Appendix 16.2.3 presents a by-patient listing of all patients with significant protocol violations and *has been reviewed.*

11. Efficacy Evaluation

11.1. Data Sets Analyzed

- The **primary analyses** were performed on an **intention-to-treat (ITT)** basis.
  - An ITT analysis is an analysis of data by the treatment groups to which patients were assigned by random allocation, even if the patient did not take the assigned treatment, did not receive the correct treatment, or otherwise did not follow the protocol.
- The **primary dataset** contains data for all patients randomly assigned to treatment with at least one baseline and one post-baseline measurement.
- **Missing values** were imputed using the last observation carried forward (LOCF).
- For **analyses of change** (or percentage change) from baseline to endpoint, **baseline observations** were defined as the last measurement from the enrollment visit through the randomization visit. **Endpoint observations** for these patients were defined as the last post-baseline measurement prior to, and including, the 48-month visit.
- All **breast cancer analyses** compare the **average annual event rate** (number of patients reporting breast cancer for a given therapy divided by the number of event-specific person-years of follow-up for patients assigned to that therapy) in **raloxifene-treated patients** with the average annual event rate in **placebo-treated patients.**
  - Section 11.4.3.4.1 identifies the cases of breast cancer **excluded** from the analyses.
- Three **interim analyses** were conducted after 6, 12, and 24 months of follow-up. These interim analyses were performed under the auspices of a Data Monitoring Board (DMB) according to the specifications set forth in the protocol (see Appendix 16.1.1).
- On 20 November 1999, the final reporting database was validated and locked.

11.2. Demographic and Other Baseline Characteristics

Table GGGK.11.1 (next page) contains a summary of baseline characteristics for all randomly assigned patients in the study. Randomization resulted in well-balanced study groups. The demographic and other baseline characteristics of the patients did not differ significantly among the three treatment groups at baseline, with the exception of height (p=0.021). The maximum difference among the three treatment groups in mean height was 0.45 cm. The difference in height among the three treatment groups was unlikely to be clinically relevant. The three groups were similar with respect to family history of breast cancer in the patient's first-generation family (mother, sisters, daughters; p=0.816).

Table GGGK.11.1. Patient Demographics (Treatment Groups at Baseline, All Randomized Patients)

Variable	PLACEBO (N=2576)	RLX060 (N=2557)	RLX120 (N=2572)	Total (N=7705)	p-Value
ORIGIN					
No. Patients	2576	2557	2572	7705	.228*
African Descent	6 (0.2)	6 (0.2)	14 (0.5)	26 (0.3)	
Western Asian	6 (0.2)	1 (0.0)	4 (0.2)	11 (0.1)	
Caucasian	2465 (95.7)	2455 (96.0)	2452 (95.3)	7372 (95.7)	
East/Southeast A	48 (1.9)	41 (1.6)	48 (1.9)	137 (1.8)	
Hispanic	44 (1.7)	48 (1.9)	41 (1.6)	133 (1.7)	
Other	7 (0.3)	6 (0.2)	13 (0.5)	26 (0.3)	
AGE: (yrs)					
No. Patients	2576	2557	2572	7705	.337**
Mean	66.60	66.48	66.31	66.47	
Median	66.92	66.86	66.73	66.85	
Standard Dev.	7.07	6.99	7.12	7.06	
Minimum	35.68	31.08	35.99	31.08	
Maximum	80.96	80.94	80.91	80.96	
HEIGHT: (cm) (VISIT: 1)					
No. Patients	2575	2557	2571	7703	.020**
Mean	158.95	158.92	159.38	159.08	
Median	159.00	159.00	159.51	159.10	
Standard Dev.	6.57	6.60	6.68	6.62	
Minimum	133.00	127.00	123.95	123.95	
Maximum	185.00	192.20	178.00	192.20	
Unspecified	0	0	1	1	
WEIGHT: (kg) (VISIT: 1)					
No. Patients	2575	2556	2572	7703	.364**
Mean	63.64	63.58	63.96	63.73	
Median	62.88	62.40	63.00	62.88	
Standard Dev.	10.52	10.35	10.73	10.53	
Minimum	33.82	34.00	35.30	33.82	
Maximum	131.21	111.00	130.75	131.21	
Unspecified	0	1	0	1	
BMI: (kg/m2) (VISIT: 2)					
No. Patients	2575	2557	2571	7703	.989**
Mean	25.24	25.23	25.22	25.23	
Median	24.82	24.66	24.78	24.77	
Standard Dev.	3.99	4.02	4.02	4.01	
Minimum	13.54	14.22	14.45	13.54	
Maximum	51.59	43.16	49.56	51.59	
Unspecified	1	0	1	2	
CURRENT SMOKER (VISIT: 2)					
No. Patients	2576	2557	2572	7705	.915*
No	2124 (83.5)	2102 (83.1)	2112 (83.2)	6338 (83.3)	
Yes	420 (16.5)	429 (16.9)	425 (16.8)	1274 (16.7)	
Unspecified	32	26	35	93	
ALCOHOL > 3 DRINKS WKLY (VISIT: 2)					
No. Patients	2576	2557	2572	7705	.606*
No	2132 (82.8)	2089 (81.7)	2134 (83.0)	6355 (82.5)	
Unknown	4 (0.2)	2 (0.1)	2 (0.1)	8 (0.1)	
Yes	440 (17.1)	466 (18.2)	436 (17.0)	1342 (17.4)	
YEARS PMP (VISIT: 1)					
No. Patients	2576	2557	2572	7705	.262**
Mean	18.89	18.76	18.51	18.72	
Median	19.00	19.00	18.00	19.00	
Standard Dev.	8.48	8.51	8.30	8.43	
Minimum	2.00	2.00	2.00	2.00	
Maximum	54.00	67.00	48.00	67.00	

Variable	PLACEBO (N=2576)	RLX060 (N=2557)	RLX120 (N=2572)	Total (N=7705)	p-Value
FAM. HIST. OF OSTPRS (VISIT: 1)					
No. Patients	2576	2557	2572	7705	.930*
No	1595 (61.9)	1561 (61.0)	1571 (61.1)	4727 (61.3)	
Unknown	299 (11.6)	304 (11.9)	295 (11.5)	898 (11.7)	
Yes	682 (26.5)	692 (27.1)	706 (27.4)	2080 (27.0)	
FAM. HIST. OF BREAST CANCER (VISIT: 1)					
No. Patients	2576	2557	2572	7705	.814*
No	2196 (85.2)	2190 (85.6)	2183 (84.9)	6569 (85.3)	
Unknown	67 (2.6)	55 (2.2)	65 (2.5)	187 (2.4)	
Yes	313 (12.2)	312 (12.2)	324 (12.6)	949 (12.3)	
HYSTERECTOMY (VISIT: 1)					
No. Patients	2576	2557	2572	7705	.252*
No	1999 (77.6)	1950 (76.3)	2010 (78.1)	5959 (77.3)	
Yes	577 (22.4)	607 (23.7)	562 (21.9)	1746 (22.7)	
TYPE OF HYSTERECTOMY (VISIT: 1)					
No. Patients	2576	2557	2572	7705	.968*
Unknown	47 (8.1)	46 (7.6)	43 (7.7)	136 (7.8)	
Uterus,0-1 Ovary	278 (48.2)	305 (50.2)	277 (49.3)	860 (49.3)	
Uterus,2 Ovaries	252 (43.7)	256 (42.2)	242 (43.1)	750 (43.0)	
Unspecified	1999	1950	2010	5959	
PREV USE OF HRT (VISIT: 1)					
No. Patients	2576	2557	2572	7705	.567*
No	1833 (71.2)	1785 (69.8)	1829 (71.1)	5447 (70.7)	
Unknown	5 (0.2)	10 (0.4)	8 (0.3)	23 (0.3)	
Yes	738 (28.6)	762 (29.8)	735 (28.6)	2235 (29.0)	
PREV USE OF THIAZ DIURETICS (VISIT: 1)					
No. Patients	2576	2557	2572	7705	.174*
No	2241 (87.0)	2224 (87.0)	2249 (87.4)	6714 (87.1)	
Unknown	24 (0.9)	14 (0.5)	29 (1.1)	67 (0.9)	
Yes	311 (12.1)	319 (12.5)	294 (11.4)	924 (12.0)	
PREV USE OF SYSTEMIC FLUORIDES (VISIT: 1)					
No. Patients	2576	2557	2572	7705	.847*
No	2531 (98.3)	2506 (98.0)	2523 (98.1)	7560 (98.1)	
Unknown	4 (0.2)	4 (0.2)	2 (0.1)	10 (0.1)	
Yes	41 (1.6)	47 (1.8)	47 (1.8)	135 (1.8)	
PREV USE OF BISPHOSPHONATES (VISIT: 1)					
No. Patients	2576	2557	2572	7705	.072*
No	2522 (97.9)	2482 (97.1)	2504 (97.4)	7508 (97.4)	
Unknown	1 (0.0)	7 (0.3)	2 (0.1)	10 (0.1)	
Yes	53 (2.1)	68 (2.7)	66 (2.6)	187 (2.4)	
MARITAL STATUS (VISIT: 2)					
No. Patients	2576	2557	2572	7705	.599*
Divorced	241 (9.4)	234 (9.2)	249 (9.7)	724 (9.4)	
Married	1522 (59.3)	1543 (60.5)	1549 (60.4)	4614 (60.1)	
Never Married	137 (5.3)	139 (5.5)	125 (4.9)	401 (5.2)	
Separated	50 (1.9)	31 (1.2)	43 (1.7)	124 (1.6)	
Single but livin	0	0	1 (0.0)	1 (0.0)	
Widowed	615 (24.0)	602 (23.6)	596 (23.3)	1813 (23.6)	
Unspecified	11	8	9	28	

Variable	PLACEBO (N=2576)	RLX060 (N=2557)	RLX120 (N=2572)	Total (N=7705)	p-Value
YEARS OF EDUCATION (VISIT: 2)					
No. Patients	2546	2530	2547	7623	.522**
Mean	11.82	11.78	11.90	11.84	
Median	12.00	12.00	12.00	12.00	
Standard Dev.	3.89	3.92	3.96	3.92	
Minimum	0.00	0.00	0.00	0.00	
Maximum	26.00	25.00	40.00	40.00	
Unspecified	30	27	25	82	
PRIOR AWARENESS OF OSTEOPOROSIS (VISIT: 2)					
No. Patients	2576	2557	2572	7705	.511*
Yes	980 (38.0)	937 (36.6)	945 (36.7)	2862 (37.1)	
No	1596 (62.0)	1620 (63.4)	1627 (63.3)	4843 (62.9)	

SOURCE IS RMP.H3SP.SASMACRO(DHSM1) DE005 01C
DATA FROM RMP.SAS.H3SM.MCGGGKSC.FINAL
* Frequencies are analyzed using a Chi-Square test.
** Means are analyzed using a Type III Sum of Squares analysis of variance (ANOVA): PROC GLM model=treatment.
XDES0001

Baseline Breast Images

- The population of women in this study was not selected based on a high risk of breast cancer.
- Baseline breast images were collected and classified as either normal or abnormal by the investigator. Abnormal breast images were further classified by the investigator as either clinically relevant or not clinically relevant.
  - At baseline, there were no statistically significant, treatment-group differences in the proportion of patients with normal, abnormal and not clinically relevant, or abnormal and clinically relevant breast imaging.

Table GGGK.11.2. Breast Imaging Results (All Randomly Assigned Patients, 48-Month Data)

	Placebo (N=2576)	RLX060 (N=2557)	RLX120 (N=2572)	p-value[a]
Baseline Breast Imaging Results[b]				
Normal	1864 (72.4%)	1864 (72.9%)	1904 (74.0%)	0.387
Abnormal, Not Clinically Relevant	671 (26.1%)	653 (25.5%)	628 (24.4%)	0.388
Abnormal, Clinically Relevant	38 (1.5%)	40 (1.6%)	39 (1.5%)	0.966
Any Abnormal Result	709 (27.5%)	693 (27.1%)	667 (25.9%)	0.411

[a] Chi-square test for total count ≥10; Fisher's Exact test for total counts 5 through 9.
[b] Patients with more than one baseline breast image were classified according to their most severe result.
Abbreviations: N = number of randomly assigned patients; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day.

Baseline characteristics of patients in sub-study I and II

Table GGGK.11.3 and Table GGGK.11.4 (not reproduced in this review) contain summaries of baseline characteristics for patients in Substudy I and Substudy II, respectively, which were conducted simultaneously to evaluate the primary endpoints.

- Within each substudy, randomization resulted in well-balanced treatment groups.
- The only statistically significant differences found among the three treatment groups were the proportion of hysterectomized patients (p=0.044) in Substudy I and height (p=0.023) in Substudy II; however, the magnitude of the difference among the three treatment groups was unlikely to be clinically relevant.
  - The maximum difference among the three treatment groups in the number of hysterectomized patients in Substudy I was 51 hysterectomized patients.
  - The maximum difference in mean height among the three treatment groups in Substudy II was 0.81 cm.

11.3. Measurements of Treatment Compliance

Study drug compliance was calculated from an accounting of returned medication. This differs from **exposure to study drug**, calculated from date of randomization to last known date of treatment, which measures theoretical maximum exposure to study drug.
This study followed ITT principles. Patients were permitted to remain in the study as long as they were compliant with study procedures, regardless of whether they continued to take study medication.

- A patient was defined as **severely noncompliant** with study drug if the number of tablets returned indicated that the patient was taking less than 70% of the study medication doses during at least two visit intervals (not necessarily contiguous).
- No statistically significant differences among the three treatment groups with respect to severe noncompliance with study drug were identified.

Counts of returned tablets could underestimate study drug noncompliance because, for example, "0" was used to record that all study drug was taken as well as to indicate that the patient failed to return any unused study drug to the site.

Only 2 patients were found to have taken a different dose than initially assigned at Visit 2.

- Patient 510-6997 was assigned to the raloxifene 60-mg group, but was dispensed raloxifene 120 mg at Visits 3 and 4. This was corrected at Visit 5, after which time the patient returned to her initially assigned dose for the remaining visits.
- Patient 085-6623 was assigned to the placebo group, but was dispensed raloxifene 120 mg at Visit 9. The patient was dispensed the correct study material 6 days after Visit 9.

Table GGGK.11.5. Patient Compliance to Treatment (All Randomized Patients, 48-Month Data)

Compliance Definition (Percentage of Study Medication Taken)	Percentage of Patients Who Were Compliant				
	Placebo (N=2576)	RLX060 (N=2557)	RLX120 (N=2572)	Total (N=7705)	p-value
Severe Noncompliance[a]	7.1%	6.8%	6.8%	6.9%	0.899
70%	96.2%	96.5%	96.5%	96.4%	0.735
75%	95.1%	95.3%	95.3%	95.2%	0.942
80%	93.3%	93.4%	93.0%	93.2%	0.848
85%	89.8%	90.1%	88.5%	89.5%	0.135
90%	83.4%	83.7%	81.7%	82.9%	0.124

Abbreviations: RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day;
N = number of randomly assigned patients.
[a] Severe noncompliance is defined as taking less than 70% of study medication during at least two visit intervals.

11.4. Efficacy Results and Tabulations of Individual Patient Data

11.4.1. Analysis of Efficacy

Breast cancer incidence was protocol-specified as a secondary endpoint, and the protocol was designed to systematically screen out preexisting breast cancers at baseline by physical examination and mammograms performed within 12 months of randomization.

- The protocol also allowed for the prospective ascertainment of breast cancer incidence through the conduct of optional mammography at year 1, along with protocol-mandated mammography at years 2, 3, and 4.

The 48-month data for the primary endpoints, which included rate of new vertebral fractures, lumbar spine and femoral neck BMD, and safety, will be presented overall and by substudy. Other secondary endpoints will also be briefly discussed.

11.4.2. Statistical/Analytical Issues

Analyses of breast cancer data are explained in the next section.

11.4.3. Breast Cancer

Eighty-two cases of primary breast carcinoma were reported to the sponsor.

- One cancer (in a raloxifene-treated patient) was assigned a diagnosis date by the investigator that was 4 days prior to her randomization date; because this patient was randomly assigned to study drug, she was included in the analyses.
- Three breast carcinoma cases were reported after the end of the study and submitted for adjudication. One of these cases has since been adjudicated and is included in these analyses.

By 48 months, raloxifene use was associated with a reduction in the incidence of breast cancer:

- The reported incidence of invasive and noninvasive breast cancers in the pooled raloxifene group was reduced by 62% compared with placebo.
- This reduction in breast cancer risk was highly statistically significant (95% confidence interval 39% to 76%).

11.4.3.1. Breast Imaging

Mammograms were required for all patients at baseline and at the 24-, 36-, and 48-month visits; mammograms were optional at the 12-month visit. If mammography was not acceptable to a patient, ultrasonography of the breast was performed instead, although patients were encouraged to undergo mammography.

Table GGGK.11.6 shows the number of eligible patients who underwent breast imaging at the baseline, 12-, 24-, 36-, and 48-month visits. For each visit interval, a patient was defined to be "eligible" for breast imaging if that patient was continuing in the study at the beginning of the visit interval (for example, a patient who had not discontinued by the 12-month visit was considered eligible for a 24-month breast image).

- Breast imaging was performed in 99.97% of patients at baseline, and 48% of women elected to have an optional breast imaging procedure at the 12-month visit.
- At the 24-, 36-, and 48-month visits, 94%, 91%, and 93% of participants continuing in the study, respectively, had breast imaging procedures performed.

- For all randomly assigned patients, there were no differences among the three treatment groups in the number of patients who had breast imaging at baseline, 12, 24, and 48 months.
- Among those patients who underwent breast imaging, there was no difference among the three treatment groups in the number of patients who elected sonography instead of mammography at baseline or at any visit.
- Among those patients eligible for breast imaging at 36-months, there was a statistically significant difference among the three treatment groups in the number of patients who had breast imaging ($p=0.004$). Although the data indicate that a slightly smaller proportion of patients in the placebo group underwent breast imaging, this difference would most likely result in an underestimate of the true rate of abnormal mammographic findings in the placebo group.

Table GGGK.11.6. Distribution of Breast Images (All Randomly Assigned Patients, 48-Month Data)

	Placebo (N=2576)	RLX060 (N=2557)	RLX120 (N=2572)	p-value[a]
Baseline Visit				
Eligible Patients[b]	2576	2557	2572	—
Patients With Breast Images	2574	2557	2571	0.778
Mammogram[c]	2500	2486	2496	—
Sonogram Only	74	71	75	0.951
12-Month Visit[d]				
Eligible Patients[b]	2576	2557	2572	—
Patients With Breast Images	1249	1234	1244	0.986
Mammogram[c]	1198	1184	1184	—
Sonogram Only	51	50	60	0.574
24-Month Visit[d]				
Eligible Patients[b]	2339	2283	2311	—
Patients With Breast Images	2176	2163	2171	0.052
Mammogram[c]	2113	2099	2121	—
Sonogram Only	63	64	50	0.328
36-Month Visit[d]				
Eligible Patients[b]	2109	2124	2148	—
Patients With Breast Images	1894	1957	1986	0.004
Mammogram[c]	1839	1899	1923	—
Sonogram Only	55	58	63	0.809
48-Month Visit[d]				
Eligible Patients[b]	1920	1964	1997	—
Patients With Breast Images	1779	1840	1861	0.444
Mammogram[c]	1731	1788	1811	—
Sonogram Only	48	52	50	0.961

[a] The p-value for "Breast Images" compares the three treatment groups with respect to the number of eligible patients who had any breast imaging during a particular visit interval. The p-value for "Sonogram Only" compares the three treatment groups with respect to the number of patients with breast imaging who only had a breast sonogram during a particular visit interval. The p-value is calculated using Fisher's Exact test, since the proportion of patients without images is very small in some cases.

[b] Eligible patients are defined as those who were continuing in the study at the beginning of the visit interval (e.g., a patient who had not discontinued by the 12-month visit was considered eligible for a 24-month breast image).

[c] Patients who had multiple breast images during any visit interval were classified as having mammography if any of the images were mammograms, otherwise, they were classified as having only sonography.

[d] Mammogram or sonogram results recorded at Visit 3 (3 months), Visit 4 (6 months), or Visit 5 (12 months) were considered 12-month breast images. Those results recorded at Visit 6 (18 months) or Visit 7 (24 months) were considered 24-month breast images. Those results recorded at Visit 8 (30 months) or Visit 9 (36 months) were considered 36-month breast images. Those results recorded at Visit 10 (42 months) or Visit 11 or 12 (48 months) were considered 48-month breast images.

Abbreviations: N = number of randomly assigned patients; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day.

For the **analyses of breast imaging**, patients with more than one post baseline breast image were **classified according to their most "severe" result**, with the result of **"abnormal, clinically relevant"** being classified as more severe than the result of **"abnormal, not clinically relevant,"** which was classified as more severe than the result of **"normal".**

By the 48-month visit, there was a statistically significant difference among the three treatment groups in the **proportion of patients with *abnormal, clinically relevant* breast imaging** (p=0.001), with a higher proportion of patients in the placebo group reporting this result:

- 120 (4.7%) patients in the placebo group
- 81 (3.2%) patients in the raloxifene 60-mg group
- 73 (2.8%) patients in the raloxifene 120-mg group

There were no statistically significant differences in ***any other* category of breast imaging** results by the 48-month visit.

For those patients with ***normal baseline* breast images**, there were no statistically significant differences in the other categories of breast imaging by the 48-month visit.

Table GGGK.11.7. Post-baseline Breast Imaging Results (All Randomly Assigned Patients, 48-Month Data)

	Placebo (N=2576)	RLX060 (N=2557)	RLX120 (N=2572)	p-value[a]
Postbaseline Breast Imaging Results[b]				
Normal	1454 (56.4%)	1453 (56.8%)	1489 (57.9%)	0.553
Abnormal, Not Clinically Relevant	731 (28.4%)	745 (29.1%)	721 (28.0%)	0.670
Abnormal, Clinically Relevant	120 (4.7%)	81 (3.2%)	73 (2.8%)	0.001
Any Abnormal Result[c]	851 (33.0%)	826 (32.3%)	794 (30.9%)	0.239
Postbaseline Breast Imaging Results for Patients with Normal Baseline Breast Images[b]				
Normal	1370 (73.5%)	1383 (74.2%)	1417 (74.4%)	0.797
Abnormal, Not Clinically Relevant	228 (12.2%)	224 (12.0%)	237 (12.5%)	0.922
Abnormal, Clinically Relevant	62 (3.3%)	37 (2.0%)	28 (1.5%)	0.000
Any Abnormal Result[c]	290 (15.6%)	261 (14.0%)	265 (13.9%)	0.276

[a] Chi-square test for total count ≥10; Fisher's Exact test for total counts 5 through 9.

[b] Patients with more than one postbaseline breast image were classified according to their most severe result. All patients who had postbaseline breast imaging at or before the 48-month visit are included in this analysis.

[c] Includes pooled categories "Abnormal, Not Clinically Relevant" and "Abnormal, Clinically Relevant."

Abbreviations: N = number of randomly assigned patients; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day.

11.4.3.2. Breast Carcinoma Adjudication Process and Results

An Adjudication Review Board (ARB) consisting of five physicians specializing in breast cancer, and chaired by a basic scientist with expertise in SERMs and breast cancer, was selected by the sponsor to adjudicate each reported case of breast cancer. The board was blinded to treatment-group assignment. For each reported case of breast cancer, the ARB was presented with as much of the following information as was available to the sponsor:

- Mammographic and other relevant radiologic reports
- Mammographic films (originals or copies)
- Estrogen receptor status
- Pathologic reports from biopsy and or surgical specimens

For each case, the ARB adjudicated:

1) Whether the case was **invasive** primary breast cancer?
2) What was the **estrogen-receptor status** (estrogen receptor-positive [ER+] or estrogen receptor-negative [ER-])?
3) Whether the cancer was **preexisting** (ie, present at the baseline visit) or **new** (occurring after the baseline visit)?

The analyses in this section summarize the findings of the ARB. The data analyzed in this section included all adjudicated cases.

It should be noted that the statistical analyses of the adjudicated breast cancer data were not prospectively defined as an efficacy endpoint in the protocol; however, safety analyses of breast cancer data showed a reduction in the incidence of breast cancer in raloxifene-treated patients compared with patients who received placebo. Thus, further analyses to determine the effect of raloxifene on the breast were conducted and are presented in this report.

Patients were asked at each visit whether they had been diagnosed with breast cancer since the previous visit. If patients reported having been diagnosed with breast cancer, evidence of breast cancer was obtained from all patients. This evidence included results of all scheduled and unscheduled breast imaging, surgical and pathology reports from breast biopsies or needle aspirations, and the documentation of tumor characteristics if a tumor was identified. All test results were recorded on the Clinical Assessment of Breast Cancer (CABC) form.

11.4.3.4. Breast Cancer Data Analyses

11.4.3.4.1. General Considerations

Event-specific patient-years of follow-up were calculated based on the following algorithm: For patients who did not experience the breast cancer event of interest, patient-year contribution was calculated as the date of final patient contact minus the randomization date plus 1 day.

For patients who experienced the event of interest, patient-year contribution was calculated as the date of the breast cancer event minus the randomization date plus 1 day.
One day was added to each patient-year contribution because patients were instructed to begin therapy on the day of randomization.

Of the 82 reported cases of breast carcinoma in Study GGGK, 4 cases were excluded from the analyses. The following is a listing of the reasons for the exclusions:
Case 081-6018 was adjudicated "metastatic adenocarcinoma of unknown primary"
Case 085-6480 was adjudicated "No cancer"
Cases 086-7631 and **068-6961** had not been adjudicated as of 22 May 2002 because required adjudication documents had not been submitted to the sponsor by the investigative sites.

11.4.3.4.2. Breast Cancer Incidence

The estimated incidence rates of breast cancer and invasive breast cancer are presented in (Table GGGK.11.8).

- Annual incidence rates of breast cancer and invasive breast cancer were much lower in patients assigned to raloxifene than in patients assigned to placebo.
- Neither the incidence of breast cancer nor invasive breast cancer was significantly different between the raloxifene 60- and 120-mg treatment groups (p=0.986 and p=0.817 respectively).
- Because treatment effects in these two groups were similar the raloxifene groups are pooled for all further analyses.

Table GGGK.11.8. Estimated Annual Incidence Rates for Breast Cancer and Invasive Breast Cancer (All Randomly Assigned Patients, 48-Month Data)

Population	Therapy	No. Randomized	Cases	Patient-years of Follow-up	Rate (per 1000)
Breast Cancer	Placebo	2576	44	8716	5.05
	RLX060		17	8756	1.94
	RLX120		17	8868	1.92
	Pooled Ralox	5129	34	17624	1.93
Invasive Breast Cancer	Placebo	2576	38	8718	4.36
	RLX060		11	8756	1.26
	RLX120		10	8869	1.13
	Pooled Ralox	5129	21	17625	1.19

Abbreviations: No. = number; Ralox = raloxifene.

11.4.3.4.2. Relative Risk

The results shown in Table GGGK.11.9 demonstrate a 62% reduction in breast cancer incidence for raloxifene-treated women compared with the placebo group.

