ratio) was stratified by baseline body weight. The Fosamax daily oral dose was 5 mg if patients weighed less than 40 kg at baseline, and 10 mg if baseline weight was at least 40 kg.

The primary analysis was performed on a modified ITT population of 112 patients. Fosamax (5 mg and 10 mg combined) showed statistically significant increases compared to placebo on the primary endpoint ($p < .001$). Mean changes from baseline were +0.99 for Fosamax and +0.09 for placebo. Least square mean changes from baseline were similar, +1.02 and +0.10. Lumbar spine BMD mean % changes from baseline, a more traditional measure of efficacy, were +37% for Fosamax and +11% for placebo. Additional analyses of the primary endpoint that included the 27 patients omitted from the sponsor's analysis attested to the robustness of the primary result.

Results were statistically significant ($p < .05$) for each dose of Fosamax compared to placebo on the primary endpoint. The 5 mg dose was numerically (but not statistically) superior to 10 mg, consistent with the fact that patients on 5 mg received a higher mg per kg dose (21 mg/kg) than patients receiving 10 mg (17 mg/kg).

There were significant treatment differences within each of the primary subgroups defined by gender, race, age, OI type and pubertal status as well as subgroups defined by median weight and height. Interactions of treatment with height and weight were statistically significant ($p < 0.01$) indicating that, while all subgroups achieved a benefit on Fosamax, smaller or lighter patients had greater treatment effects (about twice as large) compared to larger or heavier patients. These larger treatment effects were also seen in closely related subgroups, patients who were younger (<12 years) or pre-pubertal. Taken together, the larger effects seen in certain subgroups are supported by the fact that the subgroups are related clinically.

Fractures were measured radiologically and were reported separately by the investigators whether confirmed by x-ray or not. The Month-12 data show non-statistically significant treatment differences. Although the study was not powered on these secondary endpoints, the results were consistent with a null effect of Fosamax rather than a lack of statistical power.

An interim analysis was conducted on 76 patients (68% of the final sample size). The trial was not stopped at the interim analysis despite a statistically significant

result at the 1% level. The interim analysis had no appreciable effect on the overall result of the trial.

2 Introduction

The sponsor submitted data from a randomized, double-blind, placebo-controlled, multicenter trial for Fosamax (alendronate) in pediatric patients with osteogenesis imperfecta (OI). OI is a group of hereditary disorders of Type I collagen characterized by osteoporosis and increased fracture risk. The condition leads to chronic bone pain, worsening deformities, short stature and functional limitations.

The data were submitted in response to an FDA Written Request (WR) to conduct pediatric studies, dated October 27, 2000, and amended on March 8, 2002. The WR amendment clarified aspects of the study primarily involved with the interim analysis including design, timeframe and data analysis.

The original study protocol was written (b) (4) Merck contracted with (b) (4) in late 2000 to assume an active role in the daily conduct and management of the trial. The protocol was amended 3 times, Amendment 1 being the most significant of the three. Amendment 1 was a thorough revision of the original protocol that maintained the study length and primary endpoint.

The submission consists of efficacy results for the first 12 months of the 24-month controlled portion of the ongoing trial. Merck's in-house case report form cutoff was June 11, 2002. Investigator-reported fracture data through 24 months were available in the EDR. Complete data were available for 24-month completers (n=52) and discontinuations (n=22) and partial data for patients continuing their 2nd year in the trial (n=65) as of June 11, 2002. Only the 12-month fracture data were reviewed since the 24-month data were incomplete for about half the patients.

Table 1 shows major study characteristics.

Table 1. Study characteristics.

Trial # Centers Dates	Patients	# randomized [1]	Design Primary endpoint	Duration of Double – blind period
Prot 135 16 centers 15 US 1 CAN Started 5/99 Ongoing	Males and females 4-18 years with osteogenesis imperfecta (OI)	Placebo n=30 Fosamax 5mg n=84 10mg n=25	R, DB, PC, || Change from baseline in lumbar spine bone mineral density (BMD) z-score at 12 months [2]	24 months. Efficacy data reported for first 12 months (Data cutoff 6/11/02)

[1] Fosamax dose based on patient weight at baseline (5mg if < 40kg, 10mg if ≥ 40kg)
[2] BMD z-score = Number of standard deviations from the mean BMD for age-matched healthy controls

3 Design

The objective of the trial was to compare the effects of 2 doses of Fosamax combined versus placebo in pediatric patients ages 4 through 18 with severe OI for (1) change in mean lumbar spine (L1 to L4) BMD at Month 12 and for (2) safety and tolerability.

The primary endpoint was change from baseline in lumbar spine BMD z-score at 12 months. Radiologically-confirmed and investigator-reported (IR) fractures were secondary endpoints.

Femoral neck and total hip BMD were not formal endpoints but were required in patients without assessable lumbar spine vertebrae. Some patients with lumbar spine BMD data also had these data collected. These endpoints were not reviewed here since only 6 patients had femoral neck and total hip BMD measurements at Month 12.

BMD was measured with Hologic densitomiters. BMD data were read by technicians in the radiology department (b) (4)

Patients were randomized to Fosamax or placebo in an approximate 3:1 ratio. Randomization was stratified by baseline body weight. The Fosamax daily oral dose was 5 mg if patients weighed less than 40 kg at baseline, 10 mg if baseline weight was at least 40 kg.

