Statistical Review

NDA 21-042
Name of Drug: Rofecoxib
Applicant: Merck Research Laboratories
Indications: Treatment of Signs and Symptoms of Osteoarthritis, relief of pain, treatment of primary dysmenorrhea, and improvement of gastrointestinal safety
Documents Reviewed: Statistical and clinical sections of the NDA
Reviewer: Qian Li, Sc.D.
Date of Review: December 1998-April 1999

Background and Summary

Rofecoxib is a drug developed by Merck Research Laboratories and submitted for NDA for the approval of acute and chronic treatment of the signs and symptoms of osteoarthritis (OA), relief of pain, and treatment of primary dysmenorrhea. A class of nonsteroidal anti-inflammatory (NSAIDS) drug are wildly available in market for the treatment of signs and symptoms of OA. However, NSAIDS is notorious for their association with serious GI adverse events because of inhibition of both cyclooxyenase-1 (COX-1) and cyclooxyenase-2 (COX-2) in the gastric mucosa. The inhibition of COX-1 has been believed to be responsible for the association of the GI events. Rofecoxib is claimed to be COX-2 specific inhibitor. Therefore it was expected to have a better GI safety profile. For this reason, GI safety profile of Rofecoxib in comparison with NASIDS and placebo as an additional indication was also submitted for review.

In this statistical review, three aspects are discussed with regard to the treatment of Rofecoxib in osteoarthritis (OA) patients. Section I summarizes and discusses the study results on therapeutic effect of Rofecoxib in treating OA patients. Four Phase III pivotal studies were reviewed, including two short term 6 week (Studies 33 and 40) and two long term 6 month (Studies 34 and 35) studies. Two doses of Rofecoxib, 12.5 and 25 mg daily, were tested in the four studies to compare their treatment effects with those of placebo and NSAIDS (Ibuprofen and Diclofenac). In the two short term studies, both Rofecoxib doses demonstrated statistically and clinically significant improvement over placebo, as well as therapeutic effect comparable to Ibuprofen after 6 weeks of treatment of the sign and symptom of OA. In the two long term studies, the two dose groups of Rofecoxib were compared with Diclofenac. Again both Rofecoxib doses demonstrated comparable therapeutic effect to Diclofenac after 12 weeks and 6 months of treatment, although the discontinuation rates due to lack of efficacy were relatively high in Rofecoxib two dose groups compared to Diclofenac treatment group. Section II summarizes the results from four GI study reports (44, 45, 44c, 69), comments and new analyses on those studies. The two endoscopy studies compared two doses of Rofecoxib, 25 and 50 mg daily, to placebo and Ibuprofen treatment groups. Despite of the problematic endoscopy study design, the result showed that the rate of endoscopy identified ulcers in Ibuprofen treatment group was significantly higher than that in Rofecoxib two dose groups. However, the reviewer disagrees with the sponsors'

conclusion that the ulcer rates were similar among Rofecoxib 25 and 50 mg and placebo treatment groups. Study 69 covered 8 Phase III/IIb trials to monitor the upper GI events. Since the 8 studies differed in many aspects, the simple combination of the 8 studies made it impossible to interpret the study results. Section III discusses issues arising in reviewing dose-response relationship of Rofecoxib. The doses of Rofecoxib studied ranged from 5 to 125 mg daily. Sharp dose-response relationship was observed in Study 29, a Phase II dose range study. However, an integrated analysis by combining Study 29 and a pilot study (Study 10) showed a small increase of treatment benefit with increased dose of Rofecoxib. The reviewer points out that the integrated analysis was a very subjective analysis for lack of scientific and biological bases and therefore should be of less value.

I. Efficacy review:

Four pivotal study results were presented to support efficacy claim. Two short term up to 6 week studies (Studies 33 and 40) and two long term up to 6 month studies (Studies 34 and 35). Common statistical features of the four studies were summarized first in this section and followed by study results and reviewer's comments.

Study population:

Intent To Treat (ITT) analysis population included all patients with a baseline and at least one post baseline measurement. The primary analysis was based on the ITT population.