- For **only invasive tumors**, the reduction was 73%
- For **all ER+ tumors**, the reduction was 79%
  - For the subset of **invasive ER+ tumors**, raloxifene demonstrated an 83% reduction in the incidence of breast cancer compared with placebo
- Considering **only ER- tumors**, there was not a statistically significant difference between treatment groups.
- For cases of **unknown estrogen receptor (ER) status**, a non-significant reduction in relative risk was observed.

Table GGGK.11.9. Breast Cancer Relative Risk Analysis of all Cases (All Randomly Assigned Patients, 48-Month Data)

Category	Number of Cases		Rate per 1000 Women		Risk Ratio
	Placebo	Raloxifene	Placebo	Raloxifene	(95% CI)
All cases	44	34	5.05	1.93	0.38 (0.24, 0.61)
Invasive cases	38	21	4.36	1.19	0.27 (0.15, 0.48)
ER-positive cases	31	13	3.56	0.74	0.21 (0.10, 0.41)
Invasive ER-positive cases	29	10	3.33	0.57	0.17 (0.07, 0.36)
ER-negative cases	4	10	0.46	0.57	1.24 (0.36, 5.40)
Cases of unknown ER status	9	11	1.03	0.62	0.60 (0.23, 1.65)

Abbreviations: ER = estrogen receptor; CI = confidence interval.

Table GGGK.11.10 presents a secondary analysis of the relative risk reductions for raloxifene-treated patients compared with placebo for breast cancers that were adjudicated as non-preexisting at study entry.

- In this subset, raloxifene reduced all cancers by 70%, invasive cancers by 81%, and invasive ER+ cancers by 87%.

Table GGGK.11.10. Breast Cancer Relative Risk Analysis Cases Adjudicated as Non-Preexisting at Study Entry (All Randomly Assigned Patients, 48-Month Data)

Category	Number of Cases		Rate per 1000 Women		Risk Ratio
	Placebo	Raloxifene	Placebo	Raloxifene	(95% CI)
Adjudicated as non-preexisiting	33	20	3.78	1.13	0.30 (0.16, 0.54)
Adjudicated as non-preexisting invasive	29	11	3.33	0.62	0.19 (0.08, 0.39)
Adjudicated as non-preexisting invasive ER-positive	23	6	2.64	0.34	0.13 (0.04, 0.33)

Abbreviations: ER = estrogen receptor; CI = confidence interval.

11.4.3.4.3. Time-To-Event Analyses of All Reported Breast Carcinoma Cases

Kaplan-Meier analyses of percent of disease-free patients depict a continuous separation between placebo- and raloxifene-treated patients for all, invasive, and invasive ER+ breast cancer cases (Figure GGGK.11.1, Figure GGGK.11.2, and Figure GGGK.11.3).
The step-wise pattern apparent in each of the figures is due to the performance of mammograms at yearly intervals. A decrease in the relative risk of invasive breast cancer is evident by the second year of treatment ($p<.001$). These figures clearly demonstrate the sustained efficacy of raloxifene to reduce the incidence of breast cancer in postmenopausal women with osteoporosis.

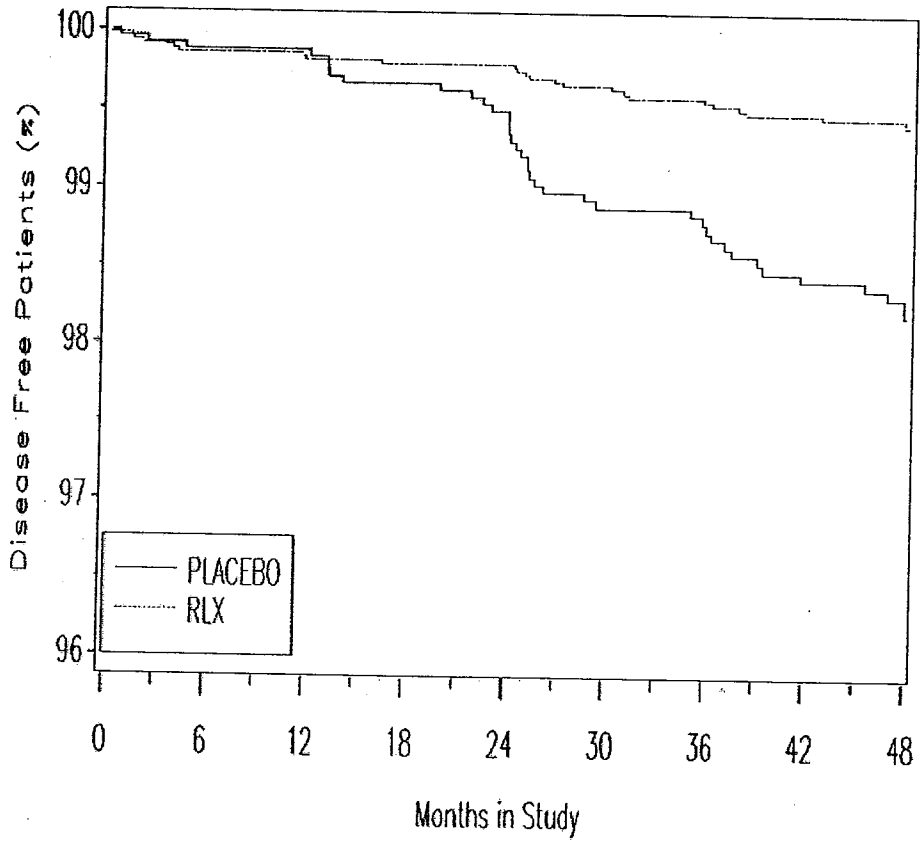

Figure GGGK.11.1. Percent of disease-free patients for all cases of breast cancer.

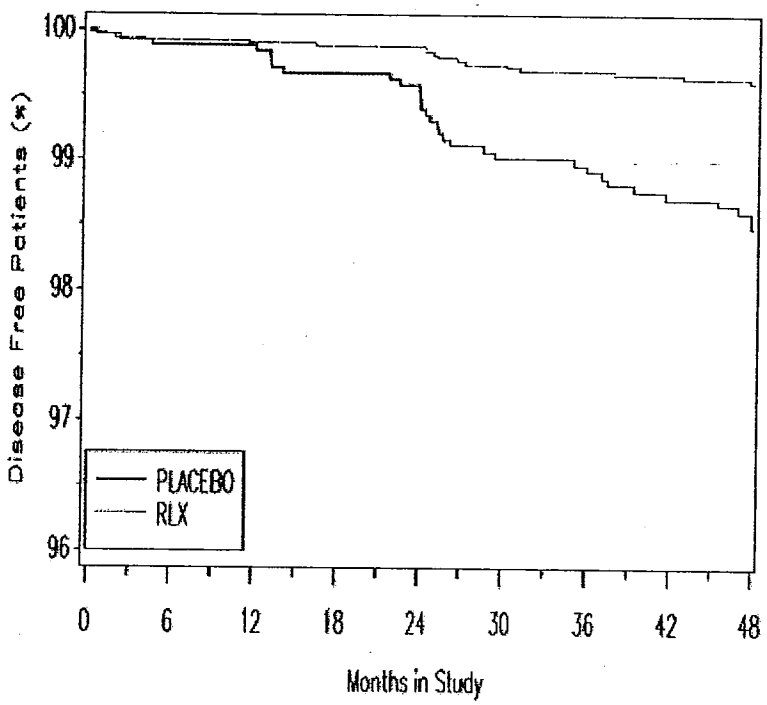

Figure GGGK.11.2. Percent of disease-free patients for cases of invasive breast cancer.

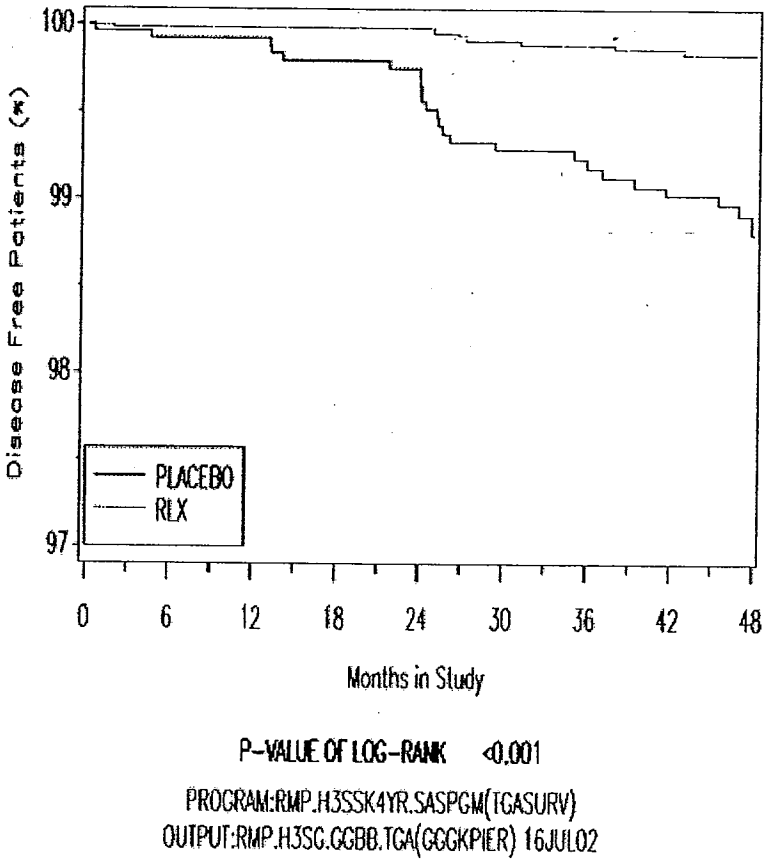

Figure GGGK.11.3. Percent of disease-free patients for cases of invasive ER+ breast cancer.

11.4.3.4.4. Number Needed to Treat (NNT)

The NNT with raloxifene to prevent one patient from developing breast cancer is calculated by taking the inverse of the difference in percentage of patients with breast cancer in the treated (21 cases in 5,129 patients, or 0.41%) and control (38 cases in 2,576 patients, or 1.48%) groups.

- Overall, 93 patients (1/[0.41-1.48]) would need to receive raloxifene to prevent one new case of invasive breast cancer.

11.4.4. Bone Efficacy

11.4.4.1. Fractures

11.4.4.1.1. Vertebral Fractures

The assessment of vertebral fractures is described in the protocol (Appendix 16.1.1). The analysis of vertebral fractures is presented in the following order:

- Results for women with adjudicated new vertebral fractures overall and by substudy
- Results after stratification by the presence or absence of prevalent fractures
- The number NTT to prevent a new vertebral fracture overall and by substudy
- Fracture incidence overall and by sub-study

11.4.4.1.1.1. New Vertebral Fractures

Table GGGK.11.11 summarizes the proportion of women and relative risk for having one or more adjudicated, new incident vertebral fractures during the trial for each dose of raloxifene and pooled raloxifene doses compared with placebo, along with 95% confidence intervals, for each substudy and for the entire study population.

- Each dose of raloxifene statistically significantly decreased the proportion of women with adjudicated, new incident vertebral fractures in each substudy and overall compared with the placebo group.
- Overall, there was a 36% reduction ($p<0.001$) in such fractures in the raloxifene 60-mg group and a 43% reduction ($p<0.001$) in the raloxifene 120-mg group compared with the placebo group.
- Overall, there was not a statistically significant difference between the two raloxifene groups in the proportion of patients with at least one new vertebral fracture.

Table GGGK.11.11. New Incident Vertebral Fracture Results Overall and by Sub-study (All Randomly Assigned Patients, 48-Month Data)

	Placebo	RLX060	RLX120	Pooled RLX Doses
Substudy I	n=1521	n=1492	n=1512	n=3004
Number of patients with ≥1 incident fracture (%)	97 (6.4%)	51 (3.4%)	57 (3.8%)	108 (3.6%)
Relative risk (95% CI) compared with placebo		0.54 (0.38, 0.75)	0.59 (0.43, 0.81)	0.56 (0.43, 0.74)
Pairwise comparison with placebo		p<0.001	p=0.001	p<0.001
Pairwise comparison with RLX060			p=0.605	
Substudy II	n=771	n=767	n=765	n=1532
Number of patients with ≥1 incident fracture (%)	191 (24.8%)	130 (16.9%)	107 (14.0%)	237 (15.5%)
Relative risk (95% CI) compared with placebo		0.68 (0.56, 0.83)	0.56 (0.46, 0.70)	0.62 (0.53, 0.74)
Pairwise comparison with placebo		p<0.001	p<0.001	p<0.001
Pairwise comparison with RLX060			p=0.109	
Pooled Substudies	n=2292	n=2259	n=2277	n=4536
Number of patients with ≥1 incident fracture (%)	288 (12.6%)	181 (8.0 %)	164 (7.2%)	345 (7.6%)
Relative risk (95% CI) compared with placebo		0.64 (0.53, 0.76)	0.57 (0.48, 0.69)	0.61 (0.52, 0.70)
Pairwise comparison with placebo		p<0.001	p<0.001	p<0.001
Pairwise comparison with RLX060			p=0.304	

Abbreviations: RLX = raloxifene; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; CI = confidence interval; n = number of patients with evaluable radiographs at baseline and endpoint.

11.4.4.1.1.2. New Vertebral Fracture Results After Stratification by the Presence or Absence of Prevalent Fractures

An analysis of new vertebral fracture incidence was performed after stratifying patients by the presence or absence of adjudicated prevalent fractures (Table GGGK.11.12).

- Each raloxifene dose group showed a statistically significant reduction in the proportion of patients with at least one new vertebral fracture compared with the placebo group.
- Reductions in the risks for the **first vertebral fractures** were 49% and 38% for raloxifene 60- and 120-mg groups, and reductions in the risks for **subsequent vertebral fractures** were 34% and 46% for the raloxifene 60- and 120-mg groups, respectively.

This analysis provided the same statistical inference and very similar estimates of relative risk for vertebral fracture reduction, as did the analysis by sub-study presented in (Table GGGK.11.11).

Table GGGK.11.12. New Vertebral Fracture Results by Presence or Absence of Prevalent Fracture (All Randomly Assigned Patients, 48-Month Data)

	Placebo	RLX060	RLX120	Pooled RLX Doses
Patients With No Prevalent Deformity	n=1459	n=1399	n=1411	n=2810
Number of patients with ≥1 incident fracture (%)	84 (5.8%)	41 (2.9%)	50 (3.5%)	91 (3.2%)
Relative risk (95% CI) compared with placebo		0.51 (0.35, 0.73)	0.62 (0.44, 0.87)	0.56 (0.42, 0.75)
Relative risk (95% CI) compared with RLX060			1.21(0.81, 1.82)	
Patients With ≥1 Prevalent Deformity	n=833	n=860	n=866	n=1726
Number of patients with ≥1 incident fracture (%)	204 (24.5%)	140 (16.3%)	114 (13.2%)	254 (14.7%)
Relative risk (95% CI) compared with placebo		0.66 (0.55, 0.81)	0.54 (0.44, 0.66)	0.60 (0.51, 0.71)
Relative risk (95% CI) compared with RLX060			0.81 (0.64, 1.02)	

Abbreviations: RLX = raloxifene; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; CI = confidence interval; n = number of patients at endpoint.

Figure GGGK.11.4 presents a Kaplan-Meier analysis of time to first vertebral fracture.

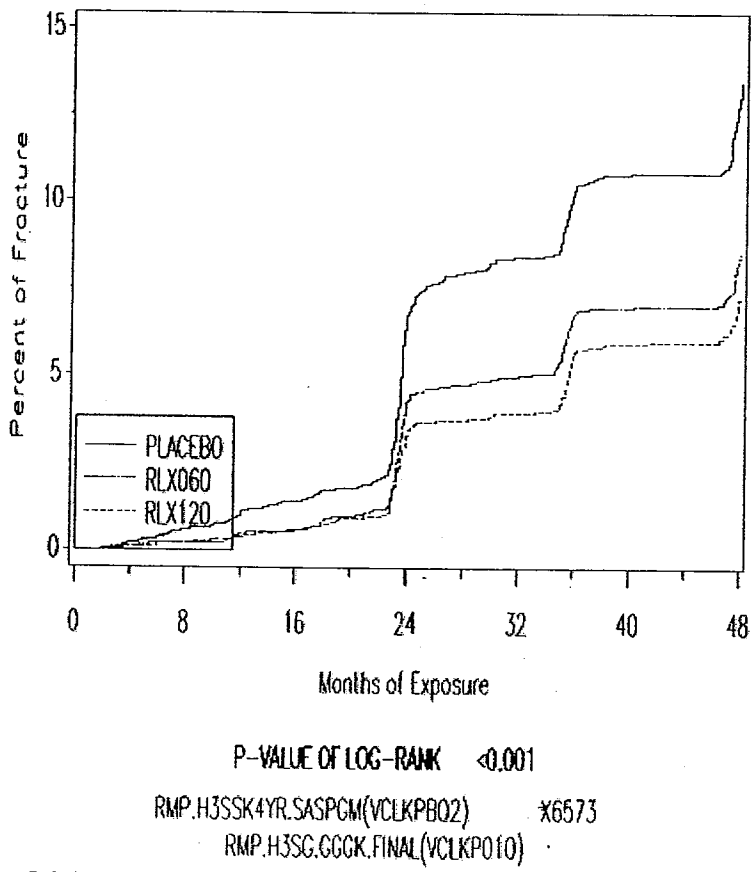

Figure GGGK.11.4. Kaplan-Meier Analysis: Time to First New Vertebral Fracture (All Randomly Assigned Patients, 48-Month Data)

11.4.4.1.1.3. Number Needed to Treat

The NNT with raloxifene to prevent a patient from having one or more new vertebral fractures is calculated by taking the inverse of the difference in percentage of patients with fractures in the treated and control group.

- Table GGGK.11.13 shows the number needed to treat overall and within each substudy for 48 months for each raloxifene group and for the pooled raloxifene group.
- Although the relative reduction in fracture rate is similar across the two substudies, fewer Substudy II patients need to be treated to prevent a new vertebral fracture because of the higher incidence of vertebral fractures in that subgroup.

Table GGGK.11.13. Number Needed to Treat With Raloxifene to Prevent a New Incident Vertebral Fracture (All Randomly Assigned Patients, 48-Month Data)

	Placebo	RLX060	RLX120	Pooled RLX Doses
Substudy I	n=1521	n=1492	n=1512	n=3004
Percentage of patients with ≥1 new incident fracture	6.4%	3.4%	3.8%	3.6%
Number needed to treat		33	38	36
Substudy II	n=771	n=767	n=765	n=1532
Percentage of patients with ≥1 new incident fracture	24.8%	16.9%	14.0%	15.5%
Number needed to treat		13	9	11
Pooled Substudies	n=2292	n=2259	n=2277	n=4536
Percentage of patients with ≥1 new incident fracture	12.6%	8.0%	7.2%	7.6%
Number needed to treat		22	19	20

Abbreviations: RLX = raloxifene; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; n = number of patients with evaluable radiographs at baseline and endpoint.

11.4.4.1.1.4. Clinical Vertebral Fractures

At each visit, whether a patient had experienced signs or symptoms suggestive of a vertebral fracture since the prior visit was to be recorded. Patients were considered to have had a clinical vertebral fracture if "yes" was recorded to this question and there was radiographic evidence of a new vertebral fracture at that visit (Table GGGK.11.14).

- Overall, 220 patients had at least one **new** clinical vertebral fracture.
- Compared with the placebo group, significantly fewer clinical fractures were reported in the patients assigned to both the raloxifene 60- and 120-mg groups and for patients in the pooled raloxifene group.
- The reduction in risk over 48 months ranged from 38% to 53% for the raloxifene 60-mg group, and from 48% to 61% for the raloxifene 120-mg group.
    - There was no difference in the proportion of patients reporting new clinical vertebral fractures in the raloxifene 120-mg group compared with the raloxifene 60-mg group.

Table GGGK.11.14. New Clinical Vertebral Fracture Results (All Randomly Assigned Patients, 48-Month Data)

	Placebo	RLX060	RLX120	Pooled RLX Doses
Pooled Substudies	N=2576	N=2557	N=2572	N=5129
Number of patients with ≥1 incident fracture (%)	107 (4.2%)	62 (2.4%)	51 (2.0%)	113 (2.2%)
Relative risk (95% CI)		0.58 (0.43, 0.79)	0.47 (0.34, 0.66)	0.53 (0.41, 0.69)
Pairwise comparison with placebo		p=0.001	p<0.001	p<0.001
Substudy I	N=1689	N=1672	N=1703	N=3375
Number of patients with ≥1 incident fracture (%)	28 (1.7%)	13 (0.8%)	11 (0.6%)	24 (0.7%)
Relative risk (95% CI)		0.47 (0.24, 0.90)	0.39 (0.19, 0.78)	0.43 (0.25, 0.74)
Pairwise comparison with placebo		p=0.02	p=0.006	p=0.002
Substudy II	N=887	N=885	N=869	N=1754
Number of patients with ≥1 incident fracture (%)	79 (8.9%)	49 (5.5%)	40 (4.6%)	89 (5.1%)
Relative risk (95% CI)		0.62 (0.44, 0.88)	0.52 (0.36, 0.75)	0.57 (0.43, 0.76)
Pairwise comparison with placebo		p=0.006	p<0.001	p<0.001

Abbreviations: RLX = raloxifene; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; CI = confidence interval; N = all randomly assigned patients.

11.4.4.1.2. Non-vertebral Fractures

A secondary objective of the trial was to establish the effects of raloxifene on the incidence of osteoporotic nonvertebral fractures, defined as a fracture at any of the following **sites:** clavicle, scapula, ribs, sternum, sacrum, coccyx, humerus, forearm, carpus, pelvis, femur, patella, tibia, fibula, ankle, calcaneus, tarsus, and metatarsus.
The following types of fractures were *excluded from analyses*: pathologic fractures, traumatic fractures (that is, fractures that are the result of a motor vehicle accident, a beating, or of being hit by a moving object), fractures of the skull, face, metacarpals, fingers, and toes. Sites were requested to confirm the fracture either by obtaining a radiologist's written report or by review of the radiograph.

- There were no statistically significant differences among the three treatment groups in the proportion of patients reporting at least one incident osteoporotic nonvertebral fracture (Table GGGK.11.15).
- A Kaplan-Meier analysis of time to first nonvertebral fracture is presented in (Figure GGGK.11.5).

Table GGGK.11.15. Osteoporotic Nonvertebral Fracture Results Overall and by Year (All Randomly Assigned Patients, 48-Month Data)

		Placebo (N=2576)	RLX060 (N=2557)	RLX120 (N=2572)
Overall (Year 0-4)	Number (proportion) reporting at least one incident fracture Relative risk (95% CI) compared with placebo Pairwise comparison with placebo	296 (11.5%)	290 (11.3%) 0.99 (0.85, 1.15) p=0.866	258 (10.0%) 0.87 (0.75, 1.02) p=0.091
Overall Substudy I	Number (proportion) reporting at least one incident fracture Relative risk (95% CI) compared with placebo Pairwise comparison with placebo	179 (10.6%)	170 (10.2%) 0.96 (0.79, 1.17) p=0.682	152 (8.9%) 0.84 (0.69, 1.03) p=0.101
Overall Substudy II	Number (proportion) reporting at least one incident fracture Relative risk (95% CI) compared with placebo Pairwise comparison with placebo	117 (13.2%)	120 (13.6%) 1.03 (0.81, 1.30) p=0.820	106 (12.2%) 0.92 (0.72, 1.18) p=0.532
Year 0-1	Number (proportion) reporting at least one incident fracture Relative risk (95% CI) compared with placebo Pairwise comparison with placebo	96 (3.7%)	107 (4.2%) 1.12 (0.857, 1.471) p=0.400	87 (3.4%) 0.91 (0.683, 1.207) p=0.505
Year 0-2	Number (proportion) reporting at least one incident fracture Relative risk (95% CI) compared with placebo Pairwise comparison with placebo	164 (6.4%)	158 (6.2%) 0.97 (0.786, 1.199) p=0.782	150 (5.8%) 0.92 (0.739, 1.135) p=0.423
Year 0-3	Number (proportion) reporting at least one incident fracture Relative risk (95% CI) compared with placebo Pairwise comparison with placebo	241 (9.4%)	228 (8.9%) 0.95 (0.802, 1.133) p=0.585	211 (8.2%) 0.88 (0.735, 1.046) p=0.144

Abbreviations: RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; CI = confidence interval; N = number of randomly assigned patients.

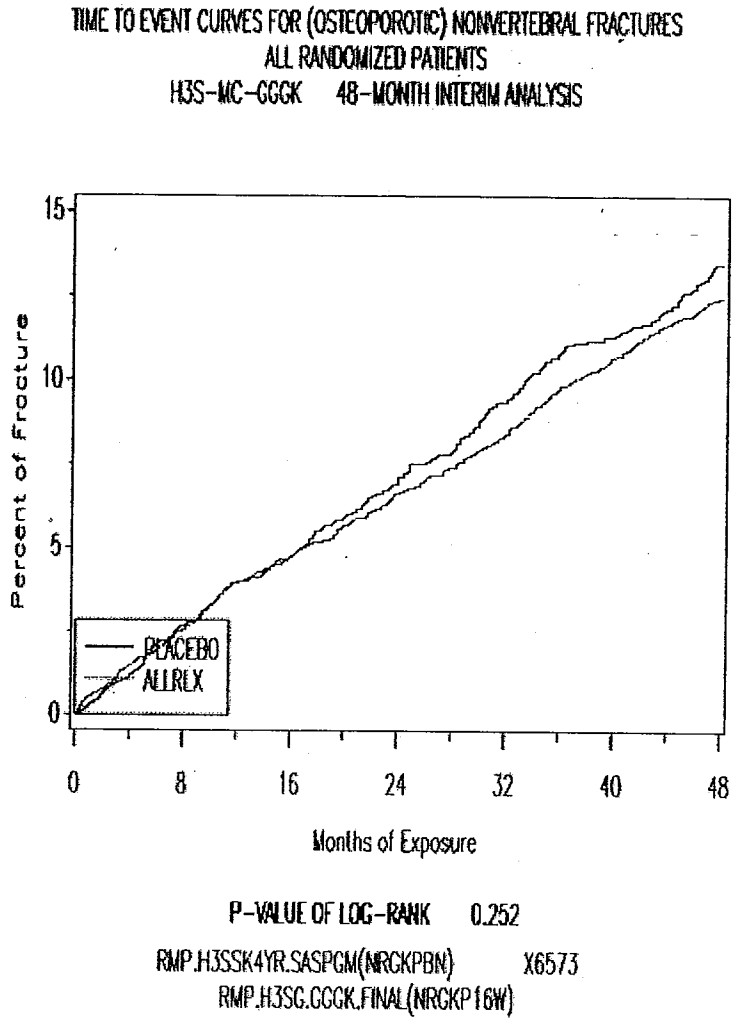

Figure GGGK.11.5. Kaplan-Meier Analysis: Time to First Osteoporotic Nonvertebral Fracture (All Randomly Assigned Patients, 48-Month Data)

11.4.4.1.3. New Vertebral and Any Nonvertebral Osteoporotic Fractures

A specified secondary endpoint in this study was the occurrence of any fracture. This was evaluated by pooling adjudicated new vertebral fractures with any nonvertebral osteoporotic fracture. Overall, there was a statistically significant decrease in the proportion of patients who reported any incident fracture (new vertebral or osteoporotic nonvertebral) in both raloxifene treatment groups and the pooled raloxifene group compared with the placebo group ($p \leq 0.001$) (Table GGGK.11.16).