Study visits were at Months 0 (baseline), 3, 6, 9 and 12 during the first 12 months of the double-blind period. DEXA measurements of lumbar spine BMD were performed at baseline and Months 6 and 12. Radiology (fractures) was performed at Baseline and Month 12. Clinical fractures and adverse experiences were recorded at each visit.

The final sample size calculations were provided in Amendment 1. The trial had 99% power on the primary endpoint to demonstrate a between-treatment difference equal to one assuming a SD = 1 and a 2-sided 5% type I error rate. The power calculation was based on a 2-stage testing procedure applied to the interim and final analyses assuming Fosamax and placebo sample sizes of 48 and 16 at the interim analysis, and 90 and 30 at the final analysis. Section 6 (Statistical Methods) describes the 2-stage procedure in more detail.

4 Baseline / demographic variables

Table 2 shows key demographic/ baseline variables for all randomized patients. There were no significant imbalances between groups.

Table 2. Key demographic / baseline variables

	Placebo (n=30)	Fosamax (n=109)	Total (n=139)
Sex			
Females	16 (53%)	62 (57%)	78 (56%)
Males	14 (47%)	47(43%)	61 (44%)
Age (yrs)			
Mean (SD)	11.1 (4.0)	11.0 (3.6)	11.0 (3.7)
Range	4 –18	4 –19	4 –19
Race			
White	25 (83%)	88 (81%)	113 (81%)
Black	2 (7%)	6 (6%)	8 (6%)
Asian	0	2 (2%)	2 (1%)
Multiracial	0	1 (1%)	1 (1%)
Hispanic	1 (3%)	11 (10%)	12 (9%)
Indian	1 (3%)	1 (1%)	2 (1%)
Native American	1 (3%)	0	1 (1%)
OI type			
Type I	26 (24%)	6 (20%)	32 (23%)
Type II	32 (30%)	7 (23%)	39 (28%)
Type IV	37 (34%)	17 (57%)	54 (39%)
Type unknown	13 (12%)	0	13 (9%)
Type not recorded	1 (1%)	0	0
Pubertal status			
Prepubertal	12 (40%)	51 (47%)	63 (45%)
Pubertal	18 (60%)	58 (53%)	76 (55%)
Body weight (kg)	(n=30)	(n=108)	(n=138)
Mean (SD)	31.9 (16.8)	32.0 (18.3)	32.0 (17.9)
Range	13 – 67	11 – 96	11 – 96

5 Disposition

Table 3 shows patient disposition during the study. Month 6 and Month 12 completers shown in the Table were patients who completed at least 180 and 360 days, respectively. Greater than 80% of patients in both arms completed 12 months. There was no single predominant reason for discontinuation for the 16 Fosamax patients discontinuing before Month 12 (sponsor's designation of completion status).

Table 3. Duration on study during double-blind period

	Fosamax	Placebo	Total
Baseline	109 (100%)	30 (100%)	139 (100%)
Month 6	103 (94%)	30 (100%)	133 (96%)
Month 12	89 (82%)	27 (90%)	116 (83%)
Month-12 completers [1]	93 (85%)	29 (97%)	122 (88%)
Sponsor's MITT endpoint [2]	86 (79%)	26 (87%)	112 (81%)

[1] Sponsor's designation

[2] MITT = modified intent to treat population consisting of all randomized patients except those excluded on the basis of 6 criteria

6 Statistical methods

The sponsor performed primary efficacy analyses on a modified intent to treatment (MITT) population consisting of 112 patients. The MITT excluded 27 randomized patients on the basis of 6 criteria listed below:

- Baseline lumbar spine scan performed on Lunar machine instead of Hologic (n=3)
- Patient unevaluable due to scoliosis (n=1)
- Baseline scan performed on Hologic 2000 and Month 6 performed on Hologic 4500. One of 2 machines sold before a cross-calibration between machines could be conducted (n=4)
- Metallic hardware present in spine at baseline or Month 6 (n=6)
- Baseline lumbar spine BMD data stored on optical disk could not be retrieved (n=3)
- Patient discontinued from study prior to the Month 6 evaluation (n=10)

The protocol-defined statistical model was ANCOVA of Month-12 lumbar spine BMD z-score change from baseline. Treatment, center and baseline weight (stratification variable) were factors in the model. Baseline lumbar spine BMD z-score was a covariate. Centers were combined so that there were at least 2 patients in each treatment at each combined center**.** The combining algorithm was defined in the Data Analysis Plan (DAP).

The primary efficacy variable was analyzed, per the DAP, within a number of subgroups: sex, race (white, other) age (< 12 years, ≥ 12 years), OI disease type (Type I, other), pubertal status (pre-pubertal, pubertal), baseline lumbar spine BMD z-score, presence of fractures during the year before the study, biochemical markers, height, weight, BMI and arm span. Consistency of the primary endpoint across subgroups was assessed by treatment-by-subgroup interaction terms added to the primary statistical model. The median value was used to determine subgroup cut points for continuous variables. Within-subgroup analyses were preformed only if there were 6 patients in the Fosamax group and 2 patients in the placebo group. The sponsor considered these analyses exploratory since it was thought the within-subgroup sample sizes too small to detect statistically significant differences between drug and placebo.