The Per-Protocol (PP) analysis population excluded patients and/or data points with clinically important protocol deviations based on prespecified criteria before blinding.

End Points:

Measurement of effectiveness was based on three primary endpoints and several other effectiveness endpoints. The primary endpoints are:

1. Investigator's global assessment of disease activity,
2. Patient's global assessment of response to therapy, and
3. Pain walking on a flat surface.

The first two endpoints are measured on a 0 to 4 point Likert scale, and the third on a 0 to 100 mm visual analog scale (VAS). The secondary endpoints include:

4. WOMAC functional subscale (0 to 100 mm VAS),
5. Patient global assessment of disease activity (0 to 100 VAS),
6. Proportion of patients discontinuing study therapy due to lack of efficacy, and
7. WOMAC stiffness subscale (0 to 100 VAS).
8. WOMAC pain subscale (0 to 100 mm VAS)
9. WOMAC questionnaire total score,
10. WOMAC total subscale average score,
11. Investigator global assessment of response to therapy (0 to 4 point Likert scale),
12. Acetaminophen tablet count

13. Study Joint Tenderness (0-3 point Likert scale)
14. Proportion of patients with swelling in study joint

Statistical Analysis:

The primary analyses were average change from baseline for each of the three primary end points. Obviously, the missing data was equivalent to be imputed by the average of the response. The primary analyses were implemented by ANCOVA, adjusting study center, baseline and GI history status.

Last observation carry forward approach was examined to check the robustness of the study result. In general, this approach can be slightly more conservative than average change from baseline when comparing a treatment group with placebo. This is because placebo treatment group usually has high drop out rate due to lack of efficacy, and last observation can be smaller than the average change from baseline. However, it is less clear in the comparison with active control. This analysis can be conservative when active control is more effective, or vice verse if the active control is less effective.

Comparability and superiority criteria:

For the purpose of demonstrating comparability between a Rofecoxib and an NASIDS treatment groups for an individual endpoints, a two sided 95% confidence interval (CI) for the mean difference between the groups will be compared to predefined clinically important comparability bound. The bounds are ±0.5 on the 0 to 4 point Likert scale, and ±10 mm on the 0 to 100 mm visual analog scale (VAS). If the CI falls within the bound, the conclusion of comparability between the two treatments will be drawn.

For the purpose of comparing between placebo and Rofecoxib, statistically significant differences were the criteria. Ideally, the criteria should be set up in consistent with superiority comparison between Rofecoxib and active control in the study design.

Multiplicity issues:

Multiple controls: There was no discussion on the multiple comparisons in studies with both placebo and active controls. In order to control the overall alpha level at 0.05, the comparison should be divided into two families. The first family was the comparison between Rofecoxib and placebo. The second was the comparison between Rofecoxib and Ibuprofen. Unless Rofecoxib showed statistically significant improvement over placebo, there would be no further consideration on the comparison between Rofecoxib and active control. Therefore, closed testing procedure was applied to the two families, and the alpha level of each family was controlled at alpha=0.05.

Multiple dose groups: Within each family, there were two dose groups of Rofecoxib, 12.5 and 25 mg. It was implicitly mentioned in Data Analysis Plan (DAP) that Rofecoxib 25 mg was the primary goal in efficacy studies. Closed testing procedure was applied to multiple dose comparisons.

Multiple end points: Three primary end points were tested for each comparison between treatment groups. For superiority test, all the three end points had to show statistical significance. For comparability comparison, the 95% CI of all the three primary endpoints must fall within the comparability bound. Therefore, no alpha adjustment was necessary. These criteria were consistent with sponsor's prespecified criteria. In sponsor's DAP, the criteria were that the 95% CI for at least two of the three primary endpoints must fall entirely within the comparability bounds, and all three endpoints must have at least 95% posterior probability with an non-informative prior that the mean difference falls within the comparability bound. Since the posterior distribution given non-informative prior under normal assumption was the same as the sample distribution, 95% CI within the comparability region implied that 95% posterior probability of the posterior mean was within the bound.