Table GGGK.11.16. New Vertebral and Any Nonvertebral Osteoporotic Fracture Results (All Randomly Assigned Patients, 48-Month Data)

	Placebo	RLX060	RLX120	Pooled RLX
Overall	(N=2576)	(N=2557)	(N=2572)	(N=5129)
Proportion of patients reporting at least one incident fracture	539 (20.9%)	438 (17.1%)	390 (15.2%)	828 (16.1%)
Relative risk (95% CI) compared with placebo		0.82 (0.73, 0.92)	0.73 (0.64, 0.82)	0.77 (0.70, 0.85)
Pairwise comparison with placebo		p=0.001	p<0.001	p<0.001
Substudy I	(N=1689)	(N=1672)	(N=1703)	(N=3375)
Proportion of patients reporting at least one incident fracture	260 (15.4%)	211 (12.6%)	196 (11.5%)	407 (12.1%)
Relative risk (95% CI) compared with placebo		0.82 (0.69, 0.97)	0.75 (0.63, 0.89)	0.78 (0.68, 0.90)
Pairwise comparison with placebo		p=0.021	p<0.001	p<0.001
Substudy II	(N=887)	(N=885)	(N=869)	(N=1754)
Proportion of patients reporting at least one incident fracture	279 (31.5%)	227 (25.6%)	194 (22.3%)	421 (24.0%)
Relative risk (95% CI) compared with placebo		0.82 (0.70, 0.95)	0.71 (0.61, 0.83)	0.76 (0.67, 0.87)
Pairwise comparison with placebo		p<0.001	p<0.007	p<0.001

Abbreviations: RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; pooled RLX = raloxifene 60 mg and raloxifene 120 mg combined; CI = confidence interval; N = number of randomly assigned patients.

11.4.4.2. Bone Mineral Density (BMD)

At 48 months, for every skeletal site measured, the mean percentage change in BMD from baseline to endpoint in each raloxifene group was significantly greater than in the placebo group ($p<0.001$ for the pair-wise comparisons between each raloxifene group and the placebo group) (Table GGGK.11.17). These results are similar to the result observed at 36 months. Table GGGK.11.18 and Table GGGK.11.19 present the percentage change from baseline to endpoint in BMD for all skeletal sites for Substudy I and Substudy II patients, respectively.

Table GGGK.11.17. Summary of Percentage Change in BMD (From Baseline to Endpoint, All Randomly Assigned Patients, 48-Month Data)

Test		Treatment Group: Placebo	RLX060	RLX120	Overall p-value*
Lumbar Spine BMD	Mean Baseline	0.814	0.817	0.816	0.555
	Mean Change	0.006	0.027a	0.026a	0.000
	Mean Percentage Change	0.740	3.293a	3.268a	0.000
Femoral Neck BMD	Mean Baseline	0.622	0.625	0.621	0.310
	Mean Change	-0.009	0.004c	0.006a	0.000
	Mean Percentage Change	-1.296	0.797a	0.970a	0.000
Trochanter BMD	Mean Baseline	0.556	0.558	0.552d	0.051
	Mean Change	-0.005	0.007c	0.008a	0.000
	Mean Percentage Change	-0.835	1.284c	1.683a	0.000
Inter-Trochanter BMD	Mean Baseline	0.837	0.838	0.836	0.921
	Mean Change	-0.010	0.006a	0.008a	0.000
	Mean Percentage Change	-1.148	0.746c	1.060a	0.000
Wards Triangle BMD	Mean Baseline	0.459	0.462	0.460	0.456
	Mean Change	-0.015	0.000a	0.001a	0.000
	Mean Percentage Change	-2.911	0.391c	0.590a	0.000
Radial Ultradistal BMD	Mean Baseline	0.309	0.309	0.306	0.441
	Mean Change	-0.006	0.002c	0.000a	0.000
	Mean Percentage Change	-1.474	1.211c	0.655a	0.000
Radial Distal 1/3 BMD	Mean Baseline	0.541	0.543	0.540	0.807
	Mean Change	-0.005	0.001c	0.001c	0.000
	Mean Percentage Change	-0.809	0.336c	0.212a	0.000
Whole Body BMD	Mean Baseline	0.893	0.891	0.888	0.606
	Mean Change	-0.004	0.006c	0.005a	0.000
	Mean Percentage Change	-0.474	0.762c	0.616a	0.000

* Using ANOVA with Unranked data
a - pairwise comparison statistically significantly ($p < 0.05$) different from placebo
b - pairwise comparison statistically significantly ($p < 0.01$) different from placebo
c - pairwise comparison statistically significantly ($p < 0.001$) different from placebo
d - pairwise comparison of RLX060 statistically significantly ($p < 0.05$) different from RLX120
DATA FROM RMP.SAS.H3SM.HCGGGKSC.FINAL
RMP.H3SSK4YR.SASPGM(BMDVBLTP) X6646

Table GGGK.11.18. Summary of Percentage Change in BMD (From Baseline to Endpoint, Randomly Assigned Substudy I Patients, 48-Month Data)

Test		Treatment Group: Placebo	RLX060	RLX120	Overall p-value*
Lumbar Spine BMD	Mean Baseline	0.823	0.830	0.825	0.319
	Mean Change	0.004	0.026c	0.034c	0.000
	Mean Percentage Change	0.446	3.156c	2.984c	0.000
Femoral Neck BMD	Mean Baseline	0.628	0.631	0.628	0.274
	Mean Change	-0.008	0.005c	0.007c	0.000
	Mean Percentage Change	-1.234	0.814c	1.093c	0.000
Trochanter BMD	Mean Baseline	0.562	0.566	0.560d	0.095
	Mean Change	-0.004	0.007c	0.009c	0.000
	Mean Percentage Change	-0.728	1.324c	1.754c	0.000
Inter-Trochanter BMD	Mean Baseline	0.850	0.853	0.848	0.564
	Mean Change	-0.010	0.006c	0.009c	0.000
	Mean Percentage Change	-1.127	0.832c	1.169c	0.000
Wards Triangle BMD	Mean Baseline	0.468	0.474	0.472	0.171
	Mean Change	-0.016	0.000c	0.002c	0.000
	Mean Percentage Change	-2.993	0.389c	0.756c	0.000
Radial Ultradistal BMD	Mean Baseline	0.312	0.314	0.311	0.601
	Mean Change	-0.006	0.002c	0.001c	0.000
	Mean Percentage Change	-1.670	1.121c	0.821c	0.000
Radial Distal 1/3 BMD	Mean Baseline	0.545	0.549	0.545	0.632
	Mean Change	-0.005	-0.000c	0.000c	0.000
	Mean Percentage Change	-0.873	0.087c	0.112c	0.000
Whole Body BMD	Mean Baseline	0.896	0.899	0.892	0.397
	Mean Change	-0.004	0.007c	0.005c	0.000
	Mean Percentage Change	-0.448	0.878c	0.582c	0.000

* Using ANOVA with Unranked data
a - pairwise comparison statistically significantly ($p < 0.05$) different from placebo
b - pairwise comparison statistically significantly ($p < 0.01$) different from placebo
c - pairwise comparison statistically significantly ($p < 0.001$) different from placebo
d - pairwise comparison of RLX060 statistically significantly ($p < 0.05$) different from RLX120
DATA FROM RMP.SAS.N38M.MCGGGESC.FINAL
RMP.N38SK4YR.SASPGM(HMDVRLS1) I6646

Table GGGK.11.19. Summary of Percentage Change in BMD (From Baseline to Endpoint, Randomly Assigned Substudy II Patients, 48-Month Data)

Test		Treatment Group			Overall p-value*
		Placebo	RLX060	RLX120	
Lumbar Spine BMD	Mean Baseline	0.795	0.794	0.791	0.812
	Mean Change	0.010	0.027c	0.030c	0.000
	Mean Percentage Change	1.317	3.478c	3.831c	0.000
Femoral Neck BMD	Mean Baseline	0.612	0.611	0.606	0.582
	Mean Change	-0.009	0.004c	0.004c	0.000
	Mean Percentage Change	-1.416	0.763c	0.723c	0.000
Trochanter BMD	Mean Baseline	0.545	0.543	0.537	0.202
	Mean Change	-0.006	0.006c	0.007c	0.000
	Mean Percentage Change	-1.044	1.108c	1.541c	0.000
Inter-Trochanter BMD	Mean Baseline	0.810	0.809	0.813	0.826
	Mean Change	-0.010	0.004c	0.006c	0.000
	Mean Percentage Change	-1.190	0.578c	0.839c	0.000
Wards Triangle BMD	Mean Baseline	0.441	0.440	0.437	0.766
	Mean Change	-0.014	0.001c	-0.000c	0.000
	Mean Percentage Change	-2.750	0.395c	0.257c	0.000
Radial Ultradistal BMD	Mean Baseline	0.301	0.297	0.295	0.436
	Mean Change	-0.004	0.001b	-0.001	0.015
	Mean Percentage Change	-1.054	0.962a	0.321	0.032
Radial Distal 1/3 BMD	Mean Baseline	0.533	0.529	0.531	0.865
	Mean Change	-0.003	0.004c	0.001a	0.001
	Mean Percentage Change	-0.672	0.897c	0.415b	0.000
Whole Body BMD	Mean Baseline	0.885	0.874	0.881	0.374
	Mean Change	-0.005	0.004c	0.006c	0.000
	Mean Percentage Change	-0.530	0.509c	0.685c	0.000

* Using ANOVA with Unranked data
a - pairwise comparison statistically significantly($p < 0.05$) different from placebo
b - pairwise comparison statistically significantly($p < 0.01$) different from placebo
c - pairwise comparison statistically significantly($p < 0.001$) different from placebo
d - pairwise comparison of RLX060 statistically significantly ($p < 0.05$) different from RLX120
DATA FROM RMP.SAS.H3SM.MCGGGXSC.FINAL
RMP.H3SSK4YR.SASPGM(HKDVRLS2) X6646

11.4.4.4.1. Total (L-1 through L-4) Lumbar Spine BMD

Figure GGGK.11.6, Figure GGGK.11.7, and Figure GGGK.11.8 summarize graphically the total (L-1 through L-4) lumbar spine BMD data:

- In the **placebo group**, lumbar spine BMD increased by 0.50% at 12 months ($p<0.001$), decreased slightly to an overall gain of 0.27% at 24 months ($p=0.003$), increased by 0.43% ($p<0.001$) at 36 months, and increased by 0.74% at 48 months ($p\leq0.001$) compared with baseline.
- Patients in the **raloxifene 60-mg** group had gains of 2.50% at 12 months, 2.85% at 24 months, 3.09% at 36 months, and 3.29% at 48 months (all within-group $p<0.001$).
- Patients in the **raloxifene 120-mg** group had gains of 2.67% at 12 months, 2.92% at 24 months, 3.12% at 36 months, and 3.26% at 48 months (each within-group $p<0.001$).

Each raloxifene group had significantly greater increases in lumbar spine BMD than the placebo group at 12, 24, 36, and 48 months:

- The **differences** between the raloxifene 60-mg group compared with the placebo group were 2.00% at 12 months, 2.58% at 24 months, 2.67% at 36 months, and 2.56% at 48 months (all pair-wise $p<0.001$).
- The **differences** between the raloxifene 120-mg group compared with the placebo group were 2.17% at 12 months, 2.66% at 24 months, 2.69% at 36 months, and 2.52% at 48 months (all pair-wise $p<0.001$).
- The raloxifene groups were not significantly different from each other at any of the time points. Similar results were seen for Substudy I (Figure GGGK.11.7) and Substudy II (Figure GGGK.11.8) individually.

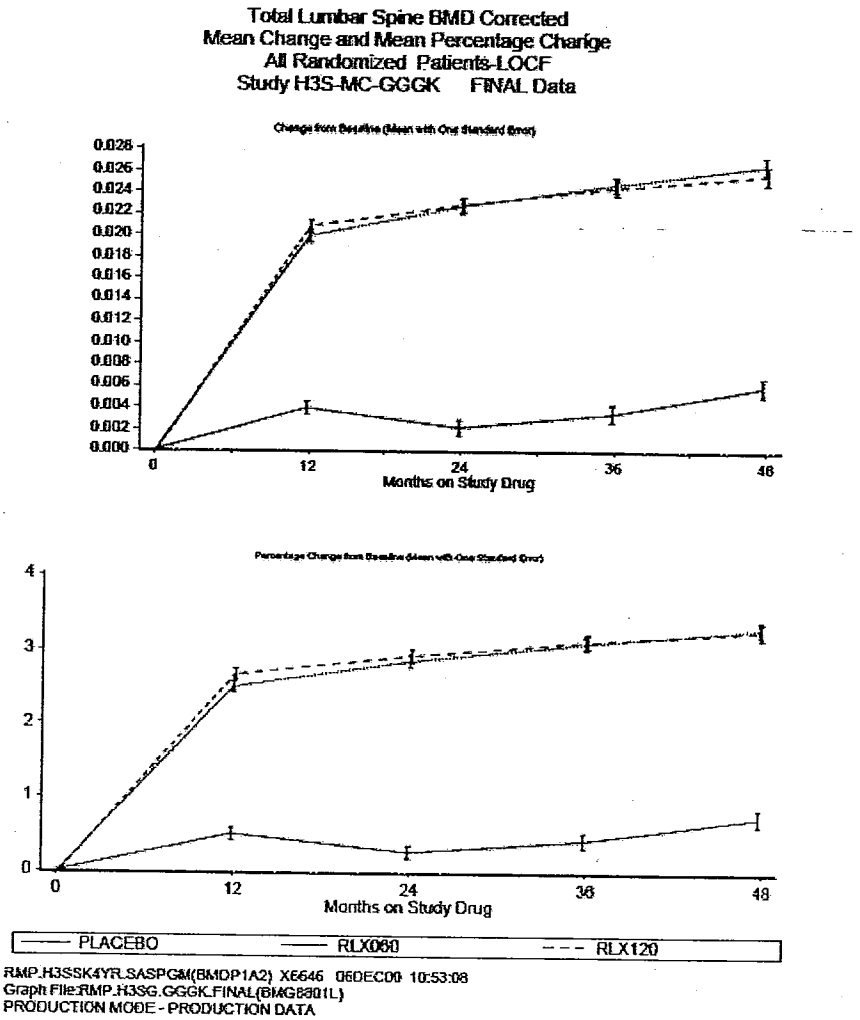

Figure GGGK.11.6. Mean Change and Mean Percentage Change in Total Lumbar Spine BMD (All Randomly Assigned Patients, 48-Month Data)

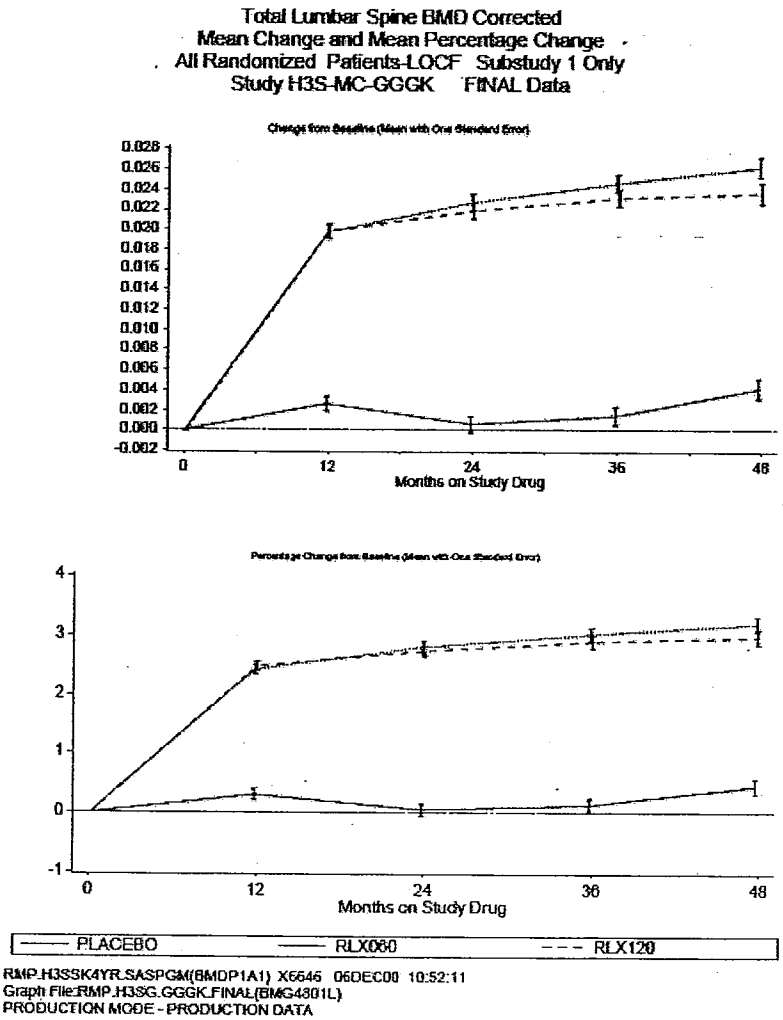

Figure GGGK.11.7. Mean Change and Mean Percentage Change in Total Lumbar Spine BMD (All Randomly Assigned Substudy I Patients, 48-Month Data)

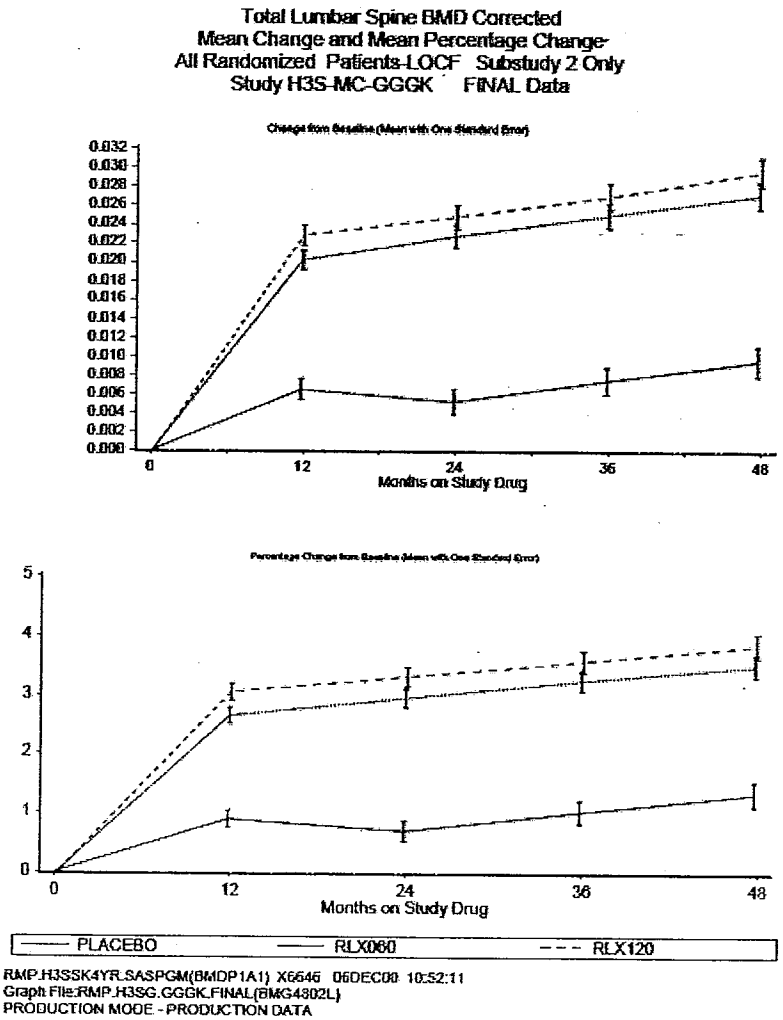

Figure GGGK.11.8. Mean Change and Mean Percentage Change in Total Lumbar Spine BMD (All Randomly Assigned Substudy II Patients, 48-Month Data)

11.4.4.4.2. Femoral Neck BMD

Femoral neck BMD data are summarized graphically in (Figure GGGK.11.9, Figure GGGK.11.10, and Figure GGGK.11.11).

- In the placebo group, femoral neck BMD increased by 0.08% at 12 months (within-group p=0.407), then decreased below baseline by 0.38% at 24 months (within-group p<0.001), and further decreased by 1.16% at 36 months and by 1.30% at 48 months (within-group p<0.001) compared with baseline.
- The raloxifene 60-mg group had gains of 1.35% at 12 months, 1.49% at 24 months, 0.94% at 36 months, and 0.80% at 48 months (all within-group p<0.001).
- The raloxifene 120-mg group had gains of 1.45% at 12 months, 1.81% at 24 months, 1.25% at 36 months, and 0.97% at 48 months (all within-group p<0.001).
- Similar results were seen for Substudy I (Figure GGGK.11.10) and Substudy II (Figure GGGK.11.11) individually.

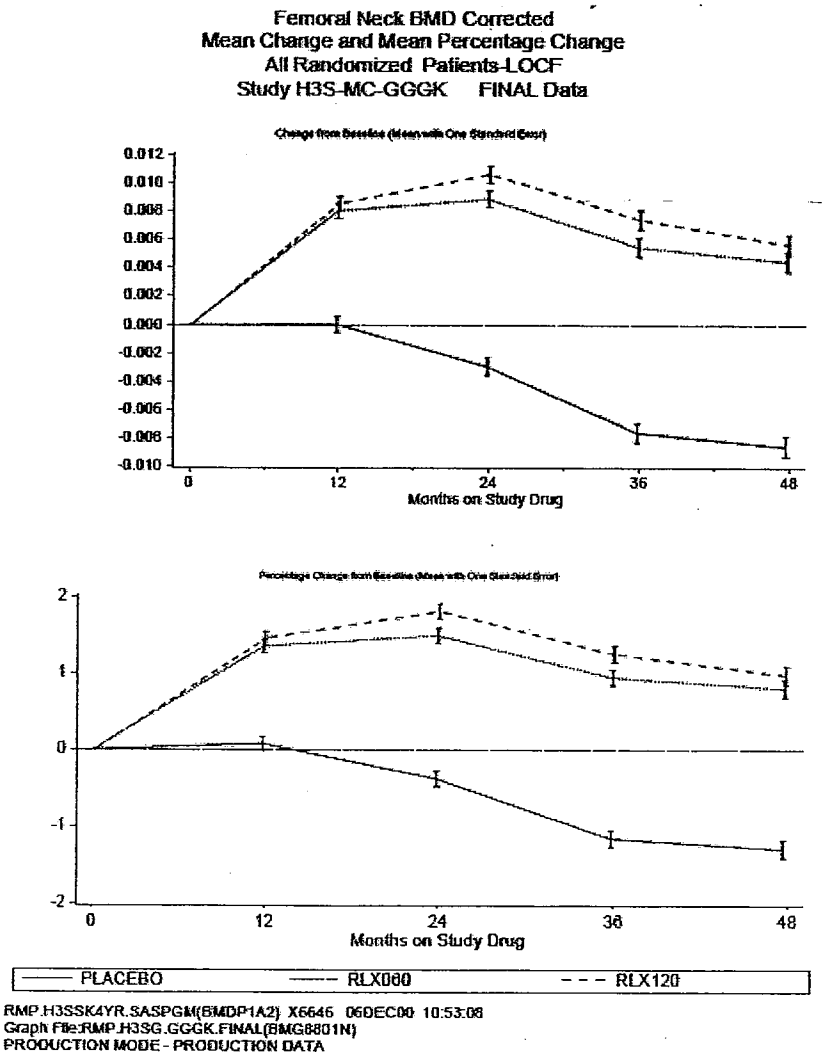

Figure GGGK.11.9. Mean Change and Mean Percentage Change in Femoral Neck BMD (All Randomly Assigned Patients, 48-Month Data)

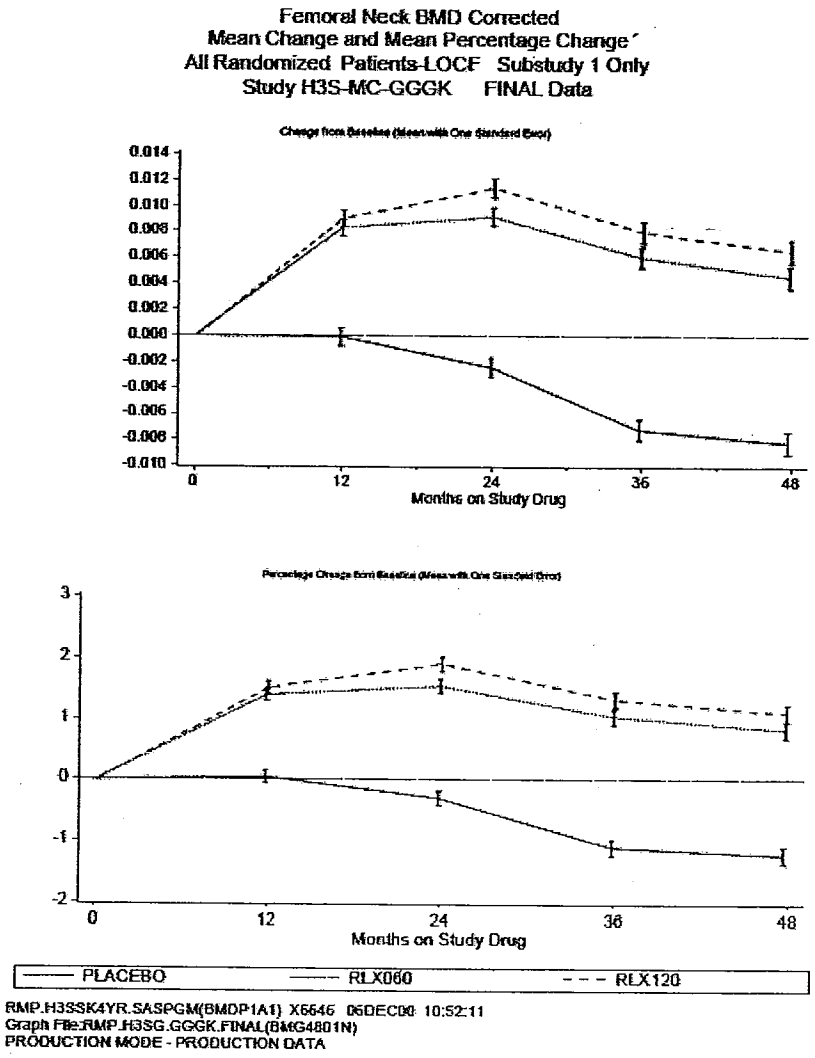

Figure GGGK.11.10. Mean Change and Mean Percentage Change in Femoral Neck BMD (All Randomly Assigned Substudy I Patients, 48-Month Data)

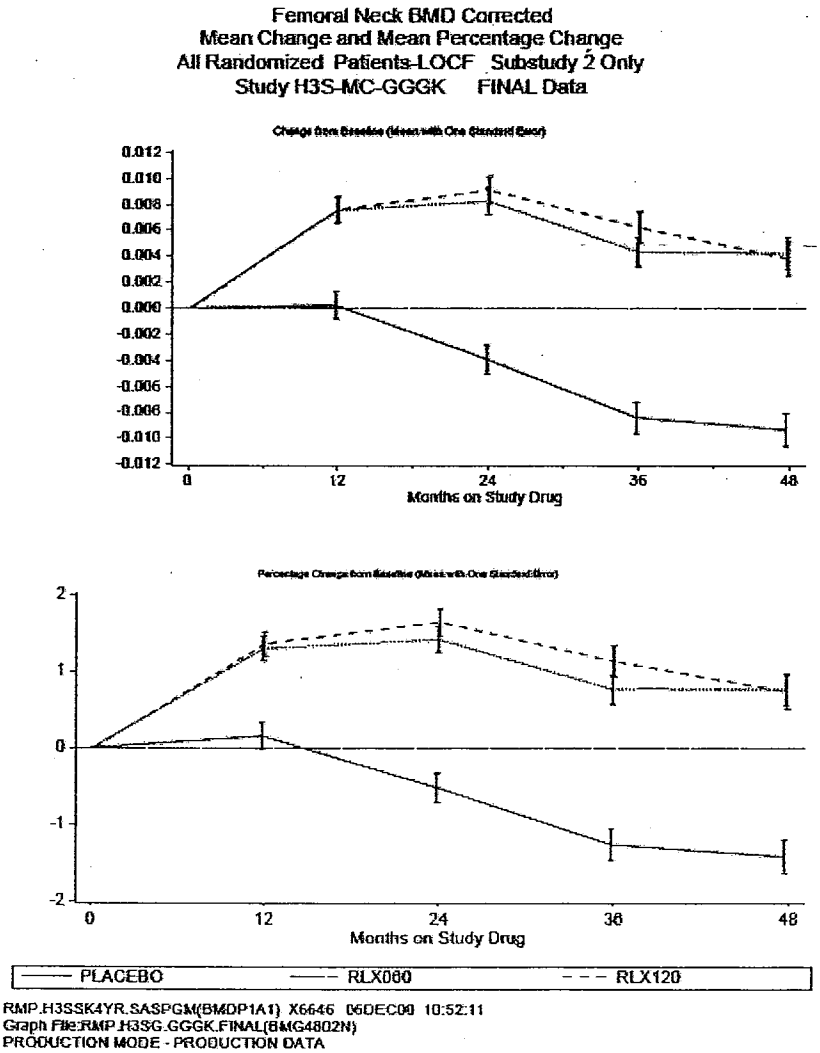

Figure GGGK.11.11. Mean Change and Mean Percentage Change in Femoral Neck BMD (All Randomly Assigned Substudy II Patients, 48-Month Data)

11.4.4.3. Biochemical Markers of Bone Metabolism

For change and percentage change in biochemical markers of bone metabolism, the distribution of residuals using ANOVA with unranked data were generally found to be non-normal. Thus, to best evaluate the central tendency, the median and statistical significance from ANOVA with ranked data is presented. Table GGGK.11.20 summarizes the median baseline and percentage change from baseline to endpoint for each biochemical marker of bone metabolism.