The sponsor applied a 2-stage testing procedure (see reference Elashoff and Reedy) at the interim and final analyses to maintain a trial-wide (overall) 5% Type I error. The interim alpha was set at 1% by the WR, the final alpha at 4.8% by the sponsor. According to the sponsor, the 4.8% alpha at the final analysis preserved the overall 5% rate given the actual sample sizes at the interim (Fosamax n=56 and placebo=20) and final analyses (Fosamax n=86 and placebo n=26).

7 Sponsor's results

Lumbar spine BMD at Month 12

Fosamax (5 mg and 10 mg combined) showed statistically significant increases compared to placebo in lumbar spine BMD z-score change from baseline ($p < .001$). The mean change from baseline was +0.99 for Fosamax and +0.09 for placebo. Least square mean changes from baseline were similar, +1.02 and +0.10. Table 4 shows the sponsor's results for lumbar spine BMD z-score, the primary endpoint in the study, and percent change from baseline, a traditional study endpoint based on the raw BMD measure, included as a reference point.

Table 4. Sponsor's 12-month lumbar spine BMD results
Modified ITT population

Lumbar spine BMD measure	Alendronate (n=86)	Placebo (n=26)	Treatment Difference	p-value
z-score (unit-less)				
Baseline mean	-4.6	-4.6		
Month 12 mean	-3.6	-4.5		
Change from baseline				
Observed mean (SE)	+0.99 (0.08)	+0.09 (0.11)		
LSM [1]	+1.02	+0.10	+0.92	< .001
95% CI for LSM			(+0.62, +1.22)	
Raw BMD score (g/cm^2)				
Baseline mean	0.37	0.36		
Month 12 mean	0.48	0.40		
% change from baseline				
Observed mean (SE)	+37.3 (3.9)	+11.2 (3.5)		
LSM [2]	+35.1	+7.4	+27.7	< .001
95% CI for LSM			(+13.3, +42.1)	

[1] Least square mean (LSM) adjusted for center, stratum and baseline lumbar spine BMD z-score.
[2] Least square mean (LSM) adjusted for center and stratum

Table 5 shows lumbar spine BMD results by randomization stratum. Treatment differences in Stratum I were numerically higher than treatment differences in Stratum II but the treatment effects in the two strata were not significantly different (treatment-by-stratum interaction p-value = 0.26).

Table 5. Sponsor's 12-month lumbar spine BMD by stratum [1]
Modified ITT population

Lumbar spine BMD	Alendronate (n=86)	Placebo (n=26)	Treatment difference
Z-score change from baseline (mean)			
Stratum I (n=88)	1.03 (n=68)	0.04 (n=20)	0.99
Stratum II (n=24)	0.81 (n=18)	0.25 (n=6)	0.56
% change from baseline (mean)			
Stratum I (n=88)	41.5% (n=68)	10.2% (n=20)	31.3%
Stratum II (n=24)	21.5% (n=18)	14.5% (n=6)	7.0%

[1] Stratum I = baseline weight < 40 kg; Stratum II = baseline weight ≥ 40 kg

Lumbar spine BMD at Month 6

Fosamax lumbar spine BMD was statistically greater than placebo at Month 6 ($p < .001$). The LS mean treatment difference for z-score change from baseline was 0.66 (95% CI = (0.43, 0.90)). The LS mean treatment difference for % change from baseline was 19.4% (95% CI = (9.0%, 29.8%)).

Subgroups

Table 6 shows the sponsor's analysis results for selected subgroups. Nominally significant subgroup-by-treatment interactions were found for weight (p=0.034), height (p=0.036) and arm span (not shown, p=0.001).

Table 6. Sponsor's analysis of lumbar Spine BMD z-score change from baseline by subgroups [1]

Subgroup (subgroup-by-treatment interaction p-value)	Fosamax			Placebo			Trt Diff [2]
	N	Mean change	SE	N	Mean change	SE	
Age (p=0.196)							
< 12 years	47	+1.12	0.10	13	+0.03	0.18	1.09
≥ 12 years	39	+0.82	0.11	13	+0.15	0.14	0.67
Gender (p=0.655)							
Male	53	+0.97	0.09	13	+0.10	0.19	0.87
Female	33	+1.02	0.14	13	+0.08	0.12	0.94
Race (p=0.457)							
White	69	+0.92	0.08	21	+0.08	0.13	0.84
Other	17	+1.24	0.18	5	+0.14	0.19	1.10
OI type (p=0.208)							
Type I	20	+0.87	0.09	5	+0.18	0.11	0.69
Other (known)	55	+1.08	0.11	21	+0.07	0.14	1.01
Pubertal status (p=0.053)							
Pre-puberty	43	+1.12	0.11	10	-0.06	0.22	1.18
Puberty	43	+0.86	0.11	16	+0.18	0.12	0.68
Z-score for weight (p=0.034)							
Above or at median	41	0.89	0.09	17	+0.18	0.11	0.71
Below median	44	1.11	0.12	9	-0.09	0.24	1.20
Z-score for height (p=0.036)							
Above or at median	40	0.86	0.09	14	+0.16	0.13	0.70
Below median	43	1.23	0.12	9	-0.17	0.23	1.40

[1] Adapted from Sponsor's Appendices 4.2.22 and 4.2.26
[2] This column added by reviewer

Interim analysis

The interim analysis was conducted on 76 patients or 68% of the final sample size. The trial was not stopped at the interim analysis despite a statistically

significant result at the 1% level. The interim analysis had no appreciable result on the overall result of the trial.