Short term studies:

Studies 33 and 40 were identically designed, multicenter, double-blind, placebo- and active-controlled, parallel-group, 6-week studies to assess the safety and efficacy of Rofecoxib in comparison with placebo, Ibuprofen in patients with osteoarthritis of the knee or hip. There were 3 to 15 day NSAIDS washout period, followed by a 6-week treatment period with 12.5, 25 mg Rofecoxib once daily, 800 mg Ibuprofen 3 times daily, or placebo. The efficacy parameters were measured at baseline(flare/confirmation week 0), week 2, 4 and 6.

Patient accounting information at week 6 was summarized in Table I-1 for studies 33 and 40. Overall the discontinuation rates were comparable between treatment groups within each study. The baseline characteristics were comparable between treatment groups within each study as well.

Table I-1: Patient accounting information of short term studies.

		Placebo	Rofecoxib 12.5 mg	Rofecoxib 25 mg	Ibuprofen 2400 mg
Study 33	Entered	69	219	227	221
	Completed	50 (72.5%)	186 (84.9%)	200 (88.1%)	625 (84.9%)
	Discontinued	19 (27.5%)	33 (15.1%)	27 (11.9%)	32 (14.5%)
	Clinical AE	4 (5.8%)	12 (5.5%)	15 (6.6%)	8 (3.6%)
	Lab AE	0	0	0	1 (0.5%)
	Lack of Efficacy	13 (18.8%)	17 (7.8%)	9 (4.0%)	19 (8.6%)
Study 40	Entered	74	244	242	249
	Completed	62 (83.8%)	217(88.9%)	217 (89.7%)	213 (85.5%)
	Discontinued	12 (16.2%)	27 (11.1%)	25 (10.3%)	36 (14.5%)
	Clinical AE	1 (1.4%)	10 (4.1%)	9 (3.7%)	21 (8.4%)
	Lab AE	0	2 (0.8%)	0	0
	Lack of Efficacy	9 (12.2%)	8 (3.3%)	7 (2.9%)	9 (3.6%)

Comparison with placebo:

In both studies 33 and 40, Rofecoxib 12.5 and 25 mg, demonstrated statistically significantly greater improvement than placebo over 6 weeks of treatment as assessed by the three primary variables. The results of the two studies were summarized in Table I-2 and Table I-3. For all secondary end points, Rofecoxib two doses demonstrated statistically significantly greater improvement than placebo.

Table I-2: Results from Study 33.

End Point	Pairwise Treatment Difference in LSMean (95% CI)		
	12.5—Placebo	25 mg—Placebo	Ibuprofen—Placebo
Pain Walking on a Flat Surface (WOMAC) (0- to 100-mm VAS)	-12.40 (-18.56, -6.24)	-16.49 (-22.68, -10.31)	-13.45 (-19.64, -7.26)
Patient Global Assessment of Response to Therapy (0- to 4-point Likert†)	-0.87 (-1.14, -0.59)	-1.01 (-1.28, -0.73)	-0.87 (-1.15, -0.60)
Investigator Global Assessment of Disease Status (0- to 4-point Likert)	-0.58 (-0.80, -0.35)	-0.66 (-0.89, -0.44)	-0.53 (-0.75, -0.31)
† Scale reversed so decreasing values indicate improvement to be consistent with presentation of the other end points.			

Table I-3: Results from Study 40.

End Point	Difference in LSMean† (95% CI)		
	25 mg-Placebo	12.5 mg-Placebo	Ibuprofen-Placebo
Pain Walking on a Flat Surface (WOMAC) (VAS)	-16.15 (-21.48, -10.83)	-15.40 (-20.72, -10.08)	-14.63 (-19.92, -9.33)
Patient Global Assessment of Response to Therapy (Likert)	-0.88 (-1.11, -0.65)	-0.72 (-0.94, -0.49)	-0.66 (-0.88, -0.43)
Investigator Global Assessment of Disease Status (Likert)	-0.58 (-0.77, -0.39)	-0.46 (-0.65, -0.28)	-0.40 (-0.59, -0.21)
† Least square mean.			