- Compared with the placebo group, raloxifene-treated patients had statistically significant median decreases in total osteocalcin, bone-specific alkaline phosphatase, carboxy-terminal pro-peptide of type I pro-collagen, urinary type I collagen fragment-to-creatinine ratio, calcium/creatinine ratio, 25-hydroxy vitamin D, and total alkaline phosphatase, respectively, at 48 months.

Table GGGK.11.20. Summary of Biochemical Markers of Bone Metabolism (Median Baseline and Percentage Change, From Baseline to Endpoint, Randomly Assigned Biochemical Markers of Bone Metabolism Subset Patients)

	Unit		Treatment Group			Overall p-value*
			Placebo	RLX060	RLX120	
Osteocalcin	ug/L	Median Baseline	24.000	24.100	24.800	0.148
		Median Change	-2.000	-6.400c	-7.400cd	<0.001
		Median Percentage Change	-8.571	-26.24c	-31.10cd	<0.001
Bone-Specific Alkaline Phosphatase	ug/L	Median Baseline	15.700	15.800	15.800	0.778
		Median Change	-2.980	-5.350c	-5.245c	<0.001
		Median Percentage Change	-19.89	-35.22c	-35.56c	<0.001
Carboxy-Terminal Propeptide of Type I procollagen	ng/ml	Median Baseline	115.00	114.35	116.00	0.930
		Median Change	10.000	3.000c	3.000c	<0.001
		Median Percentage Change	8.947	2.564c	2.727c	<0.001
Urinary Type I Collagen Fragment to Creatinine Ratio	ug/mmcrt	Median Baseline	244.47	253.80	247.30	0.178
		Median Change	-19.16	-81.20c	-75.32c	<0.001
		Median Percentage Change	-8.216	-34.23c	-31.76c	<0.001
Calcium/Creatinine Ratio	mM/mM	Median Baseline	0.420	0.430	0.400	0.283
		Median Change	0.000	-0.030b	-0.030a	0.013
		Median Percentage Change	0.000	-8.333a	-11.70a	0.025
PTH-Intact	pmol/L	Median Baseline	3.074	3.000	3.100a	0.070
		Median Change	0.100	0.400c	0.400c	<0.001
		Median Percentage Change	3.410	13.636c	13.158c	<0.001
25-Hydroxy Vitamin D	nmol/L	Median Baseline	74.000	75.000	74.000	0.179
		Median Change	14.000	10.000c	10.000c	<0.001
		Median Percentage Change	18.750	14.815c	14.286c	<0.001
Total Alkaline Phosphatase	U/L	Median Baseline	74.000	74.000	73.000	0.279
		Median Change	0.000	-8.000c	-9.000cd	<0.001
		Median Percentage Change	0.000	-11.39c	-13.25cd	<0.001

* Using ANOVA with Ranked data
a - pairwise comparison statistically significantly($p < 0.05$) different from placebo
b - pairwise comparison statistically significantly($p < 0.01$) different from placebo
c - pairwise comparison statistically significantly($p < 0.001$) different from placebo
d - pairwise comparison of RLX060 statistically significantly ($p < 0.05$) different from RLX120
PTH Intact and 25-Hydroxy Vitamin D data were collected in Visit 2 and 4 only
DATA FROM RMP.SAS.H3SM.MCGGGKSC.FINAL
RMP.H3SSK4YR.SASPGM(HMKISRE) X6646

11.4.5. Serum Lipids and Biochemical Markers of Cardiovascular Risk

For change and percentage change in biochemical markers of cardiovascular risk, the distribution of residuals using ANOVA with unranked data was generally found to be non normal. Thus, to best evaluate the central tendency, the median and statistical significance from ANOVA using ranked data is presented.

- Compared with the placebo group, raloxifene-treated patients had statistically significant median decreases of 5% to 6%, 8% to 10%, and 10% to 12% in total cholesterol, low-density lipoprotein-cholesterol (LDL-C), and fibrinogen, respectively, at 48 months.
- There were also statistically significant decreases in LDL-C/high-density lipoprotein-cholesterol (HDL-C) ratio and apolipoprotein B in patients treated in both raloxifene groups compared with the placebo group.
- Statistically significant differences in the baseline to endpoint change in triglyceride concentrations were observed in both raloxifene treatment groups compared with placebo-treated patients (p=0.004).
- No statistically significant differences were observed for total HDL-C or hemoglobin A1c (HbA1c).
- A statistically significant median increase of 2% in apolipoprotein A1 was observed in both raloxifene groups compared with the placebo group.
- The consistent lowering of total cholesterol, LDL-C, LDL-C/HDL-C ratio, and fibrinogen suggests that raloxifene might reduce the risk for major cardiovascular events. ***(Reviewer Comments:*** *RUTH trial did not support this, however.)*

Table GGGK.11.21 summarizes the median baseline, change, and percentage change from baseline to endpoint for each of these biochemical markers of cardiovascular risk.

Table GGGK.11.21. Summary of Serum Lipids and Biochemical Markers of Cardiovascular Risk (Median Baseline, Change, and Percentage Change From Baseline to Endpoint, All Randomly Assigned and Randomly Assigned Biochemical Markers of Cardiovascular Risk Subset Patients, 48-Month Data)

	Unit		Treatment Group			Overall p-value*
			Placebo	RLX060	RLX120	
Total Cholesterol	mmol/L	Median Baseline	6.077	6.100	6.051	0.228
		Median Change	-0.103	-0.410c	-0.466cd	<0.001
		Median Percentage Change	-1.780	-6.704c	-7.861cd	<0.001
LDL Cholesterol	mmol/L	Median Baseline	4.008	4.008	3.980	0.267
		Median Change	-0.100	-0.430c	-0.517cd	<0.001
		Median Percentage Change	-2.618	-10.64c	-13.28cd	<0.001
HDL Cholesterol	mmol/L	Median Baseline	1.500	1.520	1.500	0.558
		Median Change	0.052	0.052	0.051	0.685
		Median Percentage Change	3.571	3.980	3.158	0.674
LDL/HDL Ratio	No Unit	Median Baseline	2.705	2.690	2.630	0.634
		Median Change	-0.112	-0.338c	-0.418cd	<0.001
		Median Percentage Change	-4.860	-14.45c	-16.76cd	<0.001
Apolipoprotein A1	g/L	Median Baseline	1.540	1.555	1.540	0.759
		Median Change	-0.010	0.040c	0.040c	<0.001
		Median Percentage Change	-0.769	2.666c	2.540c	<0.001
Apolipoprotein B	g/L	Median Baseline	1.480	1.460	1.470	0.398
		Median Change	-0.080	-0.170c	-0.190c	<0.001
		Median Percentage Change	-5.789	-11.59c	-13.40c	<0.001
Triglyceride	mmol/L	Median Baseline	1.110	1.061a	1.061a	0.062
		Median Change	-0.030	0.010b	0.011c	0.002
		Median Percentage Change	-2.597	0.971a	1.297b	0.004
Fibrinogen	g/L	Median Baseline	3.355	3.350	3.350	0.856
		Median Change	-0.080	-0.430c	-0.470c	<0.001
		Median Percentage Change	-2.332	-12.39c	-14.24c	<0.001
Hemoglobin A1C	1	Median Baseline	0.057	0.057	0.057	0.705
		Median Change	-0.001	-0.001	-0.001	0.262
		Median Percentage Change	-1.667	-1.754	-1.724	0.267

* Using ANOVA with Ranked data
a - pairwise comparison statistically significantly(p < 0.05) different from placebo
b - pairwise comparison statistically significantly(p < 0.01) different from placebo
c - pairwise comparison statistically significantly(p < 0.001) different from placebo
d - pairwise comparison of RLX060 statistically significantly (p < 0.05) different from RLX120
DATA FROM RMP.SAS.H3SM.RCGGGK8C.FINAL
RMP.H3SSK4YR.SASPGM(LPDISRR) X6646

11.4.5.1. Triglyceride Concentration

Statistically significant median percentage change increases in triglyceride concentrations were observed in both raloxifene treatment groups compared with the reduction observed in placebo-treated patients (p=0.004) (Table GGGK.11.21).

- **Analyses of variance** with unranked data revealed similar results.

- A **categorical analysis** was performed comparing the proportion of patients at *different levels of triglyceride concentration increases* among the three treatment groups. This analysis showed that a small increase in mean triglyceride concentration in the raloxifene groups was due to a statistically significantly greater proportion of patients with triglyceride increases >1 mmol/L in the raloxifene groups compared with the placebo group (overall p=0.036) (Table GGGK.11.22). However, this statistical significance was not observed for patients with higher changes (a change of >2 or >3 mmol/L).
- **Subgroup analyses** were performed to assess if a **treatment-by-baseline interaction** effect was present for percentage change in triglyceride concentrations (Table GGGK.11.23). For percentage change, an ANOVA with treatment-by-baseline-tertile interaction effect on ranked data was performed. At the 0.10 level of statistical significance (two-sided), a treatment-by-baseline-tertile interaction effect was found for percentage change in triglyceride concentrations (p=0.054).
    - Patients with high values at baseline (in the upper tertile) had the greatest decrease in triglyceride concentration among all three treatment groups.
    - For patients in the upper tertile, the percentage change decrease in the triglyceride concentration was significantly greater in the placebo group compared with the raloxifene 60-mg group (p=0.002).
- **Subgroup analyses** were also performed to assess if treatment effect was modified by **baseline body mass index (BMI)**. The treatment-by-BMI interaction effect was significant (p=0.01; Table GGGK.11.24).
    - For patients in the upper tertile, there were significant percentage change increases in the raloxifene 60-mg group compared with the placebo group (p=0.007).
    - For patients in the middle tertile, the percentage change decrease in the triglyceride concentration was significantly greater in the placebo group compared with the raloxifene 120-mg group (p<0.001).
- For the percentage change in triglyceride concentration, a **subgroup analysis** was also performed to assess if treatment effect was modified for patients with **baseline diabetes** (defined as either a preexisting condition of diabetes mellitus, a baseline fasting glucose >7.78 mmol/L, or baseline hypoglycemic use) (Table GGGK.11.25).
    - No significant treatment-by-baseline-diabetes interaction was observed for percentage change in triglyceride concentration.

Table GGGK.11.22. Threshold Analysis of Triglyceride Concentration (Baseline, Endpoint, and Change, Randomly Assigned Biochemical Markers of Cardiovascular Risk Subset Patients, 48-Month Data)

Measurement	Threshold	PLACEBO (n=2567)	RLX060 (n=2548)	RLX120 (n=2564)	Total (N=7679)	Overall p-value
Endpoint Actual	> 0 mmol/L	929 (36.2)	913 (35.8)	914 (35.6)	2756 (35.9)	0.915
	> 1 mmol/L	537 (20.9)	497 (19.5)	531 (20.7)	1565 (20.4)	0.400
	> 2 mmol/L	106 (4.1)	136 (5.3)	118 (4.6)	360 (4.7)	0.110
	> 3 mmol/L	24 (0.9)	36 (1.4)	26 (1.0)	86 (1.1)	0.220
	> 4 mmol/L	8 (0.3)	6 (0.2)	8 (0.3)	22 (0.3)	0.841
	> 5 mmol/L	4 (0.2)	3 (0.1)	3 (0.1)	10 (0.1)	0.907
	> 6 mmol/L	4 (0.2)	1 (0.0)	2 (0.1)	7 (0.1)	0.371
Positive Change	> 0 mmol/L	434 (16.9)	461 (18.1)	462 (18.0)	1357 (17.7)	0.460
	> 1 mmol/L	23 (0.9)	36 (1.4)	44 (1.7)	103 (1.3)	0.036
	> 2 mmol/L	5 (0.2)	7 (0.3)	4 (0.2)	16 (0.2)	0.637
	> 3 mmol/L	3 (0.1)	2 (0.1)	2 (0.1)	7 (0.1)	0.869

P-VALUE FROM PEARSON CHI-SQUARE TEST
RMP.H3SSK4YR.SASPGM(TRICATPG) X6646
DATA FROM RMP.SAS.H3SM.GGGK.FINAL

Table GGGK.11.23. Analysis of Percentage Change in Triglyceride Concentrations by Baseline Tertiles (Randomly Assigned Biochemical Markers of Cardiovascular Risk Subset Patients, 48-Month Data)

Test	Subgroup Interaction Effect	Treatment	LOWER 33% Median pchange	LOWER 33% Between Group p-Value*	MIDDLE 33% Median pchange	MIDDLE 33% Between Group p-Value*	UPPER 33% Median pchange	UPPER 33% Between Group p-Value*
Triglyceride	0.054	PLACEBO	7.500	-	-.763	-	-11.7	-
		RLX060	5.489	0.714	-.962	0.314	-2.93	0.002
		RLX120	11.32	0.064	0.463	0.069	-8.22	0.158

* Compared to Placebo Group
DATA FROM RMP.SAS.H3SM.MCGGGKSC.FINAL
RMP.H3SSK4YR.SASPGM(LPDSUBPG) X6646

Table GGGK.11.24. Analysis of Percentage Change in Triglyceride Concentrations by Baseline BMI Tertiles (Randomly Assigned Biochemical Markers of Cardiovascular Risk Subset Patients, 48-Month Data)

Test	Subgroup Interaction Effect	Treatment	LOWER 33% Median pchange	LOWER 33% Between Group p-value*	MIDDLE 33% Median pchange	MIDDLE 33% Between Group p-value*	UPPER 33% Median pchange	UPPER 33% Between Group p-value*
Triglyceride	0.010	PLACEBO	-2.06	-	-4.84	-	-1.16	-
		RLX060	-4.35	0.894	0.000	0.090	6.450	0.007
		RLX120	-4.09	0.987	9.033	<0.001	3.175	0.113

* Compared to Placebo Group
DATA FROM RMP.SAS.H3SM.MCGGGKSC.FINAL
RMP.H3SSK4YR.SASPGM(LPDSUBP5) X6646

Table GGGK.11.25. Analysis of Percentage Change in Triglyceride Concentrations by Patients' Baseline Diabetes Status (Randomly Assigned Biochemical Markers of Cardiovascular Risk Subset Patients, 48-Month Data)

			Diabetes or Hypoglycemics					Others			
Subgroup Variables	Subgroup Interaction Effect	Treatment		Median Percent Change	Within Group P-Value	Between Group P-Value*		Median Percent Change	Within Group P-Value	Between Group P-Value*
Triglyceride	0.466	PLACEBO	25	-3.546	0.424	-	904	-2.477	0.150	-
		RLX060	36	5.355	0.068	0.448	877	0.833	0.518	0.028
		RLX120	27	-7.087	0.122	0.606	887	2.158	0.327	0.001

* Compared to Placebo Group
DATA FROM RMP.SAS.H3SM.NCGGGKSC.FINAL
RMP.H3SSK4YR.SASPGM(ATCSUB10) X6646

11.4.6. Cognitive and Neuropsychomotor Assessment

The effect of raloxifene on different aspects of cognitive and neuropsychomotor function was assessed using the cognitive and neuropsychomotor test and MAPS (cognitive function mapping) batteries. In addition, the effect of raloxifene on the number of falls and near falls was assessed. The Dementia Diagnostic Evaluation (DDE) was used to determine the effect of raloxifene on the prevalence of Alzheimer's disease, dementia associated with cerebrovascular disease, and all causes of dementia.

A majority of sites performed one or more of the tests in the cognitive and neuropsychomotor test battery. This battery included six tests of cognitive function and four tests of neuropsychomotor function: The Affective Rating Scale, The Short Blessed, The Trail Making A and B, The Word List Memory and Recall, The Word List Fluency,
The Static Test, The Muscle Strength, and The Gait Assessment (refer to Appendix 16.1.1, Protocol GGGK[f] for a detailed description of each test).

There were no overall statistically significant differences among the three treatment groups in the change or percentage change from baseline to endpoint using ANOVA with unranked data for any of the cognitive and neuropsychomotor batteries tests.

For the US sites only, there were two cognitive and neuropsychomotor batteries tests that showed a statistically significant difference overall in the change or percentage change from baseline to endpoint. There was a statistically significant overall difference among the three treatment groups for Affective Rating Scale score percentage change from baseline to endpoint using ANOVA with unranked data for US sites only ($p=0.010$).

There was a significantly greater decrease in the percentage change from baseline to endpoint for the placebo group compared with the increase observed in the raloxifene 120-mg group ($p<0.010$). There was a statistically significant overall difference among the three treatment groups for Short Blessed Test Time percentage change from baseline to endpoint using ANOVA with unranked data for US sites only ($p=0.045$). There was a significantly greater increase in the percentage change from baseline to endpoint for the raloxifene 60-mg group compared with the placebo group ($p<0.05$).

11.4.6.1. Falls

In addition to the neuropsychomotor tests, the number of times a patient fell and the number of times that a patient nearly fell between visits were also recorded for all patients in the cohort. For the purpose of this study, a fall was defined as falling and landing on the ground, floor, an object, or stair with or without the patient attempting to catch herself. A near fall was defined as catching oneself during a fall, thereby preventing falling all the way to the floor.

At baseline, there was no overall statistically significant difference among the three treatment groups in categorical analyses in the proportion of patients reporting falls, near falls, or falls and near falls combined during the 12 months prior to randomization. After randomization, a total of 4728 (61.4%) patients reported falling at least once and 3004 (39.0%) patients reported at least one near fall during the 48-month interval. There were no overall statistically significant differences among the three treatment groups in categorical analyses in the proportion of patients reporting falls and near falls.

11.4.6.2. MAPS Battery

In order to better delineate the effects of raloxifene on different domains of cognitive function, the MAPS battery of tests (Weingartner et al. 1983a; Weingartner et al. 1983b; Weingartner et al. 1992; Weingartner et al. 1993; Maki et al. 1999) was employed at two study sites in the United States. This battery included The Attention Vigilance Test, Vigilance Recall, Vigilance Recognition, Fragmented Pictures, Buschke, Category Retrieval 6 Versus 6 Recall, and 6 Versus 6 Recognition (refer to Appendix 16.1.1, Protocol GGGK[g] for a detailed description of each test). Sites that performed the MAPS battery were exempted from conducting the cognitive and neuropsychomotor tests described previously.

There were no overall statistically significant differences among the three treatment groups for Attention/Vigilance, Vigilance Recall, Fragmented Pictures, Category Retrieval, and 6 Versus 6 Recall at endpoint adjusted for baseline (using analysis of covariance [ANCOVA]) or in the change from baseline to endpoint (using ANOVA). Pair-wise comparisons with unranked data revealed no statistically significant between-group differences.

The Buschke consists of two scores, mean number correct from Trails 1-3 and mean consistency score from Trails 2-4. There was a statistically significant overall difference among the three treatment groups for mean number correct from Trails 2-4 in the change from baseline to endpoint (using ANOVA) with unranked data. Pair-wise comparisons with unranked data revealed statistically significant between-group differences for mean number correct from Trails 1-3 and 2-4 in the change from baseline to endpoint (using ANOVA) with unranked data (both p-values<0.05). Pair-wise comparisons with unranked data also revealed a statistically significant between-group difference for mean number correct from Trails 2-4 at endpoint adjusted for baseline ($p<0.05$).

11.4.6.3. Dementia Diagnostic Evaluation

Patients who met criteria for the DDE underwent an evaluation consistent with the NINCDS-ADRDA criteria (Appendix 16.1.1, Protocol GGGK[g] Attachment GGGK.18). This evaluation established the cognitive status of patients using the following categories: cognitively normal, mild cognitive impairment, Alzheimer's disease (AD), dementia associated with cardiovascular disease, other type of dementia, and dementia type indeterminate. The DDE consisted of two parts. Part 1 included interviews with the patient and caregiver (if appropriate), a medical history, physical and neurological examination, and a battery of cognitive and other tests. Part 2 consisted of cognitive laboratory tests which included fluorescent treponemal antibody (FTA), vitamin B_{12}, serum folate and thyroid-stimulating hormone (TSH), and a brain computed tomography (CT) or magnetic resonance imaging (MRI) scan without contrast, in order to define the type of dementia (Appendix 16.1.1, Protocol GGGK[g] Section 3.9.1.1.1 contains a detailed description of DDE Parts 1 and 2).

Data collected from the DDE was assessed primarily for the purpose of evaluating the effect of treatment with raloxifene on the prevalence of dementia and not for adverse events. Cognition-related adverse events (eg, dementia) were recorded at the discretion of the investigator based on the review of the Dementia Adjudication Committee and/or consulting clinician assessments.

Adverse events were routinely collected and reported by questioning the patient at each return visit (Appendix 16.1.1, Protocol GGGK[g] Section 3.9.4.2).
The results of the DDE were used by the Dementia Adjudication Committee to assess cognitive status of patients completing both Parts 1 and 2 of the DDE. If dementia was present, the Dementia Adjudication Committee assessed the type of dementia according to the previously indicated definitions. The criteria for the clinical diagnosis of AD were based on NINCDS-ADRDA criteria (Appendix 16.1.1, Protocol GGGK[g] Attachment GGGK.18). If dementia was present, the Dementia Adjudication Committee also determined if cognitive impairment preceded initiation of study medication.

11.4.6.3.1. Dementia Diagnostic Evaluation Assessment

Table GGGK.11.26 summarizes the characteristics of the patients who participated in the DDE. There was no difference among the three treatment groups in terms of demographics at baseline.

Table GGGK.11.26. Demographics for Dementia Data (All Randomly Assigned Patients, 48-Month Data)

	Placebo (N=250)	RLX060 (N=253)	RLX120 (N=243)	p-value
Age	67.99	68.02	67.99	0.873
Height	158.27	157.79	158.40	0.875
Weight	62.96	63.76	64.12	0.405
BMI	25.17	25.69	25.59	0.330
Years PMP	20.38	20.29	19.97	0.641
Years of Education	10.41	9.89	10.34	0.426
Current Smoker (YES)	45 (18.1%)	46 (18.3%)	37 (15.4%)	0.644
Family History of OSTPRS (YES)	53 (21.2%)	68 (26.9%)	57 (23.5%)	0.173
Family History of Breast Cancer (YES)	31 (12.4%)	30 (11.9%)	25 (10.3%)	0.697
Hysterectomy (YES)	57 (22.8%)	63 (24.9%)	63 (25.9%)	0.712
Previous Use of HRT (YES)	71 (28.4%)	70 (27.7%)	75 (30.9%)	0.823
Previous Use of THIAZ Diuretics (YES)	25 (10.0%)	37 (14.6%)	26 (10.7%)	0.141
Previous Use of Systemic Fluorides (YES)	1 (0.4%)	9 (3.6%)	5 (2.1%)	0.081
Previous Use of Bisphosphonates (YES)	7 (2.8%)	4 (1.6%)	4 (1.6%)	0.551
Prior Awareness of OSTPRS (YES)	95 (38.0%)	95 (37.5%)	86 (35.4%)	0.815

Table GGGK.11.27 summarizes the results of the effect of raloxifene on the prevalence of Alzheimer's disease and all causes of dementia. Note that all the analyses in this section use the patients who belong to the participated sites and were still in the trial at Visit 7 (total 6064 patients).

- There were no statistically significant differences among the three treatment groups in either Alzheimer's disease or the dementia category (regardless of whether it was a preexisting case or a treatment-emergent case).

Table GGGK.11.27. Dementia Diagnostic Data
Randomly Assigned Subset Patients
H3S-MC-GGGK 48-Month Data

Cognitive Category		Placebo (N=2046)[a]	RLX060 (N=2000)[a]	RLX120 (N=2018)[a]
Alzheimer's disease Treatment-emergent	Proportion of patients reported having	1 (0.0%)	3 (0.2%)	5 (0.2%)
	Relative risk (95% CI) compared with placebo		3.07(0.32, 29.5)	5.07 (0.59, 43.4)
	Pairwise comparison with placebo		p=0.31	p=0.10
Alzheimer's disease All cases	Proportion of patients reported having	15 (0.7%)	13 (0.7%)	8 (0.4%)
	Relative risk (95% CI) compared with placebo		0.89 (0.42, 1.89)	0.54 (0.23, 1.27)
	Pairwise comparison with placebo		p=0.75	p=0.15
Dementia	Proportion of patients reported having	18 (0.9%)	16 (0.8%)	15 (0.7%)
	Relative risk (95% CI) compared with placebo		0.91 (0.47, 1.78)	0.85 (0.43, 1.67)
	Pairwise comparison with placebo		p=0.78	p=0.63

Abbreviations: CI = confidence interval; N = number randomized; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120-mg/day.
[a] Includes patients who were still ongoing at Visit 7.