Fractures

Fractures were assessed by 2 methods, radiological and investigator-reported. Radiological fractures of long bones were assessed by a panel of 3 experts. [1] Endpoints were fracture incidence and fracture rate. Table 7 shows the percentage of patients with radiological fractures at Month 12. For patients whose x-rays were re-read, the Table shows the initial read. There was no difference between treatments in fracture incidence (Fosamax 55% vs placebo 48%, p=.50). An analysis of long bones without hardware in patients with at least one assessable long bone without hardware gave similar results.

Table 7. Number (%) of patients with at least one new radiologically confirmed fracture at Month 12

	Fosamax (n=94)	Placebo (n=27)	Treatment Comparison
# patients (%)	52 (55%)	13 (48%)	
Relative risk [1]			
Point estimate			1.15
95% CI			(0.75, 1.78)
p-value			0.50

[1] Adjusted by stratum

The mean numbers of assessable long bones per patient were 11.1 and 11.5 in the Fosamax and placebo groups, respectively. The mean rate of radiologic fractures was 1.11 for both groups. Table 8 shows the frequency distribution of radiologically-confirmed fractures to Month 12.

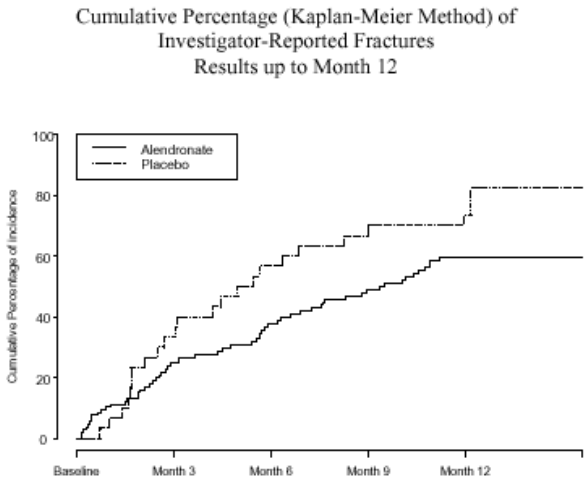
[1] Fracture data were originally read by 3 central independent radiologists. A new expert panel was constituted at the request of the DSMB to re-assess the radiologic data due to discrepancies between BMD data and radiologic fracture data, between radiologic and IR fractures, between radiologic fractures and vertebral height data, and possible ascertainment bias in the reporting of radiologic fractures.

Table 8. Frequency distribution of # of Radiologically-confirmed fractures up to Month 12

# fractures	Fosamax (n=94)			Placebo (n=27)		
	n	%	Cum %	n	%	Cum %
7	0	0	0	1	4	4
6	1	1	1	0	0	4
5	3	3	4	0	0	4
4	3	3	7	0	0	4
3	7	7	15	4	15	19
2	12	13	28	3	11	30
1	26	28	55	5	19	48
0	42	45	100	14	52	100

The sponsor performed a reliability assessment on 24 patients to measure the reproducibility of readings by the expert panel. This exercise was undertaken presumably to help explain the absence of even a numerical advantage for Fosamax in fracture reduction. The kappa statistic was 0.81 (95 CI = (0.67, 0.96)). (A kappa of zero denotes chance agreement.) The sponsor cited the observed kappa as evidence that reproducibility was “far from perfect”.

At each visit, patients or their guardians reported fractures whether radiologically-confirmed or not, that occurred since the last visit. These data (clinical fractures in any part of the body) were collected starting when Merck assumed day-to-day management of the trial. The primary IR-fracture endpoint was time to first IR-fracture. There was a statistical trend in favor of Fosamax (see graph below, p=.058).



The time-to-first IR fracture result was consistent with the incidence data in Table 9 showing showing fewer patients with at least one new IR fracture for Fosamax at Month 12 (57% vs 77%, p=.057).

Table 9. Number (%) of patients with at least one new investigator-reported fracture at Month 12

	Fosamax (n=106)	Placebo (n=30)	p-value [1]
# patients (%)	60 (57%)	23 (77%)	.057

[1] Fishers Exact test performed by reviewer

While fewer (%) Fosamax patients reported fractures, the mean annualized rate of IR fractures was numerically higher with Fosamax, 1.51 vs 1.33 for placebo. Table 10 shows the sponsor's frequency distribution of IR fractures to Month 12.

Table 10. Frequency distribution of # of Investigator-reported fractures up to Month 12

# fractures	Fosamax (n=106)			Placebo (n=30)		
	n	%	Cum %	n	%	Cum %
≥5	9	9	9	0	0	0
4	4	4	12	2	7	7
3	8	8	20	2	7	13
2	18	17	37	7	23	37
1	21	20	57	12	40	77
0	46	43	100	7	23	100