Comparison with Ibuprofen:

Rofecoxib 12.5 and 25 mg demonstrated efficacy comparable to Ibuprofen over 6 weeks of treatment as assessed by the three primary variables (Refer to Table I-4 and I-5). The results from the secondary end point analyses were consistent with those from the primary end points. Although there were no predefined comparability criteria for these end points, the 95% CIs for the pairwise treatment differences between the 3 active-treatment groups for these end points were within the range of comparability criteria defined for the primary end points.

LOCF analysis supported the conclusion based on the average change from baseline.

Table I-4: Results of comparison between Rofecoxib and Ibuprofen from Study 33

End Point	Pairwise Treatment Difference in LSMean (95% CI)		
	25 mg—Ibuprofen	12.5 mg—Ibuprofen	25 mg—12.5 mg
Pain Walking on a Flat Surface (WOMAC) (0- to 100-mm VAS)	-3.05 (-7.28, 1.18)	1.04 (-3.23, 5.31)	-4.09 (-8.36, 0.18)
Patient Global Assessment of Response to Therapy (0- to 4-point Likert†)	-0.13 (-0.32, 0.05)	0.01 (-0.18, 0.20)	-0.14 (-0.33, 0.05)
Investigator Global Assessment of Disease Status (0- to 4-point Likert)	-0.13 (-0.28, 0.02)	-0.05 (-0.20, 0.11)	-0.09 (-0.24, 0.06)

† Scale reversed so decreasing values indicate improvement to be consistent with presentation of the other end points.

Table I-5: Results of comparison between Rofecoxib and Ibuprofen from Study 40:

End Point	Difference in LSMean† (95% CI)		
	25 mg-Ibuprofen	12.5 mg-Ibuprofen	25 mg-12.5 mg
Pain Walking on a Flat Surface (WOMAC) (VAS)	-1.52 (-5.14, 2.09)	-0.77 (-4.37, 2.83)	-0.76 (-4.39, 2.88)
Patient Global Assessment of Response to Therapy (Likert)	-0.22 (-0.38, -0.07)	-0.06 (-0.21, 0.10)	-0.17 (-0.32, -0.01)
Investigator Global Assessment of Disease Status (Likert)	-0.18 (-0.31, -0.05)	-0.06 (-0.19, 0.06)	-0.12 (-0.25, 0.01)

† Least square mean.

Long term studies:

Studies 34 and 35 were multicenter double-blind, active-controlled, parallel-group, 6 month studies to assess the safety and efficacy of Rofecoxib versus Diclofenac in patients with osteoarthritis of the knee or hip. There were 3 to 15 NSAIDS washout period, followed by a 6-month treatment period with 12.5, 25 mg Rofecoxib once daily, and 50 mg Diclofenac 3 times daily. The objective of the trial was to demonstrate clinical efficacy of Rofecoxib 12.5 and 25 mg comparable to Diclofenac sodium in the treatment of osteoarthritis of the knee or hip primarily during a 12-week treatment period, secondarily to 6 months. Although the studies lasted one year, there was limited efficacy information in the second half 6 months. The efficacy parameters were evaluated at week 0 (flare/confirmation visit), 2, 4, 8, 12 and 26. Notice that since the primary analyses were average change from baseline, and there were more frequent efficacy measurement in early stage of study, it can been seen that early treatment difference could play an important role in such analysis.

In general the baseline characteristic were comparable between treatment groups within each study. Patient accounting information of the two studies at month 6 was summarized in Table I-6. As it can be seen in both studies, there were more patients discontinued due to lack of efficacy in Rofecoxib two treatment groups than Diclofenac treatment group. In study 35, the discontinuation rate was significantly higher in Rofecoxib 25 mg dose

group compared to Diclofenac treatment group. However, there were relatively higher discontinuation rate due to clinical and laboratory adverse reaction in Diclofenac treatment group compared to the two Rofecoxib treatment groups. The impact of the discontinuation due to lack of efficacy was further discussed later.

Table I-6: Patient accounting information at Month 6.