11.4.7. Drug-Drug and Drug-Disease Interactions

Concomitant medications that might confound the interpretation of efficacy and safety analyses were grouped by class based on anatomic therapeutic class (ATC). Table GGGK.11.28 lists each class, along with the proportion of patients in each treatment group who reported the use of any medication in that class.

- At baseline, there were statistically significant differences among the three treatment groups in the reported use of hypoglycemic and hypolipidemic agents.
- There were 169 patients in the placebo group, 222 patients in the raloxifene 60-mg group, and 181 patients in the raloxifene 120-mg group reporting the use of at least one **hypolipidemic** agent at baseline (p=0.010).
- There were 32 patients in the placebo group, 55 patients in the raloxifene 60-mg group, and 40 patients in the raloxifene 120-mg group reporting the use of at least one **hypoglycemic** agent at baseline (p=0.034).

Table GGGK.11.28. Baseline Use of Concomitant Medications That Might Confound the Interpretation of Efficacy and Safety Analyses (All Randomly Assigned Patients, 48-Month Data)

Class	Placebo N=2576	RLX060 N=2557	RLX120 N=2572	p-value
Estrogens	128 (5.0%)	108 (4.2%)	124 (4.8%)	0.408
Nonestrogenic bone-active drugs	158 (6.2%)	155 (6.1%)	183 (7.0%)	0.339
Hypolipidemic agents	169 (6.6%)	222 (8.7%)	181 (7.0%)	0.010
Progestins/androgens	2 (0.1%)	2 (0.1%)	6 (0.2%)	0.203
Hypoglycemic agents	32 (1.2%)	55 (2.2%)	40 (1.6%)	0.034
Antiseizure medications	14 (0.5%)	16 (0.6%)	22 (0.9%)	0.367
Corticosteroids	120 (4.7%)	111 (4.3%)	130 (5.1%)	.480

Abbreviations: RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; N = number of randomly assigned patients.

At baseline, the number of women who had received **prior hormone replacement therapy** (HRT) was similar across treatment groups.

- The treatment effect of raloxifene was similar for patients with and without prior use of HRT (p=0.597 for pooled raloxifene groups versus placebo).

Table GGGK.11.29 presents the relative risks for invasive breast cancer according to previous HRT status.

Table GGGK.11.29. Invasive Breast Cancer Incidence in Patients with Prior Hormone Replacement Therapy (All Randomly Assigned Patients, 48-Month Data)

	Placebo	Pooled Raloxifene	Relative Risk (95% CI) (Raloxifene vs Placebo)
Total Invasive Cases	n = 2576 Number of cases = 38 Incidence = 1.48%	n = 5129 Number of cases = 21 Incidence 0.41%	0.28 (0.16, 0.47)
Prior Use of HRT[a]	n = 738 Number of cases = 15 Incidence = 2.03%	n = 1497 Number of cases = 7 Incidence = 0.47%	0.23 (0.09, 0.56)
No Prior Use of HRT	n = 1833 Number of cases = 23 Incidence = 1.25%	n = 3614 Number of cases = 14 Incidence = 0.39%	0.31 (0.16, 0.60)

[a] Patients received various types of HRT.

Concomitant use of **other osteoporosis medications**, including bisphosphonates, calcitonin, or fluorides (excluding marketed raloxifene), was allowed as indicated during the extension phase of the study. Patients taking these medications were allowed to continue concomitant use of the double-blind study medication.

- After 48 months of treatment, there was a statistically significant difference among the three treatment groups in the reported use of non-estrogenic bone-active drugs (p=0.007) and hypolipidemic agents (p=0.001) (Table GGGK.11.30): fewer patients were on non-estrogenic bone-active agents in the raloxifene groups than in the placebo group.

Table GGGK.11.30. Post-baseline Use of Concomitant Medications (Visits 3 through 12) That Might Confound the Interpretation of Efficacy and Safety Analyses (All Randomly Assigned Patients, 48-Month Data)

Class	Placebo N=2576	RLX060 N=2557	RLX120 N=2572	p-value
Estrogens	324 (12.6%)	280 (11.0%)	303 (11.8%)	0.195
Nonestrogenic bone-active drugs	408 (15.6%)	332 (13.0%)	347 (13.5%)	0.007
Hypolipidemic agents	369 (14.3%)	345 (13.5%)	282 (11.0%)	0.001
Progestins/androgens	30 (1.2%)	21 (0.8%)	18 (0.7%)	0.185
Hypoglycemic agents	57 (2.2%)	76 (3.0%)	63 (2.4%)	0.210
Antiseizure medications	39 (1.5%)	54 (2.1%)	56 (2.2%)	0.163
Corticosteroids	653 (25.3%)	680 (26.6%)	686 (26.7%)	0.481

Abbreviations: RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; N = number of randomly assigned patients.

11.4.7.1. New Bone-Active Agents after 36 Months

During the extension phase of the study, there was a statistically significant difference among the three treatment groups in the reported use of non-estrogenic bone-active drugs ($p<0.001$) (Table GGGK.11.31).

Table GGGK.11.31. New Bone-Active Agent Use after 36 Months (All Randomly Assigned Patients, 48-Month Data)

Bone-Active Agent Usage	Placebo N=2576	RLX060 N=2557	RLX120 N=2572	p-value
Visit 1 through Visit 9	299	271	285	0.516
Visit 10 through Visit 12	315	243	253	<0.001
New agent usage after Visit 9	145	89	91	<0.001

Abbreviations: N = number randomized; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day.

11.4.7.2. Subgroup Analyses for Bone-Active Agents

During the fourth year of the study, a subgroup analysis was conducted to assess the effects of permitted non-estrogenic bone-active drug use on vertebral and nonvertebral fracture risk reduction, the percent change in lumber spine and femoral neck BMD, the percent change in osteocalcin, and bone-specific alkaline phosphatase.

- After 48 months of treatment, no statistically significant treatment-by-subgroup interaction was observed among the three treatment groups on these efficacy parameters in those who did versus those who did not use allowed bone-active agents (Table GGGK.11.32).

Table GGGK.11.32. Treatment-by-Subgroup Interaction Analyses for Bone-Active Agent Usage (All Randomly Assigned Patients, 48-Month Data)

Variable	Interaction p-value
Fractures[a]	
New vertebral fracture	0.82
New clinical vertebral fracture	0.83
Total nonvertebral fracture	0.78
Bone Mineral Density[b]	
Lumber spine BMD	0.83
Femoral neck BMD	0.98
Biochemical Markers of Bone Metabolism[c]	
Osteocalcin	0.38
BSAP	0.56

Abbreviations: BMD = bone mineral density; ANOVA = analysis of variance; BSAP = bone-specific alkaline phosphatase.

a For vertebral and nonvertebral fracture risk reduction, subgroup analyses were performed using the chi-square test.

b For BMD, a generalized linear model was used to test the treatment-by-subgroup interaction; subgroup analyses were performed using an ANOVA analysis on unranked data.

c For osteocalcin and BSAP, a generalized linear model was used to test the treatment-by-subgroup interaction; subgroup analyses were performed using an ANOVA analysis with ranked data.

11.4.8. Health Outcomes/Quality of Life Evaluation

11.4.8.1. Overview of Health Outcomes Results

An analysis of patients with and without incident vertebral fractures, disregarding treatment assignments, showed that *the percentages of patients who reported worsened HRQOL at study endpoint* were greater in *the group of patients who experienced incident vertebral fractures* compared with *the group of patients who did not experience incident vertebral fractures.*

There were no statistically significant differences among the three treatment groups in medical resource utilization.

11.4.8.2. Medical Resource Utilization

This section summarizes the medical resource utilization (Table GGGK.11.33). Information on medical resource utilization was collected at every visit from patients in Substudy II who were hospitalized overnight and/or sustained an osteoporotic fracture during the study. This included information on overnight hospitalizations, health care professional visits, medical procedures performed, and laboratory tests beyond the requirements of the protocol, as well as patient activity status (patients reporting paid employment). Chronic stable illnesses such as periodontal disease or routine physical examinations were not included.

- Of the 2,641 Substudy II patients included in this analysis, 638 (24.2%) patients experienced at least one post-baseline overnight hospitalization. A statistically significant difference was not found among the three treatment groups in the number of patients who experienced at least one post-baseline hospitalization. There were 82 (3.1%) patients who were hospitalized due to fracture. Neither the rate of hospitalization nor the length of hospitalization (due to either fracture or other causes) was significantly different among the three treatment groups.
- There were 816 (30.9%) patients who reported having at least one post-baseline health care professional visit beyond what was required by the protocol. A statistically significant difference among the three treatment groups was not observed. In addition, there were no differences among the three treatment groups in the variety of health care visit settings (hospital inpatient, outpatient visit, emergency room, nursing home, and other settings).
- A total of 734 (27.8%) patients received at least one medical procedure beyond the requirements of protocol. A statistically significant difference among the three treatment groups was not observed.
- A total of 472 (17.9%) patients had at least one laboratory test beyond the requirements of the protocol, and a statistically significant difference was not found among the three treatment groups.
- A total of 93 (3.5%) patients reported having paid employment at least once post-baseline.
- No statistically significant difference was found among the three treatment groups in patients who reported having paid employment at least once post-baseline.

Table GGGK.11.33. Summary of Resource Utilization Survey (Randomly Assigned Patients in Substudy II)

RU Module		Placebo (N=887)	RLX060 (N=885)	RLX120 (N=869)	Total (N=2641)	Overall p-value
	Patients who participated in the resource utilization survey	310 (34.95%)	282 (31.86%)	276 (31.76%)	868 (32.87%)	0.269
Hospitalization	Patients with ≥ 1 hospitalization	225 (25.37%)	204 (23.05%)	209 (24.05%)	638 (24.16%)	0.521
	Patients with ≥ 1 hospitalization due to fracture	31 (3.49%)	28 (3.16%)	23 (2.65%)	82 (3.10%)	0.587
Health Care	Patients with ≥ 1 health care visit	289 (32.58%)	263 (29.72%)	264 (30.38%)	816 (30.90%)	0.394
Professional Visit	Patients with ≥ 1 hospital inpatient	209 (23.56%)	191 (21.58%)	195 (22.44%)	595 (22.53%)	0.606
	Patients with ≥ 1 outpatient visit	124 (13.98%)	119 (13.45%)	120 (13.81%)	363 (13.74%)	0.946
	Patients with ≥ 1 emergency room	81 (9.13%)	65 (7.34%)	60 (6.90%)	206 (7.80%)	0.182
	Patients with ≥ 1 nursing home	2 (0.23%)	3 (0.34%)	1 (0.12%)	6 (0.23%)	0.616
Medical Procedures	Patients with ≥ 1 medical procedure	259 (29.20%)	232 (26.21%)	243 (27.96%)	734 (27.79%)	0.371
Laboratory Tests	Patients with ≥ 1 laboratory test	161 (18.15%)	156 (17.63%)	155 (17.84%)	472 (17.87%)	0.959
Activity Status	Patients with paying job	27 (3.04%)	30 (3.39%)	36 (4.14%)	93 (3.52%)	0.443

Abbreviations: RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; RU = resource utilization.

11.4.8.3. Health-Related Quality of Life

A combination of health-related quality of life (HRQOL) questionnaires was administered to patients at selected study sites in countries where appropriate language translations and validations were available (Table GGGK.11.34). A maximum of three instruments was administered at any one site. Three types of HRQOL instruments were used:

- A general health assessment—the Nottingham Health Profile (NHP)
- A preference-based instrument—either the EuroQol (currently named the EQ-5D) or the McMaster Health Utilities Index (MHUI)
- A disease-targeted instrument—either the Osteoporosis Assessment Questionnaire (OPAQ) or the European Foundation for Osteoporosis Quality-of-Life Assessment (EFFO, currently named QualEFFO)

In Substudy II, the assessments were made at the baseline visit (Visit 2), the 12-month visit (Visit 5), the 24-month visit (Visit 7), and the 36-month visit (Visit 9). In Substudy I, questionnaires were administered to patients at selected sites at baseline only. This section contains a summary of the HRQOL results for Substudy II patients.

11.4.8.3.1. Patient Population and Demographics

Table GGGK.11.35 summarizes the patient population by study site or by country for those Substudy II patients who participated in the HRQOL study.

Table GGGK.11.34. Summary of Patient Population (Substudy II Patients Who Participated in the Health-Related Quality of Life Study)

Country	NHP	EQ-5D	MHUI	QualEFFO	OPAQ
Belgium	100	39[a]	—	100	—
France	19	—	—	19	—
Germany	10	—	—	10	—
Italy	74	—	—	74	—
Netherlands	120	120	—	120	—
Norway	319	319	—	—	—
Spain	51	51	—	—	—
Sweden	55	55	—	55	—
Australia	64	—	64	—	64
Canada[b]	118[b]	—	145[b]	—	118[b]
New Zealand	21	—	21	—	21
United Kingdom	73	73	—	73	—
United States	1023	—	387[c]	—	1023
Total	2047	657	617	451	1226

a Belgium had Dutch version of EQ-5D administered at the appropriate sites (39 Dutch-speaking patients).
b Canada had the French and/or English version of the MHUI and the English version only of OPAQ and NHP administered at the appropriate sites (118 English-speaking patients).
c The USA had the MHUI administered at the following sites only 044, 055, 056, 058, 064, 071, 073, 077, 085, 092.
Abbreviations: MHUI = McMaster Health Utilities Index; NHP = Nottingham Health Profile; QOL = Quality of Life; OPAQ = Osteoporosis Assessment Questionnaire; QualEFFO = European Foundation for Osteoporosis Quality-of-Life Assessment.

Table GGGK.11.35 summarizes patient demographics for those Substudy II patients who participated in the HRQOL study. The demographic characteristics did not differ significantly among the three treatment groups.

Table GGGK.11.35. Summary of Patient Demographics (Substudy II Patients Who Participated in the Health-Related Quality of Life Study)

	Treatment Group				
	Placebo	RLX060	RLX120	Total	p-value
NHP					
N	689	685	673	2047	—
Age (Years, mean)	69.2	69.0	68.7	69.0	0.476
Years Postmenopause (mean)	21.6	21.3	21.2	21.4	0.654
BMI (kg/cm^2,mean)	25.6	25.6	25.6	25.6	0.979
Caucasian (%)	96.7	97.5	97.3	97.2	0.686
EQ-5D					
N	222	219	216	657	—
Age (Years, mean)	68.7	69.2	68.2	68.7	0.194
Years Postmenopause (mean)	21.0	22.1	20.3	21.1	0.052
BMI (kg/cm^2,mean)	25.3	25.5	25.4	25.4	0.941
Caucasian (%)	98.6	99.5	99.5	99.2	0.128
MHUI					
N	213	198	206	617	—
Age (Years, mean)	69.8	69.3	69.8	69.6	0.641
Years Postmenopause (mean)	22.0	21.7	22.0	21.9	0.912
BMI (kg/cm^2,mean)	25.8	25.7	25.8	25.8	0.948
Caucasian (%)	97.2	97.5	97.6	97.4	0.207
QualEFFO					
N	153	151	147	451	—
Age (Years, mean)	68.9	68.8	68.7	68.8	0.939
Years Postmenopause (mean)	20.9	20.8	20.5	20.8	0.921
BMI (kg/cm^2,mean)	25.4	25.3	25.5	25.4	0.904
Caucasian (%)	96.7	99.3	99.3	98.4	0.179
OPAQ					
N	411	410	405	1226	—
Age (Years, mean)	69.4	69.0	69.2	69.2	0.638
Years Postmenopause (mean)	22.0	21.1	21.9	21.7	0.244
BMI (kg/cm^2,mean)	25.8	25.8	25.7	25.8	0.953
Caucasian (%)	95.6	96.1	95.8	95.8	0.586

Abbreviations: BMI = body mass index; MHUI = McMaster Health Utilities Index; N = number of patients administered the instrument; NHP = Nottingham Health Profile; OPAQ = Osteoporosis Assessment Questionnaire; QualEFFO = European Foundation for Osteoporosis Quality-of-Life Assessment; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day.

11.4.8.3.2. Nottingham Health Profile

The NHP is a reliable and valid generic health profile instrument designed to assess a subject's perception of physical, social, and emotional distress. The instrument consists of 38 questions divided into six subscales: energy, pain, emotional reactions, sleep, social isolation, and physical mobility. Scores on each subscale range from 0 to 100; low scores indicate low levels of distress (ie, good quality of life). Table GGGK.11.36 summarizes baseline and endpoint results for Substudy II patients who were administered the NHP.

- **At baseline**, there were no statistically significant differences among the three treatment groups for the six NHP subscales.
- **At endpoint**, an overall statistically significant difference among the three treatment groups was observed for the *emotional reaction subscale* (p=0.038). The mean change from baseline to endpoint for the emotional reaction scale showed statistically significant improvement in the raloxifene 120-mg group compared with the placebo group (p=0.011).
- There were no overall or pair-wise statistically significant differences among the three treatment groups for the other five NHP subscales.
- The within-treatment-group mean changes from baseline to endpoint for the six NHP subscales ranged from -1.9 to 1.1. Although some of these within-group changes were statistically significant, all were small.

Table GGGK.11.36. Summary of Nottingham Health Profile Results (Substudy II Patients Administered Nottingham Health Profile)

		Treatment Group			
		Placebo	RLX060	RLX120	Overall p-value
Baseline		(n=673)	(n=660)	(n=651)	
Emotional reaction	Mean	9.5	9.4	9.7	0.893
Energy	Mean	19.4	16.7	19.2	0.154
Physical mobility	Mean	15.6	14.8	15.0	0.702
Pain	Mean	19.0	18.3	18.5	0.784
Sleep	Mean	23.7	23.0	24.1	0.764
Social isolation	Mean	7.0	6.1	7.3	0.254
Endpoint		(n=614)	(n=588)	(n=597)	
Emotional reaction	Mean change	0.8	-0.3	-1.5[a,b]	0.038
Energy	Mean change	-0.8	-0.3	-0.7	0.936
Physical mobility	Mean change	0.6	1.1	0.2	0.591
Pain	Mean change	-1.4	-1.9[a]	-1.9[a]	0.920
Sleep	Mean change	-1.0	0.8	-0.1	0.392
Social isolation	Mean change	-0.7	-0.9	-1.3[a]	0.716

a Significantly different from baseline ($p<0.05$).
b Significantly different from placebo ($p<0.05$).
Abbreviations: n = number of patients administered the instrument; NHP = Nottingham Health Profile; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day.

11.4.8.3.3. EQ-5D (EuroQol)

The EQ-5D, formerly EuroQol (Protocol Attachment GGGK.8 in Appendix 16.1.1), is a generic health status instrument designed for the description and valuation of HRQOL.
It consists of 25 questions divided into two parts. **Part I** includes Question 1 through 7 and **Part II** includes Question 8 through 25.

Only Part I (except Question 6) and the EQ-5D health utility score are analyzed.

Questions 1 through 5 measure the general health status of the patients (including mobility, self-care, usual activities, pain or discomfort, and anxiety or depression). Responses are rated on a three-point Likert scale: "no problem," "some or moderate problem(s)," and "unable or extreme problem." The responses to these questions provide a compact descriptive profile. A health utility score is created based on the scores of the first five questions (EuroQol Group 1996).
The EQ-5D health utility score can range from -0.59 to 1. High scores indicate a good health-related quality of life.

Question 7 is a self-rating of current HRQOL through the use of a visual analog scale (VAS). Scores for Question 7 can range from 0 to 100, with higher values indicating a better health-related quality of life.

Table GGGK.11.37 summarizes the baseline and endpoint results for Substudy II patients who were administered the EQ-5D.

- **At baseline**, there were no statistically significant differences among the three treatment groups for any of the five EQ-5D health status questions, the health state (VAS) score, and the health utility score.
- **At endpoint**, an overall statistically significant difference among the three treatment groups was observed for the mobility question (p=0.012). The percentage of patients with some **problem in mobility** was significantly higher in the raloxifene 60-mg group compared with the placebo group (p=0.036) and the raloxifene 120-mg group (p=0.004).

In addition, there was a statistically significant pair-wise difference between the placebo and raloxifene 120-mg groups for **usual activities and anxiety/depression**.

- The percentage of patients with some problem or unable to perform **usual activities** was significantly higher in the placebo group compared with the raloxifene 120-mg group (p=0.026).
- For the **anxiety/depression question**, the percentage of patients who were moderately or extremely anxious or depressed was significantly higher in the placebo group compared with the raloxifene 120-mg group (p=0.046).

At endpoint, there was a statistically significant overall difference among the three treatment groups for the EQ-5D **health utility score** (p=0.014).

- Patients in the raloxifene 120-mg group showed a statistically significant improvement at endpoint (p=0.003). The mean change from baseline to endpoint in the raloxifene 120-mg

group was significantly better than that of the placebo group (p=0.004). There was, however, no statistically significant difference in the raloxifene 60-mg group compared with placebo.

Except the treatment differences discussed above, no other overall or pair-wise statistically significant differences among the three treatment groups were observed at endpoint.

Table GGGK.11.37. Summary of EQ-5D Results
Substudy II Patients Administered EQ-5D
H3S-MC-GGGK

		Treatment Group			
EQ-5D		**Placebo**	**RLX060**	**RLX120**	**Overall p-value**
Baseline		(n=221)	(n=215)	(n=214)	
Mobility	No Problem (%)	65.9	66.2	71.4	0.382
Self-Care	No Problem (%)	91.2	90.5	92.9	0.668
Usual Activity	No Problem (%)	65.0	66.8	71.0	0.406
Pain/Discomfort	No Problem (%)	29.2	30.3	32.1	0.839
Anxiety/Depression	No Problem (%)	68.4	68.7	71.9	0.722
Health state (VAS)	Mean	72.4	71.9	73.7	0.695
EQ-5D Utility	Mean	0.738	0.725	0.745	0.712
Endpoint		(n=204)	(n=192)	(n=199)	
Mobility	No Problem (%)	66.0	55.9[b]	69.7	0.012
Self-Care	No Problem (%)	89.1	88.7	91.5	0.610
Usual Activity	No Problem (%)	55.4	61.5	66.5[b]	0.074
Pain/Discomfort	No Problem (%)	30.0	32.3	37.3	0.297
Anxiety/Depression	No Problem (%)	66.2	67.7	75.6[b]	0.100
Health State (VAS)	Mean Change	-0.9	1.4	0.2	0.411
EQ-5D Utility	Mean Change	-0.024	0.016	0.040[a,b]	0.014

a Significantly different from baseline (p<0.05).
b Significantly different from placebo (p<0.05).
Abbreviations: n = number; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; VAS = Visual Analog Scale.

11.4.8.3.4. McMaster Health Utilities Index

The MHUI, which generates preference-based measures, is a generic approach to the measurement of health status and assessment of HRQOL. It consists of 15 questions comprising six attributes: sensation, mobility, cognition, self-care, emotion, and pain.
An overall utility score is created based on the attribute scores. The MHUI health utility score can range from 0 to 1; high scores indicate good quality of life.

Table GGGK.11.38 summarizes the baseline and endpoint MHUI results for Substudy II patients who were administered the MHUI.

- **At baseline**, there were no statistically significant differences among the three treatment groups for any of the six attributes and the utility score.

- **At endpoint**, no overall or pair-wise statistically significant differences among the three treatment groups were observed for any of the six MHUI attributes and the MHUI utility score.

Table GGGK.11.38. Summary of McMaster Health Utilities Index Subscales (Substudy II Patients Administered McMaster Health Utilities Index)

		Treatment Group			
		Placebo	RLX060	RLX120	Overall p-value
Baseline		(n=190)	(n=179)	(n=177)	
Sensation	Normal/Req Equip	83.0	80.8	85.3	0.581
Mobility	Able	86.1	87.4	88.1	0.855
Cognition	Normal	63.6	67.4	68.2	0.627
Self-Care	Normal	96.3	96.1	97.2	0.851
Emotion	Generally happy	76.7	77.0	75.4	0.931
Pain	Free of pain	26.6	27.1	27.4	0.941
	Occasional pain	58.5	56.5	59.4	—
	Frequent pain	14.9	16.4	13.2	—
Utility	Mean	0.837	0.847	0.849	0.660
Endpoint		(n=172)	(n=158)	(n=165)	
Sensation	Normal/Req Equip	79.2	83.0	79.3	0.589
Mobility	Able	82.4	80.0	83.1	0.732
Cognition	Normal	59.8	65.3	64.1	0.521
Self-Care	Normal	94.2	94.2	97.8	0.156
Emotion	General happy	75.1	77.2	79.9	0.505
Pain	Free of pain	25.4	24.0	32.1	0.420
	Occasional pain	57.7	59.6	54.9	—
	Frequent pain	16.9	16.4	13.0	—
Utility	Mean change	-0.003	0.000	-0.006	0.899

Abbreviations: n = number; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day.

11.4.8.3.5. European Foundation for Osteoporosis Quality of Life Assessment

The quality of life questionnaire of the European Foundation for Osteoporosis (QualEFFO) is an instrument developed for patients with vertebral osteoporosis, specifically for use as a patient self-assessment in the clinical trial environment. The instrument consists of 54 questions that cover five domains: pain, daily activities, mobility, general health perception, and mental health. The domain scores can range from 0 to 100, with 0 representing the worst possible health status and 100 representing the best possible health status.

Table GGGK.11.39 summarizes the baseline and endpoint QualEFFO results for Substudy II patients who were administered the QualEFFO.

- **At baseline**, there were no statistically significant differences among the three treatment groups for the five QualEFFO domains and the overall QOL score.

- **At endpoint**, no overall or pair-wise statistically significant differences among the three treatment groups were observed for four of the five domains and for the overall QOL score for QualEFFO.

- In the general health perception domain, there was a statistically significant pair-wise difference for the raloxifene 120-mg group compared with raloxifene 60-mg group (p=0.050), with improved HRQOL in the raloxifene 120-mg group compared with the raloxifene 60-mg group, but there was no statistically significant overall difference among the three treatment groups in this domain.
- The *within-treatment-group mean changes* from baseline to endpoint for the five domains and the overall QOL score ranged from -4.9 to 4.4. Although some of these within-group changes were statistically significant, all were small.