8 Reviewer's analysis and comments

Lumbar spine BMD at Month 12

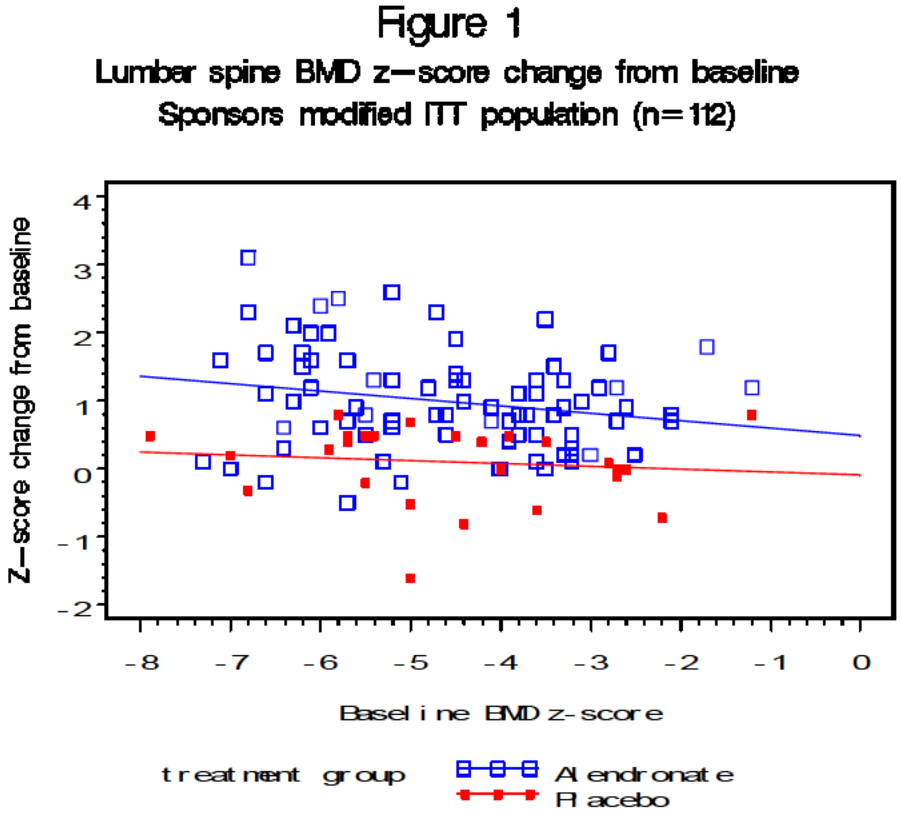
Appendix 1 shows the availability of primary endpoint data for the 27 patients excluded from the sponsor's modified ITT analysis. As shown in the Appendix, four patients with baseline and post-randomization data were excluded. Putting these patients back into the analysis had no appreciable effect on the results.

This reviewer also performed an analysis of the primary endpoint with all randomized patients including the 27 randomized patients omitted from the MITT analysis. The analysis imputed a "worst-case" treatment effect in the 27 excluded patients that, combined with the observed treatment effect in the rest of the patients, would still yield an overall significant result for the trial. The method

is described in Appendix 2. Results showed the treatment effect in the 27 excluded patients could be –1.74 (i.e., placebo could be *superior* to Fosamax by 1.74 SD units) and the overall result would still be significant at the 5% level. This statistical exercise shows the result to be extremely robust to the missing data.

Figure 1 shows individual patient data on the primary endpoint as a function of baseline. Regression lines are drawn for each treatment group. The vertical distance between the lines is the treatment effect. The graph shows that patients presenting with more severe OI (smaller baselines) had slightly larger treatment differences.

Figure 1



I repeated Figure I for each stratum (Figures 2 and 3). In stratum 1, the treatment difference was greatest for the **most** severe patients (low baseline z-score). By contrast, in stratum 2, the treatment difference was greatest for the **least** severe patients (high baseline z-score). Despite the visual differences between the 2 strata, the treatment-by-stratum interaction was not significant. The lack of significance can be attributed to (1) the small placebo sample size in Stratum 2 (n=6) and (2) the fact that the interaction looks only at the average treatment effect within each stratum and not the treatment effect as it changes with baseline.

The lack of dose response is not particularly troubling since patients receiving 10 mg received less drug on average on a mg per kg basis (0.17 mg/kg) than patients receiving 5 mg (0.21 mg/kg).

Figure 2

Lumbar spine BMD z−score change from baseline
stratum 1: fosamax 5 mg (n=88)