		Rofecoxib 12.5 mg	Rofecoxib 25 mg	Diclofenac 150 mg
Study 34	Entered	231	232	230
	Completed	169 (73.1%)	183 (78.0%)	174 (75.6%)
	Discontinued	62 (26.8%)	49 (21.1%)	56 (24.3%)
	Clinical AE	18 (7.7%)	14 (6.0%)	22 (9.6%)
	Lab AE	0	1 (0.4%)	10 (4.3%)
	Lack of Efficacy	22 (9.5%)	19 (8.2%)	13 (5.6%)
Study 35	Entered	259	257	268
	Completed	183 (70.7%)	166 (64.4%)	173 (64.6%)
	Discontinued	76 (29.3%)	91 (35.4%)	95 (35.4%)
	Clinical AE	29 (11.2%)	25 (9.7%)	35 (13.1%)
	Lab AE	1 (0.4%)	2 (0.8%)	14 (5.2%)
	Lack of Efficacy	34 (13.1%)	50 (19.5%)	27 (10.1%)

Results after 12 weeks of treatment:

The results of the least square mean change over 12 weeks of treatment from studies 34 and 35 were summarized in Table I-7 and Table I-8 respectively for the three primary end points. Most of the pairwise treatment differences in LSMean changes were statistically significant ($p<0.05$ in favor of Diclofenac, especially for Rofecoxib 12.5 mg treatment group), however, the corresponding 95% CIs were all within the predefined comparability bounds that were set for the primary efficacy endpoints (±0.5 for Likert and ±10 mm for VAS). Secondary variables showed similar results as the primary variables.

Table I-7: Result after 12 weeks of treatment from Study 34.

End Point	Pairwise Treatment Difference in LSMean (95% CI)		
	25 mg—Diclofenac	12.5 mg—Diclofenac	25 mg—12.5 mg
Pain Walking on a Flat Surface (WOMAC) (0- to 100-mm VAS)	2.98 (-0.50, 6.45)	5.27 (1.79, 8.75)	-2.29 (-5.77, 1.19)
Patient Global Assessment of Response to Therapy (0 to 4 point Likert†)	0.16 (0.01, 0.31)	0.21 (0.06, 0.36)	-0.05 (-0.20, 0.10)
Investigator Global Assessment of Disease Status (0 to 4 point Likert)	0.13 (0.00, 0.25)	0.16 (0.04, 0.29)	-0.04 (-0.16, 0.09)

† Scale reversed so decreasing values indicate improvement to be consistent with presentation of the other end points.

Table I-8: Result after 12 weeks of treatment from Study 35.

End Point	Difference in LSMean (95% CI)		
	25 mg to Diclofenac	12.5 mg to Diclofenac	25 mg to 12.5 mg
Pain Walking on a Flat Surface (0- to-100-mm VAS) (WOMAC)	2.75 (-0.93, 6.43)	3.74 (0.07, 7.40)	-0.99 (-4.69, 2.72)
Patient Global Assessment of Response to Therapy (0- to 4-point Likert†)	0.19 (0.05, 0.33)	0.24 (0.10, 0.38)	-0.05 (-0.19, 0.09)
Investigator Global Assessment of Disease Status (0- to 4-point Likert)	0.17 (0.05, 0.30)	0.18 (0.06, 0.31)	-0.01 (-0.13, 0.11)
† Scale reversed so decreasing values indicate improvement to be consistent with presentation of the other end points.			

Table I-9: Result after 6 months of treatment from Study 34.

End Point	Pairwise Treatment Difference in LSMean (95% CI)		
	25 mg—Diclofenac	12.5 mg—Diclofenac	25 mg—12.5 mg
Pain Walking on a Flat Surface (WOMAC) (0- to 100-mm VAS)	3.40 (-0.07, 6.87)	5.53 (2.06, 9.00)	-2.13 (-5.60, 1.35)
Patient Global Assessment of Response to Therapy (0 to 4 point Likert†)	0.16 (0.01, 0.31)	0.21 (0.06, 0.36)	-0.05 (-0.20, 0.10)
Investigator Global Assessment of Disease Status (0 to 4 point Likert)	0.14 (0.01, 0.26)	0.17 (0.04, 0.29)	-0.03 (-0.16, 0.09)
† Scale reversed so decreasing values indicate improvement to be consistent with presentation of the other end points.			