Table GGGK.11.39. Summary of the European Foundation for Osteoporosis Quality of Life Questionnaire Results (Substudy II Patients Administered the European Foundation for Osteoporosis Quality of Life Questionnaire)

		Treatment Group			
		Placebo	RLX060	RLX120	Overall p-value
Baseline		(n=152)	(n=151)	(n=145)	
Pain	Mean	59.4	59.7	62.2	0.596
Daily Activity	Mean	85.4	86.8	86.2	0.764
Mobility	Mean	70.4	74.5	73.7	0.185
General Health	Mean	48.7	51.2	51.5	0.406
Mental Health	Mean	69.1	68.9	72.1	0.276
Overall QOL	Mean	63.6	64.4	67.7	0.192
Endpoint		(n=133)	(n=130)	(n=130)	
Pain	Mean change	2.9	4.4[a]	1.1	0.475
Daily Activity	Mean change	-4.5[a]	-2.1	-1.6	0.166
Mobility	Mean change	-2.6	-4.2[a]	-2.5[a]	0.575
General Health	Mean change	-0.2	-4.6[a]	0.1	0.089
Mental Health	Mean change	-1.5	-3.1[a]	-1.2	0.581
Overall QOL	Mean change	-0.9	-1.3	-4.9[a]	0.324

a Significantly different from baseline (p<0.05).
Abbreviations: n = number; QOL = quality of life score; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day.

11.4.8.3.6. Osteoporosis Assessment Questionnaire

The OPAQ is an osteoporosis disease-targeted questionnaire designed to assess HRQOL in osteoporosis patients. It measures osteoporosis-specific health status in four major dimensions: physical function, emotional status, symptoms, and social interactions. The dimension score can range from 0 to 100, with 0 representing the worst possible health status and 100 representing the best possible health status.

Table GGGK.11.40 summarizes the baseline and endpoint OPAQ results for Substudy II patients who were administered the OPAQ.

- **At baseline**, there were **no statistically significant differences** among the three treatment groups for the four OPAQ dimensions and the overall HRQOL score.
- **At endpoint**, an overall statistically significant difference among the three treatment groups was observed for the **emotional status dimension** (p=0.036).
  - The mean change from baseline to endpoint for the emotional status dimension showed statistically significant worsening in the raloxifene 60-mg group compared with the placebo group (p=0.039) and compared with the raloxifene 120-mg group (p=0.017).
  - The raloxifene 120-mg group was not significantly different from placebo in the emotional status dimension.
- There were no overall or pair-wise statistically significant differences among the three treatment groups for the other three OPAQ dimensions and the overall HRQOL score.
- The within-treatment-group mean changes from baseline to endpoint for the five domains and the overall QOL score ranged from -2.3 to 2.0. Although some of these within-group changes were statistically significant, all were small.

Table GGGK.11.40. Summary of Osteoporosis Assessment Questionnaire Results (Substudy II Patients Administered Osteoporosis Assessment Questionnaire)

		Treatment Group			
		Placebo	RLX060	RLX120	Overall p-value
Baseline		(n=407)	(n=404)	(n=398)	
Physical Function	Mean	87.0	87.4	86.6	0.748
Emotional Status	Mean	70.7	71.7	69.9	0.236
Symptoms	Mean	66.0	66.0	64.9	0.672
Social Interaction	Mean	62.2	64.0	62.0	0.120
Overall HRQOL	Mean	78.2	79.4	79.5	0.501
Endpoint		(n=371)	(n=365)	(n=361)	
Physical Function	Mean change	-1.1	-1.1	-1.6[a]	0.795
Emotional Status	Mean change	-0.7	-2.3[a,b,c]	-0.4	0.036
Symptoms	Mean change	-0.2	0.5	0.9	0.572
Social Interaction	Mean change	2.0[a]	1.0	1.7[a]	0.572
Overall HRQOL	Mean change	-1.6	-0.8	-0.8	0.762

a Significantly different from baseline (p<0.05).
b Significantly different from placebo (p<0.05).
c Significantly different from RLX120 (p<0.05).
Abbreviations: HRQOL = health-related quality of life; n = number; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day.

11.4.8.3.7. Impact of Incident Vertebral Fractures on Health-Related Quality of Life

For evaluation of impact of *incident vertebral fractures* on patients' HRQOL, data were analyzed by the **presence or absence of at least one incident vertebral fracture**, not by the treatment assignments. Patients who had at least one or none incident vertebral fracture during the first 36-months were pooled across the three treatment groups.

Patients whose HRQOL status was significantly worse at endpoint compared with baseline were defined by the following:

- **For HRQOL data measured on a categorical scale** (EQ-5D Questions 1 through 5, and MHUI Mark II attributes), patients with worsening HRQOL were those whose rating at endpoint dropped by at least one category (for example, from "no problem" to "some problem," or from "free of pain" to "occasional pain").
- **For HRQOL data summarized by a composite score** (NHP subscales, EQ-5D and MHUI utility scores, QualEFFO domains and OPAQ dimensions, and overall QOL VAS), patients with worsening HRQOL were those whose scores at endpoint were worse than baseline by at least one standard deviation (for NHP, "worse" is indicated by a "higher" score; for the other four questionnaires, "worse" is indicated by a "lower" score). The standard deviation at baseline is usually about 20 on a 0 to 100 scale, or 0.20 on a 0.0 to 1.0 scale.

The percentages of patients in each group (the group with incident vertebral fractures and the group without incident vertebral fractures) whose endpoint HRQOL status was significantly worse than baseline were compared using a CMH test adjusted for age and country. Table GGGK.11.41 summarizes the results of this analysis.

- Of the 31 HRQOL scales from the five questionnaires examined, except for QualEFFO general health and OPAQ social interaction, **the percentage of patients who reported their HRQOL being significantly worse at endpoint was higher for patients with at least one incident vertebral fracture** (compared to those without incident vertebral fractures).
  - The differences were statistically significant for the following scales: NHP pain and physical mobility (both $p<0.001$); EQ-5D mobility ($p=0.006$); MHUI self-care ($p=0.005$), pain ($p=0.005$), and utility ($p=0.003$); QualEFFO daily activity ($p<0.001$), mobility ($p=0.003$) and mental health ($p<0.001$); and OPAQ physical function, symptoms, and overall HRQOL (all $p<0.001$). Across the five questionnaires, differences were consistently found in pain (NHP, MHUI, and OPAQ), physical function (NHP, EQ-5D, QualEFFO, and OPAQ), and self-care or daily activity (MHUI and QualEFFO).

Table GGGK.11.41. Impact of Incident Vertebral Fractures on HRQOL (Substudy II Patients Administered QOL Questionnaires)

	Without incident fractures			With incident fractures			
HRQOL Scale	**N**	**n**	**Percent**	**N**	**n**	**Percent**	**p-value[a]**
NHP							
Emotional Reaction	1459	94	6.4	255	23	9.0	0.436
Energy	1492	126	8.4	266	30	11.3	0.407
Pain	1451	96	6.6	255	33	12.9	<0.001
Physical Mobility	1464	135	9.2	263	48	18.3	<0.001
Sleep	1484	122	8.2	264	28	10.6	0.475
Social Isolation	1473	120	8.1	265	28	10.6	0.326
EQ-5D							
Mobility	494	61	12.3	92	22	23.9	0.006
Self-Care	493	31	6.3	93	9	9.7	0.431
Usual Activity	492	73	14.8	92	21	22.8	0.064
Pain/Discomfort	495	60	12.1	93	12	12.9	0.950
Anxiety/Depression	492	62	12.6	92	16	17.4	0.579
Health State (VAS)	470	62	13.2	89	14	15.7	0.466
EQ-5D Utility	481	28	5.8	89	9	10.1	0.220
MHUI							
Sensation	404	42	10.4	63	8	12.7	0.667
Mobility	418	45	10.8	67	11	16.4	0.169
Cognition	419	57	13.6	69	12	17.4	0.535
Self-Care	423	10	2.4	69	6	8.7	0.005
Emotion	422	43	10.2	69	10	14.5	0.398
Pain	421	66	15.7	68	20	29.4	0.005
MHUI Utility	390	32	8.2	60	13	21.7	0.003
QualEFFO							
Pain	308	23	7.5	76	8	10.5	0.330
Daily Activity	314	27	8.6	76	18	23.7	<0.001
Mobility	315	20	6.3	76	13	17.1	0.003
General Health	310	46	14.8	75	10	13.3	0.918
Mental Health	314	28	8.9	76	20	26.3	<0.001
Overall QOL (VAS)	296	49	16.6	71	15	21.1	0.300
OPAQ							
Physical Function	926	61	6.6	148	31	20.9	<0.001
Emotional Status	926	77	8.3	151	20	13.2	0.062
Symptoms	940	52	5.5	151	22	14.6	<0.001
Social Interaction	937	85	9.1	151	12	7.9	0.592
Overall HRQOL	926	132	14.3	153	39	25.5	<0.001

a p-values are from a CMH test adjusted for age and country.

Abbreviations: CMH = Cochran-Mantel-Haenszel; HRQOL = health-related quality of life; MHUI = McMaster Health Utilities Index; n = number of patients who had significantly worse HRQOL at endpoint compared with baseline; percent = 100 x (n / N); NHP = Nottingham Health Profile; OPAQ = Osteoporosis Assessment Questionnaire; QOL = Quality of Life; QualEFFO = European Foundation for Osteoporosis Quality-of-Life Assessment; N = number of patients who completed questionnaire; VAS = visual analog scale.

11.4.9. Summary of Pharmacokinetic Evaluation

Steady-state concentrations of *raloxifene in plasma* and *total raloxifene in hydrolyzed plasma (TRHP)* were evaluated in patients following 3, 6, 12, 18, 24, and 36 months of raloxifene treatment. Data in Figure GGGK.11.12 and Figure GGGK.11.13 and Table GGGK.11.42 and Table GGGK.11.43 were obtained from patients included in the pharmacokinetic study group population (ie, those for whom plasma concentrations, time of dose, and time of sample draw were available). *Total raloxifene in hydrolyzed plasma* represents the combined concentrations of *raloxifene* and all known *metabolites* (raloxifene-4'-glucuronide, raloxifene-6-glucuronide, and raloxifene-6,4'-diglucuronide).
Separate analytical methods were used to analyze raloxifene plasma and TRHP concentrations.

11.4.9.1. Raloxifene Concentrations

Figure GGGK.11.12 shows the steady-state raloxifene concentrations for individual patients in each raloxifene treatment group.

- Raloxifene plasma concentrations for most patients tended to be consistent over time.
- The overall mean steady-state raloxifene plasma concentrations were 1.09 ng/mL and 2.05 ng/mL for 60-mg and 120-mg raloxifene doses, respectively (Table GGGK.11.42). The overall mean steady-state concentration for the 120-mg dose group was 1.88 times the steady-state value obtained for the 60-mg dose group.
- Raloxifene concentrations were highly variable between individuals. The within-patient variability was 40% and was similar to that observed in a previous study (H3SLC-GGGV [Raloxifene: Gender-Specific Metabolism, Within-Subject Variability, and Analytical Methods Development]).

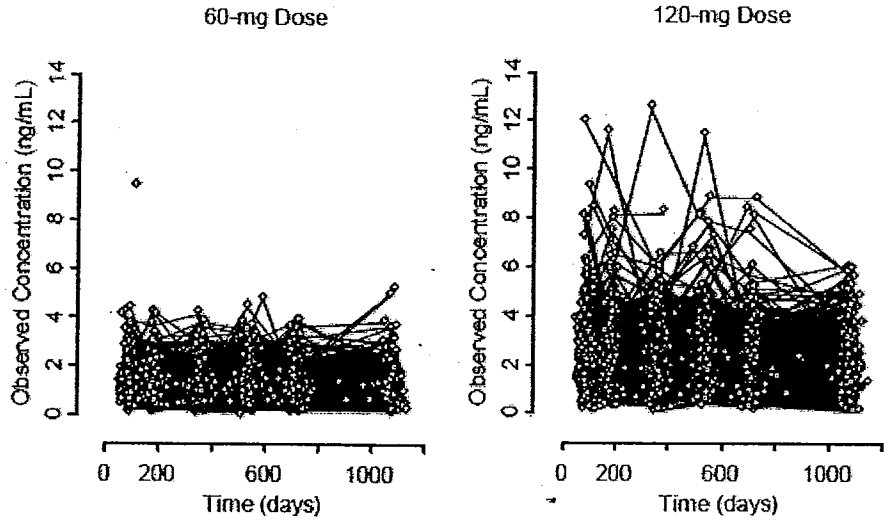

Figure GGGK.11.12. Observed Raloxifene Plasma Concentration versus Time for Individual Patients (Pharmacokinetic Study Population)

Table GGGK.11.42. Mean Observed Plasma Raloxifene Concentrations at Steady-State (Pharmacokinetic Study Population)

Raloxifene Dose	3 Months	6 Months	12 Months	18 Months	24 Months	36 Months	Overall [a]
60 mg							
Mean (ng/mL)	1.17	1.06	1.04	1.19	1.04	1.06	1.09
CV (%) [b]	58.6	54.9	55.3	55.2	55.4	56.8	56.4
n [c]	819	780	764	739	721	671	4498 [d]
120 mg							
Mean (ng/mL)	2.20	2.02	2.01	2.21	1.91	1.96	2.05
CV (%) [b]	56.4	56.8	57.1	54.8	56.2	51.3	55.9
n [c]	792	770	757	728	703	665	4416 [d]

[a] May include analytical results obtained at an unscheduled visit.

[b] Statistics for coefficient of variation (CV) are from UNIVARIATE procedure in SAS®.

[c] Number of patients for whom plasma concentrations, as well as dose and sample draw times, were available.

[d] The overall number (n) represents the total number of blood samples and may be greater than the sum of patient number for 3, 6, 12, 18, 24, and 36 months.

11.4.9.2. Total Raloxifene in Hydrolyzed Plasma (TRHP) Concentrations

Figure GGGK.11.13 shows the steady-state TRHP concentrations for individual patients in each raloxifene treatment group.

- Total raloxifene in hydrolyzed plasma concentrations for most patients tended to be consistent over time.
- The overall mean steady-state TRHP concentrations were 188.0 and 323.5 ng/mL for doses of 60-mg and 120-mg raloxifene, respectively (Table GGGK.11.43). The overall mean steady-state concentration for the 120-mg dose group was 1.72 times the steady-state value obtained for the 60-mg dose group. These results are consistent with findings from a previous study (H3S-LCGGHR: Raloxifene Hydrochloride: Dose Linearity).
- Total raloxifene in hydrolyzed plasma concentrations were highly variable between individuals. The within-patient coefficient of variation was 47% and was similar to that observed in Study GGGV.

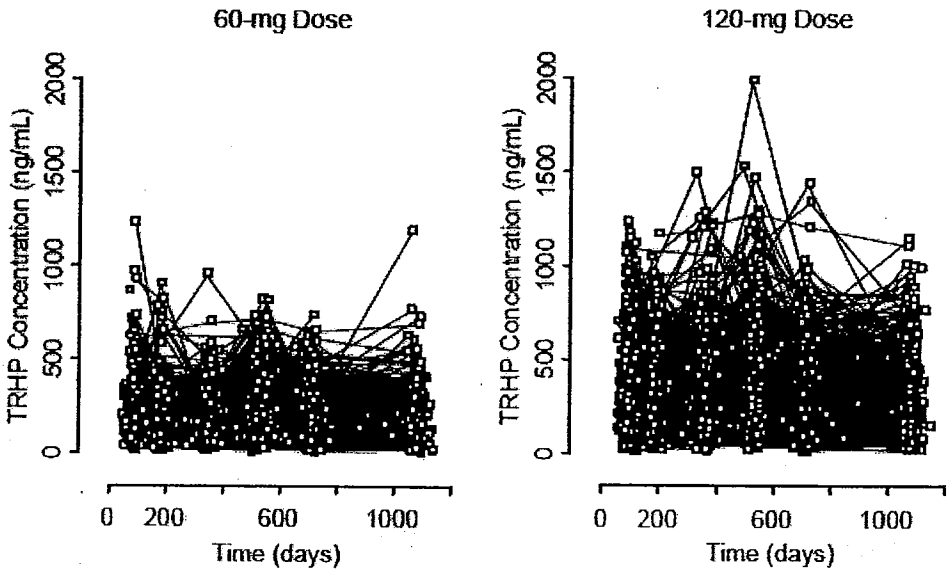

Figure GGGK.11.13. Total Raloxifene in Hydrolyzed Plasma Concentration versus Time for Individual Patients (Pharmacokinetic Study Population)

Table GGGK.11.43. Mean Total Raloxifene in Hydrolyzed Plasma Steady-State Plasma Concentrations (Pharmacokinetic Study Population)

Raloxifene Dose	3 Months	6 Months	12 Months	18 Months	24 Months	36 Months	Overall [a]
60 mg							
Mean (ng/mL)	197.1	170.8	178.7	221.8	176.7	182.7	188.0
CV (%) [b]	63.8	65.2	59.4	60.0	62.6	65.1	63.4
n [c]	818	779	763	739	720	671	4494 [d]
120 mg							
Mean (ng/mL)	338.0	297.8	318.3	376.6	299.0	309.2	323.5
CV (%) [b]	63.1	60.2	63.2	62.1	65.4	60.3	63.2
n [c]	791	768	756	728	703	665	4412 [d]

[a] May include analytical results obtained at an unscheduled visit.
[b] Statistics for coefficient of variation (CV) are from UNIVARIATE procedure in SAS®.
[c] Number of patients for whom plasma concentrations, as well as dose and sample draw times, were available.
[d] The overall number (n) represents the total number of blood samples and may be greater than the sum of patient number for 3, 6, 12, 18, 24, and 36 months.

11.4.10. Efficacy Conclusions

- Forty-eight months of treatment with raloxifene decreased the incidence of all cases of breast cancer by 62% and invasive breast cancer by 73% (compared with placebo). This reduction was largely due to the effect of raloxifene on cases of invasive ER+ breast cancers, which were reduced 83%.
- In contrast, treatment with raloxifene had no effect on ER- breast cancers.
    - This decreased incidence is consistent with the estrogen-antagonist activity of raloxifene on the breast.
- A decrease in the relative risk of invasive breast cancer was evident by the second year of treatment ($p<.001$).
- Overall, 93 patients would need to receive raloxifene to prevent one new case of invasive breast cancer.
- Treatment with raloxifene decreased the rate of new vertebral fractures in osteoporotic, postmenopausal women with and without prevalent vertebral fractures. It also increased BMD and decreased biochemical markers of bone turnover significantly relative to placebo and baseline without an effect on nonvertebral fractures.
- Raloxifene reduced total cholesterol, LDL-C, LDL-C/HDL-C ratio, and fibrinogen, while not affecting HDL-C, or HbA1c. An analysis of triglyceride changes revealed no clinically relevant changes.
- Raloxifene did not appear to have a significant effect on cognitive or neuro-psychomotor function after 48 months of treatment.

The sponsor concludes that 48 months of raloxifene is an effective treatment for osteoporosis that significantly reduces the risk of breast cancer in postmenopausal women with osteoporosis. In addition, raloxifene has positive effects on several markers of cardiovascular disease risk.

12. Safety Evaluation

12.1. Extent of Exposure

There were no statistically significant differences among the three treatment groups in exposure to study drug. Exposure in patient-years for each treatment group was as follows:

Treatment Group Exposure

Placebo	8,346 patient-years
Raloxifene 60 mg/day	8,423 patient-years
Raloxifene 120 mg/day	8,523 patient-years

Total patient exposure was 25,292 patient-years, of which 16,946 patient-years were exposure to raloxifene.
The median exposure for all randomly assigned patients was 1,441 days for all three treatment groups (Table GGGK.12.1). The mean exposure was 1,198 days with no statistically significant difference among the three treatment groups.

Table GGGK.12.1. Exposure to Study Drug (Comparison of Treatment Groups, All Randomized Patients)

Variable	PLACEBO (N=2576)	RLX060 (N=2557)	RLX120 (N=2572)	Total (N=7705)
Exposure (days)				
No. Patients	2576	2557	2572	7705
Mean	1182.61	1202.40	1209.54	1198.17
Median	1439.00	1442.00	1442.00	1441.00
Standard Dev.	471.57	474.00	471.84	472.55
Minimum	1.00	1.00	1.00	1.00
Maximum	1685.00	1653.00	1671.00	1685.00

SOURCE IS EMP.H3SP.SASMACRO(DESM1) CP007 01C
DATA FROM EMP.SAS.H3SM.HCGGGKSC.FINAL
Note: Exposure = Date of Final Dose - Date of First Dose + 1.
Note: Date of final dose is date of cutoff visit for ongoing patients.
KDES0001

12.2. Adverse Events

12.2.1. Brief Summary of Adverse Events

The data in this section are based on data contained in the clinical trial database up to and including the 48-month visit for all randomly assigned patients.
At each visit, patients were questioned regarding the occurrence of adverse events. All events and their severity as reported by patients were recorded in the case report forms, regardless of potential causality or severity. Adverse events were classified as mild, moderate, or severe. For information on the intent-to-treat (ITT) analysis plan for this study, refer to Protocol(g) (Appendix 16.1.1).
Abnormal laboratory test results, vital signs, or other diagnostic procedure results which

the investigator thought to be clinically significant also were recorded as adverse events, regardless of potential causality. Adverse events were first coded at each study site using classification terms from the World Health Organization Adverse Reaction Terminology (WHOART) dictionary and subsequently automatically coded using the Coding Symbol and Thesaurus for Adverse Reaction Terminology (COSTART).
Table GGGK.12.2 contains an overview of adverse events reported during the study.

Table GGGK.12.2. Overview of Adverse Events
Number and Percentage of Patients
All Randomized Patients
H3S-MC-GGGK 48-Month Data

	Number (%) of Patients		
Adverse Event[a]	Placebo n=2576	RLX060 n=2557	RLX120 n=2572
Deaths	36 (1.4)	23 (0.9)	41 (1.6)
Serious adverse events	794 (30.8)	732 (28.6)	744 (28.9)
Discontinuations due to an adverse event	285 (11.1)	327 (12.8)	298 (11.6)
Treatment-emergent adverse events	2425 (94.1)	2417 (94.5)	2449 (95.2)

Abbreviations: RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day.
[a] Patients may be counted in more than one category.

12.2.2. Display and Analysis of Adverse Events

Adverse events have been categorized and will be discussed in the following order: 1) secondary conditions, and 2) treatment-emergent adverse events.

12.2.2.1. Secondary Conditions

Secondary conditions are ongoing medical conditions that were present prior to the Visit 2 date (randomization). Of the 7705 randomly assigned patients, 6978 (90.6%) reported at least one secondary condition (2344 [91.0 %] in the placebo group, 2302 [90.0 %] patients in the raloxifene hydrochloride (HCl) 60-mg group, and 2332 [90.7%] patients in the raloxifene 120-mg group). There were no overall statistically significant differences in the proportion of patients reporting at least one secondary condition among the three treatment groups.
When analyzed by body system, a statistically significant difference among the three treatment groups was observed in the reporting of secondary conditions of the respiratory system (p=0.0090) and metabolic and nutritional disorders (p=0.041). A summary of secondary conditions by body system is located in Table GGGK.14.3 (Section 14.3.1). A total of 1390 (18.0%) patients reported secondary conditions in the respiratory system (513 [19.9%] patients in the placebo group, 430 [16.8%] patients in the raloxifene 60-mg group, and 447 [17.4%] patients in the raloxifene 120-mg group). Sinusitis was the most frequently reported secondary condition in the respiratory system occurring in 358 (4.6%) patients, with no statistically significant difference among the three treatment groups. Statistically significant differences were noted among the three treatment groups for emphysema and respiratory disorder.

12.2.2.2. Treatment-Emergent Adverse Events

Treatment-emergent adverse events are events that began after randomization or were

preexisting and worsened in severity after randomization. Table GGGK.12.3 summarizes the COSTART terms for treatment-emergent adverse events overall and by body system in order of decreasing frequency. This summary only includes treatment-emergent adverse events that occurred in ε5% of drug-treated patients.

Of the 7705 randomly assigned patients, 7291 (94.6%) reported at least one treatment-emergent adverse event (2425 [94.1%] the placebo group, 2417 [94.5%] in the raloxifene 60-mg group, and 2449 [95.2%] patients in the raloxifene 120-mg group). There were no differences in the proportion of patients reporting at least one treatment-emergent adverse event among the three treatment groups or in the pooled raloxifene group compared with the placebo group (Table GGGK.12.3).

Significant differences among the three treatment groups were observed overall in the reporting of treatment-emergent adverse events in the cardiovascular system (p=0.015). Cardiovascular system, treatment-emergent, adverse events were reported by 981 (38.1%) patients in the placebo group, 983 (38.4%) patients in the raloxifene 60-mg group, and 1072 (41.7%) patients in the raloxifene 120-mg group. The raloxifene 120-mg group reported cardiovascular system treatment-emergent adverse events more frequently compared with both the placebo group (p<0.01) and the raloxifene 60-mg group (p<0.05). This difference was due primarily to a greater incidence of vasodilatation in the raloxifene groups (183 [7.1%] patients in the placebo group, 272 [10.6%] patients in the raloxifene 60-mg group and 319 [12.4%] patients in the raloxifene 120-mg group). When vasodilatation is excluded, there is no statistically significant difference among the three treatment groups in the reporting of cardiovascular system adverse events.

Section 14 contains a listing of treatment-emergent adverse events by body system. Appendix 16.2.1 contains a by-patient listing of all adverse events.