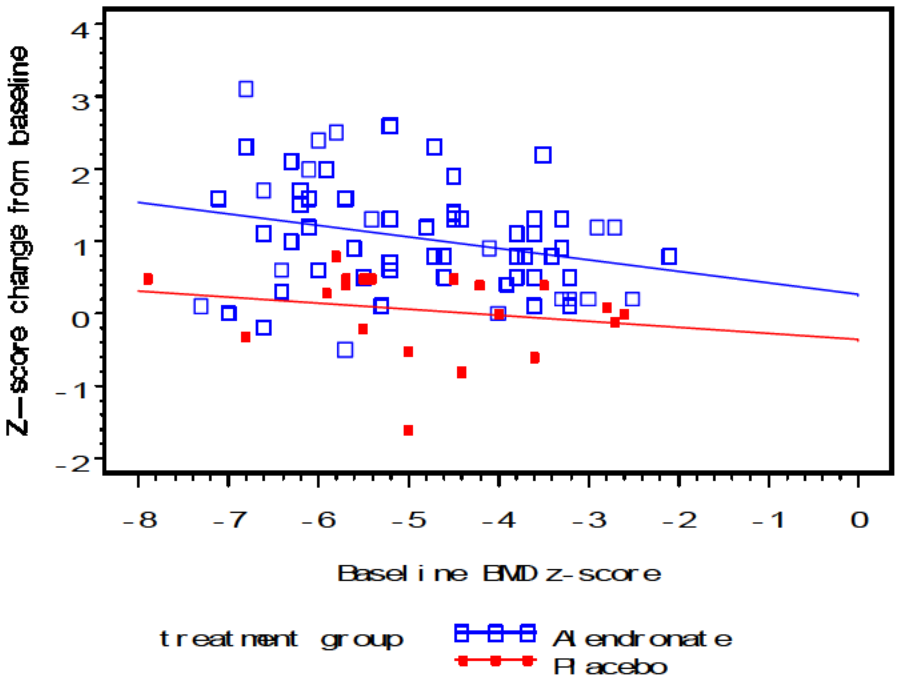
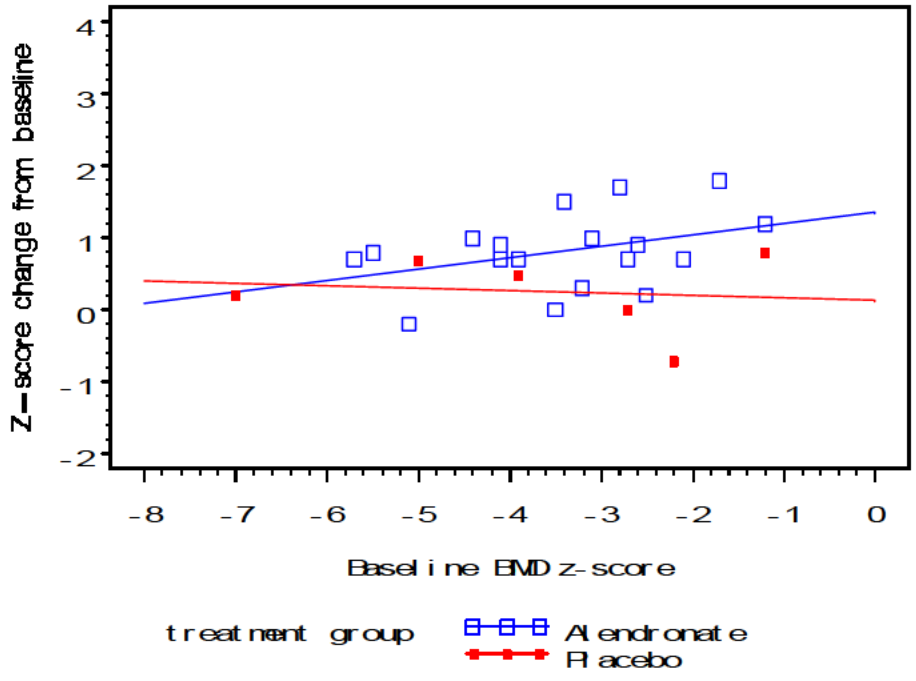


Figure 3

Lumbar spine BMD z−score change from baseline
stratum 2: fosamax 10 mg (n=24)



Subgroups

The power of the trial was 99% on the primary endpoint. The high power of the trial resulted in reasonable power to examine subgroup effects. For example, to detect an effect equal to one Z-score unit (equal to the treatment effect in the original sample size calculation), the power for tests of interaction had approximately 65% power for subgroups of roughly equal size. This 65% figure would apply to subgroups for continuous variables defined by the median but not necessarily to categorical variables with pre-defined categories where the power would be less than 65% if the subgroups were highly unbalanced. Analyses *within* subgroups of approximately ½ the total sample size (for example, males which were 56% of all randomized patients) had at least 90% power.

Therefore, it was not surprising there were significant treatment differences ($p<.05$) within each of the primary subgroups for gender, race, age, OI type and pubertal status.

The sponsor reported the results of subgroup analyses by weight and height using z-scores to define the subgroups. I performed subgroup analyses for baseline weight and height using the raw values to define the subgroups.

Table 11 shows subgroup results for weight and height. Interaction p-values were more significant than those obtained by the sponsor. The interactions were quantitative, not qualitative (i.e., not characterized by a change in sign of the treatment effect). Within subgroup p-values were also significant. Statistical results for height and weight were quite similar due to the high correlation between the variables (Spearman's $\rho = 0.87$).

Table 11. Lumbar Spine BMD z-score change from baseline by baseline weight and height

Subgroup (subgroup-by-treatment interaction p-value)	Fosamax			Placebo			Trt Diff	Within-Subgroup p-value
	N	Mean change	SE	N	Mean change	SE		
Weight (.009)								
Above or at median	40	+0.83	0.09	16	+0.23	0.11	0.60	.002
Below median	45	+1.16	0.12	10	-0.14	0.21	1.30	<.001
Height (.001)								
Above or below median	41	+0.86	0.10	12	+0.25	0.14	0.61	.027
Below median	42	+1.12	0.12	11	-0.20	0.18	1.32	<.001

Alpha at the final analysis

The planned sample size at the interim analysis was 50% of the final sample size. Given a 1% alpha and the planned sample size at the interim analysis, the final alpha would have been 4.5%. The actual sample size at the interim analysis was ~ 2/3 of the final sample. An appropriate final alpha is 4.7%, close to the sponsor's stated alpha of 4.8%. In any case, the results for the primary endpoint at the 12-Month time point are not in question.