Table GGGK.12.3. Summary of Common Treatment-Emergent Adverse Events Overall and by Body System (>5%)
All Randomly Assigned Patients
H3S-MC-GGGK 48-Month Data

Body System: OVERALL

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
PATIENTS WITH >= 1 TESS	2425 (94.1%)	2417 (94.5%)	2449 (95.2%)		.178

Body System: BODY AS A WHOLE

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
PATIENTS WITH >= 1 TESS	2112 (82.0%)	2092 (81.8%)	2107 (81.9%)		.897
ACCIDENTAL INJURY	1129 (43.8%)	1141 (44.6%)	1055 (41.0%)ad	.023	.397
SURGICAL PROCEDURE	966 (37.5%)	941 (36.8%)	951 (37.0%)		.600
PAIN	804 (31.2%)	818 (32.0%)	810 (31.5%)		.541
BACK PAIN	687 (26.7%)	660 (25.8%)	662 (25.7%)		.399
INFECTION	463 (18.0%)	466 (18.2%)	501 (19.5%)		.349
FLU SYNDROME	360 (14.0%)	415 (16.2%)a	429 (16.7%)b	.017	.005
ABDOMINAL PAIN	293 (11.4%)	264 (10.3%)	315 (12.2%)		.911
HEADACHE	255 (9.9%)	265 (10.4%)	257 (10.0%)		.702
ASTHENIA	235 (9.1%)	236 (9.2%)	253 (9.8%)		.559
CHEST PAIN	207 (8.0%)	212 (8.3%)	208 (8.1%)		.817
ALLERGIC REACTION	143 (5.6%)	135 (5.3%)a	147 (5.7%)		.923
FEVER	134 (5.2%)	123 (4.8%)	131 (5.1%)		.636

Body System: CARDIOVASCULAR SYSTEM

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
PATIENTS WITH >= 1 TESS	981 (38.1%)	963 (37.6%)	1072 (41.7%)bd	.015	.093
HYPERTENSION	289 (11.2%)	231 (9.0%)b	264 (10.3%)	.034	.032
VASODILATATION	183 (7.1%)	272 (10.6%)c	319 (12.4%)cd	<.001	<.001

Body System: DIGESTIVE SYSTEM

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
PATIENTS WITH >= 1 TESS	1102 (42.8%)	1073 (42.0%)	1152 (44.8%)		.615
NAUSEA	246 (9.5%)	259 (10.1%)	269 (10.5%)		.305
DIARRHEA	217 (8.4%)	225 (8.8%)	226 (8.8%)		.587
DYSPEPSIA	224 (8.7%)	189 (7.4%)	203 (7.9%)		.106
CONSTIPATION	193 (7.5%)	181 (7.1%)	201 (7.8%)		.944
GASTROINTESTINAL DISORDER	160 (6.2%)	144 (5.6%)	159 (6.2%)		.597
VOMITING	142 (5.5%)	148 (5.8%)	155 (6.0%)		.483

Body System: HEMIC AND LYMPHATIC SYSTEM

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
PATIENTS WITH >= 1 TESS	398 (15.5%)	344 (13.5%)	369 (14.3%)		.068
PURPURA	257 (10.0%)	208 (8.1%)a	223 (8.7%)	.059	.022

Body System: METABOLIC AND NUTRITIONAL DISORDERS

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
PATIENTS WITH >= 1 TESS	639 (24.8%)	570 (22.3%)	615 (23.9%)		.097
PERIPHERAL EDEMA	156 (6.1%)	182 (7.1%)	203 (7.9%)a	.047	.026
HYPERCHOLESTEREMIA	147 (5.7%)	82 (3.2%)c	76 (3.0%)c	<.001	<.001

Body System: MUSCULOSKELETAL SYSTEM

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
PATIENTS WITH >= 1 TESS	1102 (42.8%)	1157 (45.2%)	1124 (43.7%)		.158
ARTHRALGIA	444 (17.2%)	468 (18.3%)	486 (18.9%)		.143
MYALGIA	237 (9.2%)	225 (8.8%)	228 (8.9%)		.592
ARTHRITIS	239 (9.3%)	221 (8.6%)	219 (8.5%)		.307
LEG CRAMPS	154 (6.0%)	234 (9.2%)c	218 (8.5%)c	<.001	<.001
ARTHROSIS	162 (6.3%)	156 (6.1%)	151 (5.9%)		.599

Body System: NERVOUS SYSTEM

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
PATIENTS WITH >= 1 TESS	1047 (40.6%)	983 (38.4%)	1057 (41.1%)		.462
DIZZINESS	260 (10.1%)	229 (9.0%)	244 (9.5%)		.219
DEPRESSION	197 (7.6%)	188 (7.4%)	202 (7.9%)		.946
ANXIETY	167 (6.5%)	163 (6.4%)	189 (7.3%)		.530
INSOMNIA	165 (6.4%)	141 (5.5%)	157 (6.1%)		.300
VERTIGO	121 (4.7%)	114 (4.5%)	131 (5.1%)		.877

Body System: RESPIRATORY SYSTEM

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
PATIENTS WITH >= 1 TESS	1033 (40.1%)	1012 (39.6%)	1071 (41.6%)		.654
RHINITIS	307 (11.9%)	300 (11.7%)	300 (11.7%)		.778
COUGH INCREASED	289 (11.2%)	297 (11.6%)	292 (11.4%)		.730
BRONCHITIS	266 (10.3%)	291 (11.4%)	306 (11.9%)		.085
SINUSITIS	223 (8.7%)	230 (9.0%)	211 (8.2%)		.931
PHARYNGITIS	159 (6.2%)	160 (6.3%)	193 (7.5%)		.238
DYSPNEA	177 (6.9%)	145 (5.7%)	168 (6.5%)		.192
PNEUMONIA	136 (5.4%)	123 (4.8%)	128 (5.0%)		.381

Body System: SKIN AND APPENDAGES

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
PATIENTS WITH >= 1 TESS	726 (28.2%)	731 (28.6%)	725 (28.2%)		.851
RASH	194 (7.5%)	183 (7.2%)	187 (7.3%)		.614
SKIN DISORDER	123 (4.8%)	133 (5.2%)	129 (5.0%)		.526

Body System: SPECIAL SENSES

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
PATIENTS WITH >= 1 TESS	534 (20.7%)	536 (21.0%)	516 (20.1%)		.823
CATARACT SPECIFIED	141 (5.5%)	135 (5.3%)	122 (4.7%)		.386

Body System: UROGENITAL SYSTEM

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
PATIENTS WITH >= 1 TESS	1102 (42.8%)	1055 (41.3%)	1044 (40.6%)		.119
URINARY TRACT INFECTION	267 (10.4%)	264 (10.3%)	245 (9.5%)		.544
VAGINITIS	216 (8.4%)	196 (7.7%)	203 (7.9%)		.355
CYSTITIS	137 (5.3%)	144 (5.6%)	146 (5.7%)		.543
UTERINE DISORDER	80 (3.1%)	108 (4.2%)a	130 (5.1%)c	.002	.001

NOTE: Chi-square tests were used when total count >= 10, else Fisher's exact test was used. Pairwise comparisons were performed when the overall or pooled raloxifene comparison was significant ($p < 0.05$).
a - pairwise comparison statistically significant ($p < 0.05$) different from placebo
b - pairwise comparison statistically significant ($p < 0.01$) different from placebo
c - pairwise comparison statistically significant ($p < 0.001$) different from placebo
d - pairwise comparison of RLX060 statistically significant ($p < 0.05$) different from RLX120

Data: RMP.SAS.H3SM.MCGGGKSC.FINAL
Source: RMP.H3SSE4TR.SASPGM(ARTCS001) X7164 04JAN01
Output: RMP.H3SO.GGGK.FINAL(ARTCSeYN)

12.2.2.2.1. Analyses by Event Term

Table GGGK.12.4 summarizes the COSTART terms for which there was a statistically significant difference in the proportion of patients reporting the treatment-emergent adverse event either 1) among the three treatment groups or 2) between the pooled raloxifene group and the placebo group.

- Statistically significant **treatment-emergent adverse events** were categorized (by COSTART term) into one of the following groups: *Greater Incidence with Raloxifene and Potentially Clinically Relevant* or *Lower Incidence with Raloxifene and Potentially Clinically Relevant*.
  - **Breast analyses** (excluding breast cancer, which is presented as an efficacy endpoint are reported separately (Section 12.3.3.3).
  - Specific **uterine analyses**, which exclude women with a prior hysterectomy, and **venous thromboembolism analyses** are also reported separately (Section 12.3.3.4 and Section 12.3.3.6, respectively).

Section 14.3.5 contains a listing of actual terms for all statistically significant, treatment-emergent, adverse event COSTART terms (Table GGGK.14.12); a summary of preexisting conditions that worsened in severity for statistically significant, treatment-emergent, adverse events (Table GGGK.14.13); a summary of statistically significant treatment-emergent adverse events that were first reported post-baseline (Table GGGK.14.14); and a summary of onset of statistically significant treatment-emergent adverse events by 6-month visit interval (Table GGGK.14.15).

Table GGGK.12.4. Summary of Statistically Significant COSTART Terms Greater Incidence and Lower Incidence in Raloxifene-Treated Patients (All Randomly Assigned Patients, 48-Month Data)

HIGHER INCIDENCE IN POOLED RALOXIFENE GROUPS

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
FLU SYNDROME	360 (14.0%)	415 (16.2%)a	429 (16.7%)b	.017	.005
VASODILATATION	183 (7.1%)	272 (10.6%)c	319 (12.4%)cd	<.001	<.001
LEG CRAMPS	154 (6.0%)	234 (9.2%)c	218 (8.5%)c	<.001	<.001
PERIPHERAL EDEMA	158 (6.1%)	182 (7.1%)	203 (7.9%)a	.047	.026
UTERINE DISORDER	80 (3.1%)	108 (4.2%)a	130 (5.1%)c	.002	.001
ENDOMETRIAL DISORDER	27 (1.0%)	48 (1.9%)a	42 (1.6%)	.044	.017
DIABETES MELLITUS	17 (0.7%)	38 (1.5%)b	36 (1.4%)b	.011	.003
DRY MOUTH	20 (0.8%)	30 (1.2%)	36 (1.4%)a	.058	.044
CERVIX NEOPLASM	17 (0.7%)	30 (1.2%)	28 (1.1%)	.133	.047
PHLEBITIS	13 (0.5%)	21 (0.8%)	29 (1.1%)a	.046	.031
DEEP THROMBOPHLEBITIS	6 (0.2%)	20 (0.8%)a	23 (0.9%)b	.023	.007
REDUCTION	5 (0.2%)	16 (0.6%)a	18 (0.7%)b	.025	.008
HYPERGLYCEMIA	7 (0.3%)	9 (0.4%)	20 (0.8%)ad	.017	.075
FEMALE LACTATION	0 (0.0%)	4 (0.2%)a	9 (0.3%)b	.009	.011
ABNORMAL STOOLS	1 (0.0%)	7 (0.3%)a	2 (0.1%)	.042	.116
HEMATEMESIS	0 (0.0%)	3 (0.1%)	7 (0.3%)b	.025	.025
HYPERVOLEMIA	0 (0.0%)	1 (0.0%)	5 (0.2%)a	.028	.188

LOWER INCIDENCE IN POOLED RALOXIFENE GROUPS

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
HYPERTENSION	289 (11.2%)	231 (9.0%)b	264 (10.3%)	.034	.032
PURPURA	257 (10.0%)	208 (8.1%)a	223 (8.7%)	.059	.022
HYPERCHOLESTEREMIA	147 (5.7%)	82 (3.2%)c	76 (3.0%)c	<.001	<.001
BREAST NEOPLASM	110 (4.3%)	81 (3.2%)a	77 (3.0%)a	.025	.007
PRURITUS	90 (3.5%)	74 (2.9%)	62 (2.4%)a	.070	.039
HEMATURIA	70 (2.7%)	50 (2.0%)	43 (1.7%)a	.026	.009
BRADYCARDIA	44 (1.7%)	20 (0.8%)b	29 (1.1%)	.009	.004
BREAST CARCINOMA	43 (1.7%)	16 (0.6%)c	16 (0.6%)c	<.001	<.001
EPISTAXIS	31 (1.2%)	16 (0.6%)a	22 (0.9%)	.086	.042
EMPHYSEMA	17 (0.7%)	17 (0.7%)	4 (0.2%)bd	.011	.139
VISUAL FIELD DEFECT	13 (0.5%)	5 (0.2%)	4 (0.2%)a	.037	.011
B12 DEFICIENCY ANEMIA	11 (0.5%)	5 (0.2%)	4 (0.2%)a	.067	.021
ALBUMINURIA	10 (0.4%)	4 (0.2%)	4 (0.2%)	.138	.046
COMA	9 (0.3%)	1 (0.0%)a	3 (0.1%)	.019	.006
THROMBOCYTHEMIA	6 (0.2%)	1 (0.0%)	1 (0.0%)	.093	.020
BREAST ENGORGEMENT	5 (0.2%)	2 (0.1%)	0 (0.0%)	.059	.046
INJECTION SITE PAIN	4 (0.2%)	1 (0.0%)	0 (0.0%)	.094	.046

INCONSISTENT EFFECT IN RALOXIFENE GROUPS

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
ACCIDENTAL INJURY	1129 (43.8%)	1141 (44.6%)	1055 (41.0%)ad	.023	.397
HERNIA	81 (3.1%)	70 (2.7%)	106 (4.1%) d	.019	.568
ANEMIA	79 (3.1%)	62 (2.4%)	98 (3.8%) d	.017	.900
CHILLS	32 (1.2%)	17 (0.7%)a	36 (1.4%) d	.030	.408
OTITIS MEDIA	28 (1.1%)	33 (1.3%)	15 (0.6%)ad	.031	.527
SEPSIS	5 (0.2%)	4 (0.2%)	13 (0.5%) d	.036	.286
ADENOMA	9 (0.3%)	0 (0.0%)b	11 (0.4%) d	.006	.272
SHOCK	4 (0.2%)	2 (0.1%)	10 (0.4%) d	.039	.474
PERICARDIAL EFFUSION	1 (0.0%)	5 (0.2%)	0 (0.0%) d	.020	.671

NOTE: Chi-square tests were used when total count >= 10, else Fisher's exact test was used.
a - pairwise comparison statistically significant (p < 0.05) different from placebo
b - pairwise comparison statistically significant (p < 0.01) different from placebo
c - pairwise comparison statistically significant (p < 0.001) different from placebo
d - pairwise comparison of RLX060 statistically significant (p < 0.05) different from RLX120

Data: RMP.SAS.H3SM.MCGGGKSC.FINAL
Source: RMP.H3SSK4YR.SASPGM(ARTESS@1) 95761 04JAN01
Output: RMP.H3SO.GGGK.FINAL(ARTESS@1)

12.3. Deaths, Other Serious Adverse Events, and Clinically Significant Adverse Events

The analyses of SAEs presented in this report are based on the reporting database and include randomized patients who reported an adverse event that met any of the serious criteria *whether the event was judged to be related to study drug or not.*

- The reporting database contains only SAE information collected while the patient was participating in the study.
- Deaths and serious adverse events were also collected in a Pharmacovigilance database (ClinTrace). The listing of SAEs from the ClinTrace database may differ from the reporting database, as ClinTrace also contains SAEs that occurred after the patient discontinued from the trial and that were judged by the investigator to be potentially related to study drug.

Appendix 16.3 of this CSR contains the case report forms for any patient who died, experienced other serious adverse events, or discontinued due to an adverse event.

12.3.1. Deaths

As of the 48-month visit, 100 deaths were reported among the randomly assigned patients in the clinical trial database (36 [1.4%] in the placebo group, 23 [0.9%] in the raloxifene 60-mg group, and 41 [1.6%] in the raloxifene 120-mg group) (Table GGGK.12.5).

- There were no statistically significant treatment differences in the number of patients who died overall or between the pooled raloxifene groups and the placebo group.
- Neither raloxifene group was statistically different from the placebo group; however, there was a statistically significant difference between the raloxifene 60-mg group and the raloxifene 120-mg group, with the lower incidence in the raloxifene 60-mg group.
- When **analyzed by event term**, there was no overall statistically significant difference among the three treatment groups in the proportion of deaths due to **any single event.**
- When **analyzed by body system**, there was an overall statistically significant difference among the three treatment groups in the proportion of deaths due to respiratory system events (p=0.037). For events in the respiratory system, there were fewer deaths in the raloxifene 60-mg group (no deaths) compared with both the placebo group (7 deaths representing 0.3%) and the raloxifene 120-mg group (6 deaths, representing 0.2%).

Table GGGK.12.5. Adverse Events Reported for Patients Who Died (All Randomly Assigned Patients, 48-Month Data)

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
OVERALL	36 (1.4%)	23 (0.9%)	41 (1.6%) d	.077	.584
BODY AS A WHOLE	11 (0.4%)	12 (0.5%)	12 (0.5%)	.969	.801
CARCINOMA	4 (0.2%)	6 (0.2%)	5 (0.2%)	.927	.819
ACCIDENTAL INJURY	2 (0.1%)	3 (0.1%)	2 (0.1%)	.807	1.000
DEATH	1 (0.0%)	2 (0.1%)	2 (0.1%)	.753	.670
SEPSIS	1 (0.0%)	1 (0.0%)	1 (0.0%)		
SUICIDE ATTEMPT	2 (0.1%)	0 (0.0%)	0 (0.0%)		
HYDROCEPHALUS	1 (0.0%)	0 (0.0%)	0 (0.0%)		
OVERDOSE	0 (0.0%)	1 (0.0%)	0 (0.0%)		
PERITONITIS	0 (0.0%)	0 (0.0%)	1 (0.0%)		
SCLERODERMA	0 (0.0%)	0 (0.0%)	1 (0.0%)		
SUDDEN DEATH	0 (0.0%)	1 (0.0%)	0 (0.0%)		
CARDIOVASCULAR SYSTEM	12 (0.5%)	5 (0.2%)	11 (0.4%)	.220	.290
MYOCARDIAL INFARCT	3 (0.1%)	2 (0.1%)	4 (0.2%)	.850	1.000
HEART ARREST	3 (0.1%)	0 (0.0%)	2 (0.1%)	.383	.342
CEREBROVASCULAR ACCIDENT	2 (0.1%)	1 (0.0%)	1 (0.0%)		
HEART FAILURE	1 (0.0%)	2 (0.1%)	1 (0.0%)		
CORONARY THROMBOSIS	0 (0.0%)	0 (0.0%)	2 (0.1%)		
ARRHYTHMIA	0 (0.0%)	0 (0.0%)	1 (0.0%)		
COR PULMONALE	1 (0.0%)	0 (0.0%)	0 (0.0%)		
MESENTERIC ARTERIAL OCCLUSION	1 (0.0%)	0 (0.0%)	0 (0.0%)		
SHOCK	1 (0.0%)	0 (0.0%)	0 (0.0%)		
DIGESTIVE SYSTEM	3 (0.1%)	4 (0.2%)	6 (0.2%)	.584	.426
GASTROINTESTINAL CARCINOMA	3 (0.1%)	1 (0.0%)	5 (0.2%)	.293	1.000
HEPATOMA	0 (0.0%)	1 (0.0%)	1 (0.0%)		
GASTROINTESTINAL HEMORRHAGE	0 (0.0%)	1 (0.0%)	0 (0.0%)		
INTESTINAL GANGRENE	0 (0.0%)	1 (0.0%)	0 (0.0%)		
NERVOUS SYSTEM	3 (0.1%)	2 (0.1%)	5 (0.2%)	.499	.818
CEREBRAL HEMORRHAGE	1 (0.0%)	0 (0.0%)	4 (0.2%)	.115	.670
SUBARACHNOID HEMORRHAGE	1 (0.1%)	1 (0.0%)	0 (0.0%)		
ENCEPHALITIS	0 (0.0%)	1 (0.0%)	0 (0.0%)		
PARALYSIS	0 (0.0%)	0 (0.0%)	1 (0.0%)		
RESPIRATORY SYSTEM	7 (0.3%)	0 (0.0%)b	6 (0.2%) d	.037	.116
RESPIRATORY DISORDER	3 (0.1%)	0 (0.0%)	2 (0.1%)	.383	.342
CARCINOMA OF LUNG	1 (0.1%)	0 (0.0%)	1 (0.0%)		
PNEUMONIA	2 (0.1%)	0 (0.0%)	0 (0.0%)		
APNEA	0 (0.0%)	0 (0.0%)	1 (0.0%)		
HYPOVENTILATION	0 (0.0%)	0 (0.0%)	1 (0.0%)		
LUNG DISORDER	0 (0.0%)	0 (0.0%)	1 (0.0%)		
SKIN AND APPENDAGES	0 (0.0%)	0 (0.0%)	1 (0.0%)		
SKIN MELANOMA	0 (0.0%)	0 (0.0%)	1 (0.0%)		

NOTE: Chi-square tests were used when total count >= 10, else Fisher's exact test was used.
a - pairwise comparison statistically significant (p < 0.05) different from placebo
b - pairwise comparison statistically significant (p < 0.01) different from placebo
c - pairwise comparison statistically significant (p < 0.001) different from placebo
d - pairwise comparison of RLX060 statistically significant (p < 0.05) different from RLX120

Source: EMP.H3S8K4YR.SASPGM(DCDTH@@5) 86167 15NOV00
Data: EMP.SAS.H3SM.MCGGGKSC.FINAL
Output: EMP.H3SO.GGGK.FINAL(DCDTHSWN)

12.3.2. Other Serious Adverse Events

Serious adverse events are collected both in the **Lilly safety database** and **the clinical trial reporting database**. The Lilly safety database is used to facilitate the monitoring and filing of serious adverse event reports occurring in Lilly-sponsored clinical trials worldwide, as well as spontaneous reports of adverse events with Lilly-marketed products. The clinical trial reporting database is an archived database used for analysis purposes and it contains adverse event data collected via remote data entry systems.

The 48-month analyses of serious adverse events presented here are based on data from the clinical trial reporting database and include only randomly assigned patients who reported an adverse event that met one or more of the serious criteria.

Serious adverse events were analyzed by body system (Table GGGK.12.6) and also by COSTART term (Table GGGK.12.7). All analyses were performed on the proportion of patients reporting at least one serious adverse event after randomization for the body system or COSTART term.

Table GGGK.12.6. Summary of Serious Adverse Events Reported by Body System (All Randomly Assigned Patients, 48-Month Data)

Body System	Placebo (N= 2574)	RLX060 (N= 2557)	RLX120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
OVERALL	794 (30.8%)	732 (28.6%)	744 (28.9%)	.173	.063
BODY AS A WHOLE	574 (22.4%)	542 (21.2%)	536 (20.8%)	.342	.152
CARDIOVASCULAR SYSTEM	191 (7.4%)	185 (7.2%)	206 (8.0%)	.547	.744
DIGESTIVE SYSTEM	141 (5.5%)	120 (4.7%)	141 (5.5%)	.345	.474
ENDOCRINE SYSTEM	7 (0.3%)	5 (0.2%)	5 (0.2%)	.795	.498
HEMIC AND LYMPHATIC SYSTEM	26 (1.0%)	17 (0.7%)	27 (1.0%)	.280	.509
METABOLIC AND NUTRITIONAL DISORDERS	19 (0.7%)	32 (1.3%)	29 (1.1%)	.174	.076
MUSCULOSKELETAL SYSTEM	69 (2.7%)	55 (2.2%)	75 (2.9%)	.210	.707
NERVOUS SYSTEM	74 (2.9%)	74 (2.9%)	76 (3.0%)	.984	.898
RESPIRATORY SYSTEM	98 (3.8%)	75 (2.9%)	94 (3.7%)	.190	.249
SKIN AND APPENDAGES	77 (3.0%)	78 (3.1%)	84 (3.3%)	.835	.686
SPECIAL SENSES	33 (1.3%)	31 (1.2%)	31 (1.2%)	.964	.786
UROGENITAL SYSTEM	127 (4.9%)	84 (3.3%)b	78 (3.0%)c	<.001	<.001

NOTE: Chi-square tests were used when total count >= 10, else Fisher's exact test was used.
a - pairwise comparison statistically significant (p < 0.05) different from placebo
b - pairwise comparison statistically significant (p < 0.01) different from placebo
c - pairwise comparison statistically significant (p < 0.001) different from placebo
d - pairwise comparison of RLX060 statistically significant (p < 0.05) different from RLX120

Data: EMP.SAS.H3SM.MCGGGKSC.FINAL
Source: EMP.H3SSK4YR.SASPGM(ARSB@@@1) 60167 04JAN01
Output: EMP.H3SO.GGGK.FINAL(ARSB@@NN)

Table GGGK.12.6 summarizes the proportion of patients reporting at least one serious adverse event overall, listed by body system. A total of 2,270 (29.5%) of the 7,705 patients reported at least one serious adverse event after randomization (794 [30.8%] in the placebo group, 732 [28.6%] in the raloxifene 60-mg group, and 744 [28.9%] in the raloxifene 120-mg group).

Table GGGK.12.7. Summary of Statistically Significant Serious Adverse Events Reported by COSTART Term (All Randomly Assigned Patients, 48-Month Data)

Event Classification	Placebo (N= 2576)	Rlx060 (N= 2557)	Rlx120 (N= 2572)	Overall p-Value	Pooled Rlx p-Value
BREAST CARCINOMA	43 (1.7%)	16 (0.6%)c	16 (0.6%)c	<.001	<.001
INFECTION	26 (1.0%)	12 (0.5%)a	16 (0.6%)	.084	.032
DEEP THROMBOPHLEBITIS	8 (0.3%)	20 (0.8%)a	23 (0.9%)b	.023	.007
GASTROINTESTINAL DISORDER	10 (0.4%)	22 (0.9%)a	10 (0.4%) d	.030	.185
DIZZINESS	15 (0.6%)	5 (0.2%)a	8 (0.3%)	.061	.024
ANEMIA	10 (0.4%)	2 (0.1%)a	12 (0.5%) d	.031	.392
RESPIRATORY DISORDER	9 (0.3%)	1 (0.0%)a	7 (0.3%) d	.068	.084
ARRHYTHMIA	1 (0.0%)	7 (0.3%)a	5 (0.2%)	.113	.049
BREAST NEOPLASM	8 (0.3%)	3 (0.1%)	1 (0.0%)a	.039	.015
EDEMA	0 (0.0%)	7 (0.3%)b	4 (0.2%)a	.034	.019
LUNG NEOPLASM	7 (0.3%)	1 (0.0%)a	3 (0.1%)	.080	.034
COMA	7 (0.3%)	1 (0.0%)a	2 (0.1%)	.046	.014
SUBARACHNOID HEMORRHAGE	3 (0.1%)	6 (0.2%)	0 (0.0%) d	.033	1.000
ANOREXIA	5 (0.2%)	1 (0.0%)	0 (0.0%)	.045	.018

NOTE: Chi-square tests were used when total count >= 10, else Fisher's exact test was used.
a - pairwise comparison statistically significant (p < 0.05) different from placebo
b - pairwise comparison statistically significant (p < 0.01) different from placebo
c - pairwise comparison statistically significant (p < 0.001) different from placebo
d - pairwise comparison of RLX060 statistically significant (p < 0.05) different from RLX120

Data: rmp.SAS.H3SM.MCGGGHSC.FINAL
Source: RMP.H3SSK4YR.SASPGM(ARSAR021) 95761 04JAN01
Output: RMP.H3SO.GGGK.FINAL(ARSAR021)

12.3.3. Clinically Significant Adverse Events

12.3.3.1. Greater Incidence in Raloxifene Patients and Potentially Clinically Relevant

12.3.3.1.1. Phlebitis and Venous Thromboembolic Events (VTE)

There was an increased incidence of **phlebitis** overall and in the pooled raloxifene group compared with the placebo group, with the highest incidence reported in the raloxifene 120-mg group (29 patients, 1.1%). In the raloxifene 60-mg group, 21 patients reported phlebitis (0.8%), and in the placebo group 13 (0.5%). A statistically significant increase was observed in the proportion of patients reporting **new not worsening phlebitis** overall and in the pooled raloxifene group.

- **Phlebitis superficial** was the most frequently reported term for this treatment-emergent adverse event.
- Only 1 patient in the raloxifene 60-mg group reported phlebitis as severe.

Of the 10 patients with a **preexisting condition of phlebitis**, 1 patient in the raloxifene 60-mg group reported phlebitis as a treatment-emergent adverse event (worsened in severity).

Two patients in the raloxifene 120-mg group reported **both phlebitis and deep thrombophlebitis** as treatment-emergent adverse events. In these patients, phlebitis was reported 21 and 47 days prior to the report of deep thrombophlebitis.

At the end of the study, **the incidence of VTE**, which was *defined as deep thrombophlebitis (or deep vein thrombosis [DVT]), retinal vein thromboses (RVT), and pulmonary embolism (PE),* was increased in the pooled raloxifene group compared with the placebo group.

- There was an increased incidence of deep thrombophlebitis overall and in the pooled raloxifene group compared with the placebo group, as previously reported in the 36-month clinical study report.
- Similarly, higher incidences of deep thrombophlebitis occurred in both the raloxifene 60-mg group (20 patients, 0.8%) and the raloxifene 120-mg group (23 patients, 0.9%) compared with the placebo group (8 patients, 0.3%). Section 12.3.3.6.1 presents a detailed discussion of VTE.

12.3.3.1.2. Vasodilation (Hot Flushes)

There was an increased incidence of vasodilatation, or hot flushes, overall and in the pooled raloxifene group compared with the placebo group.

A significantly higher incidence of **vasodilatation** was reported in both the raloxifene 60-mg group (272 patients, 10.6%) and the raloxifene 120-mg group (319 patients, 12.4%) compared with the placebo group (183 patients, 7.1%).

A significantly higher incidence in **severe vasodilatation** was reported by the raloxifene 60-mg group (18 patients, 0.7%) compared with both the raloxifene 120-mg group (7 patients, 0.3%) and the placebo group (3 patients 0.1%).

An **overall difference** was observed in the proportion of patients reporting vasodilatation during the first 6-month visit interval only. During that period, a significantly higher incidence in treatment-emergent adverse events was reported by both the raloxifene 60-mg group (179 patients, 7.0%) and the raloxifene 120-mg group (198 patients, 7.7%) compared with the placebo group (94 patients, 3.6%) (Table GGGK.14.15).

Of the 415 patients with a **preexisting condition of vasodilatation**, 94 patients (24 patients taking placebo, 34 patients taking raloxifene 60 mg, and 36 patients taking raloxifene 120 mg) reported vasodilatation as a treatment-emergent adverse event (worsened in severity). An overall statistically significant difference among the three treatment groups was noted in the reporting of vasodilatation for the subgroup of patients without, but not the subgroup with, preexisting vasodilatation ($p<0.001$).

12.3.3.1.3. Leg Cramps

An increase was observed in the proportion of patients reporting **new leg cramps** between the pooled raloxifene group compared with the placebo group for each 6-month visit interval through the 18-month visit.

- An overall increase was observed during the first 6 months and during the 12- to 18-month interval.
- There was no increase in the risk of venous thromboembolism (VTE) among patients reporting leg cramps.

Of the 293 patients with a **preexisting condition of leg cramps**, 77 patients reported leg cramps as a treatment-emergent adverse event (worsened in severity).

- An overall statistically significant difference among the three treatment groups was noted in the reporting of leg cramps for the subgroup of patients without, but not the subgroup with, preexisting leg cramps ($p<0.001$).

There was an increased incidence of leg cramps overall and in the pooled raloxifene group compared with the placebo group.

- Both the raloxifene 60-mg group.(234, 9.2%) and the raloxifene 120-mg group (218, 8.5%) reported an increased frequency of leg cramps compared with the placebo group (154 patients, 6.0%).
- The majority of patients reported leg cramps as mild to moderate in severity with 26 randomly assigned patients (7 patients taking placebo, 12 patients taking raloxifene 60 mg, and 7 patients taking raloxifene 120 mg) reporting severe leg cramps.
- No statistically significant difference was noted among the three treatment groups in the reporting of severe leg cramps.

12.3.3.1.4. Peripheral Edema

There was an increased incidence of peripheral edema overall and in the pooled raloxifene group compared with the placebo group, with the highest incidence reported in the raloxifene 120-mg group (203 patients, 7.9%). In the raloxifene 60-mg group,
182 patients reported peripheral edema (7.1%), and the placebo group had 158 patients (6.1%) reporting this event.

- Peripheral edema was most frequently reported as ankle edema and usually referred to as edema in the lower extremities.
- The majority of patients reported peripheral edema as **mild to moderate** in severity, with only 15 (0.2%) of the randomly assigned patients reporting **severe** peripheral edema.
- No statistically significant differences were noted among the three treatment groups in the reporting of severe peripheral edema.
- There were no statistically significant differences observed among the three treatment groups in the reporting of peripheral edema when reported by 6-month visit intervals.

Of the 237 patients with a **preexisting condition of peripheral edema**, 40 patients (11 taking placebo, 17 taking raloxifene 60 mg, and 12 taking raloxifene 120 mg) reported peripheral edema as a treatment-emergent adverse event (worsened in severity).

12.3.3.1.5. Diabetes Mellitus

There was an increased incidence of diabetes mellitus overall and in the pooled raloxifene group compared with the placebo group. Similarly, higher incidences of diabetes mellitus occurred in both the raloxifene 60-mg group (38 patients, 1.5%) and the raloxifene 120-mg group (36 patients 1.4%), compared with the placebo group (17 patients, 0.7%).

Diabetes mellitus was reported as severe in 9 patients, with no difference among the three treatment groups.

The increased reporting of diabetes mellitus as a treatment-emergent adverse event among raloxifene-treated patients is considered to be due, in part, to **the baseline imbalance among treatment groups in the prevalence of diabetes mellitus**, as indicated by the baseline imbalances in fasting glucose and use of hypoglycemic agents.

- ***At baseline,** the mean value for **fasting glucose** was 5.22, 5.29, and 5.24 mmol/L for the placebo, raloxifene 60- and raloxifene 120-mg groups respectively; the differences between raloxifene 60-mg and placebo and between raloxifene 60 and 120 mg were both significant.*
- *Also, **at baseline**, 32, 55, and 39 patients in the placebo, raloxifene 60- and 120-mg groups reported **use of hypoglycemic agents**; the difference between raloxifene 60 mg and placebo was significant.*

Of the 195 patients with a **baseline condition of diabetes mellitus** (60 placebo patients, 77 raloxifene 60-mg patients, and 58 raloxifene 120-mg patients), 24 patients reported diabetes mellitus as a treatment-emergent adverse event (worsened in severity) (Table GGGK.14.13).

- Both the raloxifene 60-mg group (12 patients, 15.6%) and the raloxifene 120-mg group (10 patients, 17.2%) had an increased incidence of diabetes mellitus (worsened in severity) compared with the placebo group (2 patients, 3.3%) in the subgroup with preexisting diabetes.
- In contrast, there was no significant increase in reporting of diabetes in those without diabetes at baseline (Table GGGK.14.14).

12.3.3.1.6. Hyperglycemia

There was an increased incidence of hyperglycemia overall, with the highest frequency reported in the raloxifene 120-mg (0.8%) group.

(Of patients reporting hyperglycemia, only 6 patients (1 patient taking placebo, 3 patients taking raloxifene 60 mg, and 2 patients taking raloxifene 120 mg) **also reported diabetes.** This event became significantly different at the 48-month time point.)

Review of glucose levels, reported at the time the treatment-emergent adverse event of hyperglycemia was noted, does not confirm the diagnosis of hyperglycemia in most of these cases.

- Of 7 patients in the placebo group, 3 patients had glucose levels <7.0 mmol/L (126 mg/dL), 2 patients had no concurrent glucose level reported, and 2 patients had glucose levels >7.0 mmol/L (126 mg/dL).
- Of 9 patients in the raloxifene 60-mg group, 2 patients had glucose levels <7.0 mmol/L (126 mg/dL), 4 patients had no concurrent glucose level reported, and 3 patients had glucose levels >7.0 mmol/L (126 mg/dL).
- Of 20 patients in the raloxifene 120-mg group, 3 patients had glucose levels <7.0 mmol/L (126 mg/dL), 10 patients had no concurrent glucose level reported, and 7 patients had glucose levels >7.0 mmol/L (126 mg/dL).

12.3.3.1.7. Female Lactation

There was an increased incidence of female lactation overall and in the pooled raloxifene group compared with the placebo group. Both the raloxifene 60-mg group (4, 0.2%) and the raloxifene 120-mg group (9, 0.3%) reported a higher incidence compared with the placebo group (no patients). None of these patients was subsequently diagnosed with breast cancer.

Female lactation is the COSTART term used for various types of breast discharge. The characteristics of the breast discharge among the patients reporting the COSTART term female lactation varied, without any predominant type identified. The variability in the type of breast discharge underscores the heterogeneity of actual terms that are assigned to the COSTART term "female lactation."

Section 12.3.3.3 presents further data regarding breast tissue-related events.

12.3.3.1.8. Hematemesis

There was an increased incidence of hematemesis overall and in the pooled raloxifene group compared with the placebo group, with the highest incidence reported in the raloxifene 120-mg group (7 patients, 0.3%). In the raloxifene 60-mg group, 3 patients reported hematemesis (0.1%), while the placebo group had no patients reporting this event.

- This event became significantly different at the 48-month time point.

12.3.3.2. Lower Incidence in Raloxifene Patients and Potentially Clinically Relevant

12.3.3.2.1. Breast Carcinoma

There was a lower incidence of breast carcinoma and severe breast carcinoma overall and in the pooled raloxifene group compared with the placebo group. The number of cases of breast carcinoma reported was lower than the number of cases adjudicated and included for the efficacy analyses (based on intent-to-treat principles). Section 11.4.3 presents efficacy analyses for breast cancer.

12.3.3.2.2. Breast Neoplasm

There was a lower incidence of **breast neoplasm** overall and in the pooled raloxifene group compared with the placebo group and a lower incidence of **severe breast neoplasm** observed in the pooled raloxifene group compared with the placebo group.

- Similarly low incidence of breast neoplasm occurred in the raloxifene 60-mg group (81 patients, 3.2%) and the raloxifene 120-mg group (77 patients, 3.0%) compared with the placebo group (110 patients, 4.3%).
- Forty-three patients reported **both a treatment-emergent adverse event of breast neoplasm and breast carcinoma** (26, 1.06%) in the placebo group, 8 (0.3%) patients in the raloxifene 60-mg group, and 9 (0.3%) patients in the raloxifene 120-mg group ($p<0.001$ for overall and pooled comparisons).
- When patients who reported breast carcinoma were excluded from the treatment-emergent adverse event of breast neoplasm, there was no difference between treatment groups in non-cancerous breast neoplasms.

Of the 163 patients with a **preexisting condition of breast neoplasm**, 6 patients reported breast neoplasm as a treatment-emergent adverse event (worsened in severity).

- A statistically significant difference among the three treatment groups was noted in the reporting of breast neoplasm for the subgroup of patients without, but not the subgroup with, preexisting breast neoplasm ($p=0.021$).

12.3.3.2.3. Breast Engorgement

There was a lower incidence of breast engorgement reported in the pooled raloxifene group compared with the placebo group. The lowest incidence occurred in the raloxifene 120-mg group

(0 patients, 0.0%). In the raloxifene 60-mg group, 2 patients reported breast engorgement (0.1%), and the placebo group had 5 patients (0.2%) reporting this event.

- One patient in the placebo group reported severe breast engorgement and was diagnosed with breast cancer.
- One patient reported breast engorgement and also reported breast pain.
- No patients who reported breast engorgement also reported lactation.
- One patient reported a preexisting condition of breast engorgement that did not worsen in severity.

When patients who reported breast pain were excluded from the treatment-emergent adverse event of breast engorgement, there was no difference among the treatment groups.

12.3.3.2.4. Hypercholesterolemia

There was a lower incidence of **hypercholesterolemia** overall and in the pooled raloxifene group compared with the placebo group, and a lower incidence of **severe hypercholesterolemia** observed in the pooled raloxifene group compared with the placebo group.

- Both the raloxifene 60-mg group (82 patients, 3.2%) and the raloxifene 120-mg group (76 patients, 3.0%) reported a similarly low incidence of the treatment-emergent adverse event of hypercholesterolemia compared with the placebo group (147 patients, 5.7%).

Of the 1,137 patients with a **preexisting condition of hypercholesterolemia**, 47 patients (26 patients taking placebo, 14 patients taking raloxifene 60 mg, and 7 patients taking raloxifene 120 mg) reported hypercholesterolemia as a treatment-emergent adverse event (worsened in severity).

- A statistically significant difference among the three treatment groups was noted in the reporting of hypercholesterolemia for both subgroups of patients with ($p=0.009$) and without ($p<0.001$) preexisting hypercholesterolemia.

12.3.3.2.5. Hypertension

There was a lower incidence of hypertension overall and in the pooled raloxifene group compared with the placebo group, with the lowest incidence reported in the raloxifene 60-mg group (231 patients, 9.0%). In the raloxifene 120-mg group, 264 patients reported hypertension (10.3%), and the placebo group had 289 patients (11.2%) reporting this event.

- A statistically significant difference was observed in the proportion of patients reporting new or worsening hypertension among the three treatment groups for the 18- to 24-month and the 30- to 36-month intervals only.

12.3.3.3. Breast Tissue

By 48 months, raloxifene use was associated with a reduction in all types of **breast cancer** (see efficacy analyses in Section 11.4.3). In addition, raloxifene was not associated with the adverse events of **breast pain, breast enlargement, or breast engorgement.**

At each visit, patients were questioned regarding the occurrence of adverse events.

Breast-related adverse events are a subset of these reported events. The only breast-related serious adverse events reported after baseline were breast carcinoma, breast neoplasm, and fibrocystic breast (Table GGGK.12.8).

Table GGGK.12.8. Breast-Related Serious Adverse Events (All Randomly-Assigned Patients, 48-Month Data)

Serious Adverse Event	Placebo (N=2576)	RLX060 (N=2557)	RLX120 (N=2572)	Total (N=7705)	Overall p-value[a]	Pooled RLX p-value[a]
Breast Carcinoma[b]	43(1.7%)	16(0.6%)[c]	16(0.6%)[c]	75 (1.0%)	0.001	<0.001
Breast Neoplasm	8(0.3%)	3(0.1%)	1 (0.0%)d	12 (0.2%)	0.039	0.015
Fibrocystic Breast	1(0.0%)	3(0.1%)	2 (0.1%)	6 (0.1%)	0.464	0.671

a Chi-square test for total count ≥10; Fisher's Exact test for total counts 5 through 9.
b The number of cases of breast carcinoma reported was lower than the number adjudicated and included in the efficacy analyses (see Section 11.4.3).
c Pairwise comparison statistically significantly ($p<0.01$) different from placebo.
d Pairwise comparison statistically significantly ($p<0.05$) different from placebo.
Abbreviations: N = number of randomly assigned patients; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; RLX = raloxifene.

A summary of all treatment-emergent adverse events related to the breast is presented in (Table GGGK.12.9).

Table GGGK.12.9. Summary of Breast-Related Treatment-Emergent Adverse Events (All Randomly Assigned Patients, 48-Month Data)

Treatment-Emergent Adverse Event	Placebo (N=2576)	RLX060 (N=2557)	RLX120 (N=2572)	Total (N=7705)	Overall p-value[a]	Pooled RLX p-value[a]
Any Breast-Related Adverse Event	255(9.9%)	217(8.5%)	215(8.4%)	687(8.9%)	0.099	0.032
Breast Pain	80(3.1%)	74(2.9%)	79(3.1%)	233(3.0%)	0.893	0.767
Breast Neoplasm	110(4.3%)	81(3.2%)[b]	77(3.0%)[b]	268(3.5%)	0.025	0.007
Fibrocystic Breast	51(2.0%)	46(1.8%)	43(1.7%)	140(1.8%)	0.708	0.448
Breast Enlargement	19(0.7%)	24(0.9%)	22(0.9%)	65(0.8%)	0.731	0.471
Breast Carcinoma[c]	43(1.7%)	16(0.6%)[d]	16(0.6%)[d]	75(1.0%)	<0.001	<0.001
Female Lactation	0(0.0%)	4(0.2%)[b]	9(0.3%)[d]	13(0.2%)	0.009	0.011
Breast Engorgement	5(0.2%)	2(0.1%)	0(0.0%)	7(0.1%)	0.059	0.046
Breast Atrophy	1(0.0%)	1(0.0%)	2(0.1%)	4(0.1%)	—	—
Mastitis	1(0.0%)	3(0.1%)	4(0.2%)	8(0.1%)	0.423	0.282

[a] Chi-square test for total count ≥10; Fisher's Exact test for total counts 5 through 9.

[b] Pairwise comparison statistically significantly ($p<0.05$) different from placebo.

[c] The number of cases of breast carcinoma reported was lower than the number adjudicated and included in the efficacy analyses (see Section 11.4.3).

[d] Pairwise comparison statistically significantly ($p<0.01$) different from placebo.

Abbreviations: N = number of randomly assigned patients; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; RLX = raloxifene.

12.3.3.4. Uterine Corpus

Only those patients who did not have a hysterectomy at baseline were included in the analyses in this section. Patients who reported vaginal bleeding of uterine origin and patients who had endometrial thickness measurements of >5.0 mm by transvaginal ultrasound (TVU) underwent a uterine evaluation. A patient could undergo evaluation, according to the appropriate uterine evaluation algorithms, multiple times during the study. A special database was developed to capture the data generated on any uterine assessment procedures. The form on which investigators entered uterine evaluation data was termed "uterine packet." The decision to enter the data obtained into a uterine packet for conditions other than vaginal bleeding or increased endometrial thickness was not made in a systematic fashion. Therefore, data collected under the category "Other" were not analyzed further, except for those patients who experienced uterine bleeding or endometrial thickness >5.0 mm and were analyzed in their respective groups.

12.3.3.4.1. Assessment of the Effects of Raloxifene on the Uterus

Investigators were not required to provide a final clinical diagnosis accounting for uterine bleeding or increased endometrial thickness. The assignment of one or more final clinical diagnoses was based on a review of all data entered into a uterine packet and any additional diagnostic assessment provided by the study site. A sponsor gynecologist, without knowledge of treatment assignment, performed this data review.

Diagnostic assignment was based on the findings from TVU, saline-infusion sonohysterography (SIS), hysteroscopy, and/or any form of endometrial sampling/biopsy.

The clinical diagnosis was recorded according to an adaptation of the International Classification of Diseases, 9th revision (ICD-9-CM). Table GGGK.14.17 in Section 14.3.5 lists the possible ICD-9-CM assignments. All clinical diagnostic assignments were done for the entire dataset at the 48-month time point, not just for additional packets since the 36-month analysis (data on file).

Any histologic sample obtained through a blind biopsy, hysteroscopy and guided biopsy, or dilatation and curettage (D&C) was assigned a pathological diagnosis according to the World Health Organization (WHO)/Blaustein criteria for endometrial pathology (Kurman 1994). For patients with more than one histologic diagnosis or multiple samplings (whether obtained locally or centrally), only the highest assigned code for the WHO/Blaustein classification was used in the analysis in order to represent the most advanced stage of endometrial stimulation (Table GGGK.14.18 in Section 14.3.5). For this analysis, Classification Codes 7 (Hyperplasia) and 8 (Benign Neoplasia) were considered comparable in severity.

In many instances, samples were analyzed locally and were also forwarded for central reading. However, not all local reports could be systematically collected. For this reason, the diagnoses from all local pathology reports that were collected and received by the sponsor were assigned a WHO/Blaustein diagnostic classification by a sponsor gynecologist blinded to treatment assignment.

Patients who had a histologic diagnosis of insufficient tissue, surface endometrium, or inactive/atrophic endometrium result were grouped under the category normal, as has been done in the Postmenopausal Estrogen/Progestin Interventions (PEPI) trial involving estrogen and hormone replacement therapy (PEPI Trial Writing Group 1996).

12.3.3.4.1.1. Results for Patients with Uterine Bleeding

The proportion of patients reporting postmenopausal bleeding was calculated for women who had not undergone a prior hysterectomy at baseline.

- Of the 5,959 patients who had not undergone a prior hysterectomy, 220 (3.7%) reported vaginal bleeding (74 [3.7%] in the placebo group, 81 [4.2%] in the raloxifene 60-mg group, and 65 [3.2%] in the raloxifene 120-mg group) (Table GGGK.12.10).
- Any patients who experienced both bleeding and any other uterine event (including endometrial thickness >5.0 mm) were analyzed in the bleeding category.
- For each category of non-uterine- or uterine- bleeding presented in Table GGGK.12.10, there were no statistically significant treatment-group differences in the number of patients reporting bleeding.

Table GGGK.12.10. Incidence of Uterine and Non-uterine Bleeding (All Randomly Assigned Patients without Prior Hysterectomy, 48-Month Data)

Bleeding Category	Placebo (N=1999)	RLX060 (N=1950)	RLX120 (N=2010)	Total (N=5959)	Overall p-value[a]	Pooled RLX p-value[a]
Any Vaginal Bleeding (AE or Uterine Surveillance) After Baseline	74 (3.7%)	81 (4.2%)	65 (3.2%)	220 (3.7%)	0.308	0.977
Nonuterine-Related Bleeding	5 (0.3%)	7 (0.4%)	8 (0.4%)	20 (0.3%)	0.704	0.417
Uterine-Related Bleeding	69 (3.5%)	74 (3.8%)	57 (2.8%)	200 (3.4%)	0.236	0.771
Algorithm Follow-Up	46 (2.3%)	55 (2.8%)	32 (1.6%)	133 (2.2%)	0.032	0.797
No Algorithm Follow-Up	23 (1.2%)	19 (1.0%)	25 (1.2%)	67 (1.1%)	0.717	0.891
Expected Bleeding	11 (0.6%)	9 (0.5%)	9 (0.5%)	29 (0.5%)	0.880	0.616
Nonalgorithm Follow-Up	4 (0.2%)	7 (0.4%)	3 (0.2%)	14 (0.2%)	0.365	0.693
No Follow-Up	8 (0.4%)	3 (0.2%)	13 (0.7%)	24 (0.4%)	0.050	0.982

[a] Chi-square test.

Abbreviations: AE = adverse event; N = number of randomly assigned patients without prior hysterectomy; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; HRT = hormone replacement therapy.

Note: Expected bleeding refers to bleeding after procedures or administration of hormone replacement therapy; nonalgorithm follow-up included any evaluation by a physician; and no follow-up refers to no documented follow-up of procedures or evaluation.

Among the patients presenting with uterine bleeding, the most common finding after entry into the uterine algorithm due to uterine bleeding was **endometrial atrophy**, a normal finding in postmenopausal women (Table GGGK.12.11). Atrophy was commonly observed **along with an**

identifiable pathologic finding (eg, endometrial atrophy and submucosal myoma). The occurrence of endometrial atrophy was comparable in all three treatment groups.

The most common abnormal finding was that of **benign endometrial polyp**, found more frequently in the raloxifene treatment groups than in the placebo group. There was a statistically significant difference between the pooled raloxifene groups compared with the placebo group (p=0.049). There was no statistically significant difference for any other clinical diagnosis.

Table GGGK.12.11. Clinical Diagnoses for Patients with Bleeding (International Classification of Diseases, All Randomly Assigned Patients without Prior Hysterectomy, 48-Month Data)

Clinical Diagnosis	Placebo (N=1999)	RLX060 (N=1950)	RLX120 (N=2010)	Total (N=5959)	Overall p-value[a]	Pooled RLX p-value[a]
Myoma	4	5	5	14	0.924	0.693
Polyp	6	17	11	34	0.057	0.049
Atrophy	35	49	33	117	0.100	0.401
Unspecified Disorder	0	0	2	2	---	---
Proliferative	7	1	4	12	0.111	0.069
Secretory[b]	0	0	1	1	---	---
Hyperplasia	1	3	0	4	---	---
Normal	3	0	2	5	0.339	0.342
SEETUS	5	1	2	8	0.269	0.128
Adenocarcinoma	2	4[c]	1	7	0.277	1.000
Cystic Focus (microcysts)	0	0	1	1	---	---
Adenofibroma	0	1	0	1	---	---
Hematometra	0	0	1	1	---	---
Mucocele	0	1	0	0	---	---
Endometrial Fluid	1	0	2	3	---	---
Sarcoma	1	0	1	2	---	---
Adenomyosis	2	1	1	4	---	---
Total Clinical Diagnoses[d]	47	59	44	150	0.206	0.561
Incomplete Data[e]	3	4	3	10	0.887	0.812
Algorithm Complete[f]	1	4	2	7	0.323	0.436
Algorithm Violation[g]	15	20	11	46	0.226	0.892
Lost to Follow-Up[h]	1	0	0	1	---	---

a Chi-square test for total count ≥10; Fisher's Exact test for total count 5 through 9.

b One case of secretory endometrium was identified in a patient in the raloxifene 120-mg group who received combined estrogen/progestin therapy.

c One case of adenocarcinoma was reported in the 36-month dataset, but was excluded from the same table at 36 months because she did not have a uterine packet completed by the 36-month visit. Refer to Section 14.3.5, Table GGGK.14.19 for further detail.

d Counts patients with multiple clinical diagnoses only once.

e Insufficient data to establish a clinical diagnosis.

f Patients who completed all mandated steps in the uterine surveillance algorithm, but showed no clinical evidence to pursue further evaluation.

g Algorithm was not followed. No clinical diagnosis made despite the follow-up reported.

h Patient was lost to follow-up.

Abbreviations: N = number of randomly assigned patients without prior hysterectomy; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; SEETUS = spurious elevation of endometrial thickness of undetermined significance; RLX = raloxifene.

- An overall statistically significant difference was observed among the three treatment groups, and between the pooled raloxifene group and the placebo group, for the histopathology based diagnosis of **benign endometrial polyps** (p=0.042 and p=0.022, respectively).
- There was no apparent dose effect in the proportion of patients who were identified as having **both uterine bleeding and polyps**.
- There was an overall statistically significant difference between the pooled raloxifene group and the placebo group for the **pathological diagnosis of proliferative tissues** (p=0.020) with most cases found in the placebo group.

Table GGGK.12.12. Combined Centrally and Locally Read Endometrial Biopsy Results (WHO/Blaustein Biopsy Classification, For Patients with Bleeding, All Randomly Assigned Patients without Prior Hysterectomy, 48-Month Data)

Biopsy Result	Placebo (N=1999)	RLX060 (N=1950)	RLX120 (N=201)	Total (N=5959)	Overall p-value[a]	Pooled RLX p-value[a]
Normal (PEPI Definition)[b]	27	35	22	84	0.168	0.784
Insufficient Tissue	16	14	12	42	0.741	0.531
Surface Endometrium Only	3	10	4	17	0.069	0.164
Inactive/Atrophic Endometrium	8	11	6	25	0.428	0.870
Nonproliferative	1	0	5	6	0.053	0.671
Proliferative Tissues	6	1	1	8	0.083	0.020
Polyps	3	13	9	25	0.042	0.022
Hyperplasia	0	2	0	2	---	---
Benign Neoplasia	1	1	1	3	---	---
Carcinoma	3	4[c]	2	9	0.592	1.000
Endometritis	0	0	0	0	---	---

a Chi-square test for total count ≥10; Fisher's Exact test for total count 5 through 9.
b Normal includes insufficient tissue, surface endometrium, and inactive/atrophic endometrium results (Langer et al. 1997).
c One case of adenocarcinoma was reported in the 36-month dataset, but was excluded from the same table at 36 months because she did not have a uterine packet completed by the 36-month visit. Refer to Table GGGK.14.19 in Section 14.3.5 for further detail.
Abbreviations: PEPI = Postmenopausal Estrogen/Progestin Interventions Trial; N = number of randomly assigned patients without prior hysterectomy; RLX = raloxifene; RLX060 = raloxifene 60 mg/day; RLX120 = raloxifene 120 mg/day; WHO = World Health Organization.

12.3.3.4.1.2. Results for Patients with Increased Endometrial Thickness

The majority of ultrasonograms were performed by TVU while few sites used transabdominal techniques. All images were interpreted locally and were recorded to the nearest 0.1 mm.

Among the subset of 2,157 patients who were scheduled at the onset of the study to have both a baseline and at least one post baseline endometrial thickness measurement, a total of 1,644 patients fulfilled this requirement. There was no statistically significant difference among the