Fracture data

Partial IR fracture data were available electronically through 24 months. The 24-month data were complete for 52 patients who completed 24 months and 22 patients who discontinued. The 24-month data were not complete for 65 patients continuing their 2nd year in the trial as of June 11, 2002. This reviewer did not analyze IR fracture data beyond 12 months.

There was a 15% increased risk in having at least one new radiologic fracture with Fosamax compared to placebo at Month 12. The mean numbers of fractures per patient were the same in both groups (1.11). The overall effect of Fosamax on radiologic fractures was null. The "far from perfect" kappa value of 0.81 is actually "almost perfect" using the designations established by Landis and Koch (see reference). In the opinion of this reviewer, the kappa statistic cannot be used as justification for the null result.

The time to first IR fracture trended in favor of Fosamax (p=.058). An analysis of the frequency distribution, however, shows aspects of the fracture data favoring placebo. While the mean numbers of IR fractures in the 2 groups were similar (Fosamax 1.51 vs placebo 1.33), the SD in the Fosamax group was almost twice that in the placebo group, 2.13 vs 1.12. Clearly some Fosamax patients had a high number of multiple IR fractures. Table 12 is a modified version of Table 9 breaking out the frequency distribution for patients with $\geq$ 5 IR fractures. The maximum number of fractures reported was 10. The mean numbers of fractures for Fosamax and placebo patients who had at least one IR fracture were 2.67 and 1.74, respectively (t test; p=.008). While this is not a randomized comparison and therefore the p-value is of limited value, this finding may still be of clinical interest.

Table 12. Frequency distribution of # of Investigator-reported fractures up to Month 12

# fractures	Fosamax (n=106)			Placebo (n=30)		
	n	%	Cum %	n	%	Cum %
10	1	1	1	**0**	**0**	**0**
9	1	1	2	**0**	**0**	**0**
8	1	1	3	**0**	**0**	**0**
7	2	2	5	**0**	**0**	**0**
6	2	2	7	**0**	**0**	**0**
5	2	2	9	**0**	**0**	**0**
4	4	4	12	2	7	7
3	8	8	20	2	7	13
2	18	17	37	7	23	37
1	21	20	57	12	40	77
0	46	43	100	7	23	100

10 Reference

Elashoff J and Reedy TJ. Two-stage clinical trial stopping rules. *Biometrics* 40, 791-795, 1984

Landis RJ and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 33:159-174, 1977

Appendix 1. Accounting of available data for the 27 patients excluded from randomized (ITT) population (n=139)

Obs number	Patient ID	Trt group F=fosamax P=placebo	Patient has data at Month [1]			Reason for exclusion from ITT [2]	12-month completion status (Y=1/N=0 (last day))
			0	6	12		
1	287	F	◦	◦	◦	1	1
2	288	F	◦	◦	◦	1	1
3	253	**P**	◦	◦	◦	1	1
4	416	F	◦	◦	◦	2	1
5	088	F	●	●	●	3	1
6	091	F	●	●	●	3	1
7	307	F	●	●	●	3	1
8	320	F	●	●	◦	3	0 (day 269)
9	321	F	◦	◦	◦	4	1
10	382	F	◦	◦	◦	4	1
11	347	F	◦	◦	◦	4	1
12	423	F	◦	◦	◦	4	1
13	404	**P**	◦	◦	◦	4	1
14	421	**P**	◦	◦	◦	4	1
15	019	F	◦	◦	◦	5	1
16	269	F	◦	●	◦	5	1
17	286	**P**	◦	◦	●	5	1
18	254	F	●	◦	◦	6	0 (day 243)
19	060	F	●	◦	◦	6	0 (day 128)
20	087	F	◦	◦	◦	6	0 (day 255)
21	319	F	●	◦	◦	6	0 (day 256)
22	305	F	●	◦	◦	6	0 (day 364)
23	366	F	●	◦	◦	6	0 (day 103)
24	150	F	●	◦	◦	6	0 (day 86)
25	004	F	●	◦	◦	6	0 (day 166)
26	346	F	◦	◦	◦	6, 7	0 (day 56)
27	348	F	●	◦	◦	6	0 (day 103)

[1] ● = yes, ◦ = no

[2] Reason for exclusion from ITT

1 = Baseline lumbar spine scan performed on Lunar instead of Hologic

2 = Patient unevaluable due to scoliosis

3 = Baseline scan performed on Hologic 2000 and Month 6 performed on Hologic 4500. One of 2 machines sold before a cross-calibration between machines could be conducted.

4 = Metallic hardware present in spine at baseline or Month 6

5 = Baseline lumbar spine BMD data stored on optical disk could not be retrieved

6 = Discontinued before Month 6 lumbar spine BMD was performed

7 = Patient was improperly positioned during baseline scan and also had metal sutures in her spine at baseline. Patient also found to have fibrous dysplasia not OI.

Appendix 2. An imputation method for the 27 patients excluded by sponsor from analysis

Objective

Find the smallest effect size for the primary endpoint (lumbar spine BMD z-score) one could impute for the missing data (n=27) and still retain statistical significance for the all-randomized dataset (n=139).

Methods

Assume the mean response in the Fosamax missing cohort is smaller than the observed mean Fosamax response by an amount Δ, and the mean response in the placebo missing cohort is greater than or equal to than the observed mean placebo response by an amount $c\Delta$ ($0 \leq c \leq 1$). C is a proportionality factor that allows the mean response in the missing placebo cohort to depart from the observed response by a different magnitude than the corresponding means in the Fosamax group. The objective is to calculate Δ (i.e., the smallest effect size) for $c \in [0,1]$ such that the overall standardized treatment difference corresponds to a p-value of exactly .05

Let

observed data

y_{AO} = observed mean response in the Fosamax group
y_{PO} = observed mean response in the placebo group
sd_{AO} = observed standard deviation in the Fosamax group
sd_{PO} = observed standard deviation in the placebo group
n_{AO} = sample size in the observed Fosamax cohort
n_{PO} = sample size in the observed placebo cohort

missing data

$y_{AO} - \Delta$ = mean response in the Fosamax missing group [2]
$y_{PO} + c\Delta$ = mean response in the placebo missing group ($0 \leq c \leq 1$) [3]
n_{AM} = sample size in the missing Fosamax cohort
n_{PM} = sample size in the missing placebo cohort

all data

n_A = total randomized sample size in the Fosamax group
n_P = total randomized sample size in the placebo group

---

[2] Δ = mean decrement ($\Delta>0$) in the imputed Fosamax response for the missing cohort with respect to the observed data.

[3] $c\Delta$= mean increment in the imputed placebo response for the missing cohort with respect to the observed data ($0 \leq c \leq 1$).

The overall treatment difference $D(\Delta)$ equals:

$$D(\Delta) = (n_{AO}y_{AO} + n_{AM}(y_{AO}-\Delta))/ n_A - (n_{PO}y_{PO} + n_{PM}(y_{PO}+c\Delta))/ n_P$$

$$= (y_{AO} - y_{PO}) - \Delta (n_{AM}/ n_{A+} c\, n_{PM} / n_P)$$

The SE of $D(\Delta)$ is the usual 2-sample or pooled estimate using the observed SD sd_{AO} and sd_{PO} and the total randomized sample sizes n_A and $n_{P..}$ The calculation of SE(D) involves the implicit (and unverifiable but reasonable) assumption that the SD's of the missing data are the same as those for the observed data. Also note that SE(D) is independent of Δ:

$$SE(D) = [(1/n_A +1/n_P) ((n_A – 1)sd_{AO}^2 + (n_P – 1)sd_{PO}^2) / (n_A+ n_A - 2)]^{1/2}$$

Set the standardized treatment difference $D(\Delta) / SE(D)$ equal to $Z_{alpha/2}$ and solve for Δ:

$$Z_{alpha/2} = D(\Delta) / SE(D)$$
$$= [(y_{AO} - y_{PO}) - \Delta (n_{AM}/ n_{A+} c\, n_{PM} / n_P)] / SE(D)$$

so

$$\Delta = [(y_{AO} - y_{PO}) – Z_{alpha/2}\, SE(D)] / (n_{AM}/ n_A + c\, n_{PM}/ n_P]$$

Therefore, the missing cohort could have a treatment difference $D^*(\Delta)$ equal to

$$D^*(\Delta) = (y_{AO} - \Delta) -- (y_{PO} + c\Delta)$$
$$= y_{AO} – y_{PO} – (c+1)\Delta$$

and still maintain a nominally statistically significant difference between groups (p=.05) for the all-randomized dataset.

The value of Δ is chosen to give the smallest overall effect size such that the standardized treatment difference corresponds to a p-value of precisely .05. We chose c in [0,1] so that $D^*(\Delta)$ is maximized. This value of c gives the value of $D^*(\Delta)$ that is closest to the observed treatment difference and so is a conservative estimate of how small the effect in the missing cohort could be and still yield an overall significant result. The value of $c \in [0,1]$ that maximizes $D^*(\Delta)$ depends on the relative values of n_{AM}/ n_A and n_{PM}/ n_P. If $n_{AM}/ n_A = n_{PM}/ n_P$, then $D^*(\Delta)$ is independent of the value of c. If $n_{AM}/ n_A > n_{PM}/ n_P$, then c=0 gives the

maximum value for $D^*(\Delta)$. If $n_{AM}/ n_A < n_{PM}/ n_P$, then c=1 gives the maximum value for $D^*(\Delta)$.

Results

	Sample sizes			LS mean	SD of LS mean [1]
	Observed	Missing	Total		
Fosamax	$n_{AO} = 86$	$n_{AM} = 23$	$n_A = 109$	$y_{AO} = +1.02$	$Sd_{AO} = 0.93$
Placebo	$n_{PO} = 26$	$n_{PM} = 4$	$n_P = 30$	$y_{PO} = + 0.10$	$Sd_{AO} = 0.71$

[1] $SD = SE * n^{1/2}$ where SE is generated by SAS

$SE(D) = ((1/109 + 1/30) (108 (0.93)^2 + 29 (0.71)^2) / 137))^{1/2}$
$= 0.18$

Use c = 0 since 23/109 > 4/30

$\Delta = [(1.02 - 0.10) – 1.96(0.18)] / [23/109]$
$= 2.66$

$y_{AO} - \Delta = 1.02 – 2.66$
$= - 1.64$

$y_{PO} + (0)\Delta = 0.10$

$D^*(\Delta) = - 1.64 – 0.10$
$= -1.74$ (favoring placebo)

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Todd Sahlroot
7/21/03 09:55:34 AM
BIOMETRICS

S. Edward Nevius
7/21/03 10:10:21 AM
BIOMETRICS
Concur with review.