Table I-10: Result after 6 months of treatment from Study 35:

End Point	Difference in LSMean (95% CI)		
	25 mg to Diclofenac	12.5 mg to Diclofenac	25 mg to 12.5 mg
Pain Walking on a Flat Surface (0- to 100-mm VAS) (WOMAC)	2.06 (-1.59, 5.71)	2.62 (-1.02, 6.26)	-0.56 (-4.24, 3.12)
Patient Global Assessment of Response to Therapy (0- to 4-point Likert†)	0.18 (0.05, 0.32)	0.21 (0.07, 0.34)	-0.02 (-0.16, 0.12)
Investigator Global Assessment of Disease Status (0- to 4-point Likert)	0.16 (0.04, 0.28)	0.15 (0.03, 0.28)	0.01 (-0.12, 0.13)
† Scale reversed so decreasing values indicate improvement to be consistent with presentation of the other end points.			

Results after 6 months of treatment:

Results at Month 6 were similar to the results after 12 weeks of treatment which were summarized in Table I-9 and Table I-10. This was not surprising because there was only one efficacy treatment measurement at Week 26 after week 12's assessment. Some differences were observed between study 34 and 35. Although within prespecified comparable range, in study 34, Diclofenac was consistently better than Rofecoxib 12.5

and 25 mg treatment groups and the treatment difference became slightly larger at Month 6; In study 35, however, the treatment difference became smaller at Month 6 as compared to the treatment difference at Week 12.

The results from LOCF analyses supported the conclusion of the primary analysis. However, the LOCF analyses showed larger treatment difference in favor of Diclofenac than the results of average change from baseline in Study 34 (Again, the difference was larger between Rofecoxib 12.5 mg and Diclofenac than that of Rofecoxib 25 mg). While in study 35, the difference became smaller than the results using average change from baseline analysis. Overall the analyses using average change from baseline were closer to the analysis at Week 12.

Comments on discontinuation due to lack of efficacy:

Since discontinuation due to lack of efficacy had direct impact on the efficacy assessment, it is worthwhile to discuss further because imbalances of drop out due to lack of efficacy were observed between treatment groups in the two long term studies.

In study 34, the percents of patients discontinued due to lack of efficacy at Month 6 were 9.52, 8.19, and 5.65% for the 12.5-mg, 25-mg Rofecoxib, and Diclofenac groups, respectively. The differences were not statistical significant. For study 35, the percent of patients who discontinued the study due to lack of efficacy was 13.13, 19.46, and 10.07% for 12.5, 25 mg Rofecoxib and Diclofenac, respectively. There were statistically significant fewer patient discontinuations due to lack of efficacy in the Diclofenac group as compared with the 25-mg Rofecoxib group ($p<0.003$).

Despite of the comparability results shown above, many efficacy end points showed that Rofecoxib treatment groups were statistically significantly inferior to declofenac, especially in 12.5 mg dose group of Rofecoxib. However, the comparability result was quite robust since the 97.5% CI was still within the comparability range.

To further understand the drop out pattern, Table I-11 listed the cumulative discontinuation rate at different time point. In study 35, an unusual phenomena was observed, i.e., Rofecoxib 25 mg treatment group had a higher drop out rate than that of Rofecoxib 12.5 mg through out the study. This drop out pattern was not observed in other OA studies.

Although drop out due to adverse event was a safety issue, it is worthwhile to discuss here in reviewing the whole picture of therapeutic effect of Rofecoxib. In both studies, Diclofenac group had larger portion of patient withdrawal due to clinical and laboratory adverse reaction than Rofecoxib dose groups. The number of patient withdrawal at Week 8, 12, and Month 6 by treatment groups were summarized in Table I-12. From efficacy point view, those patients should be considered as treatment failure as well. Taking account of this factor, it was comparable in terms of treatment failure between treatment groups.

In general, considering the discontinuation due to lack of efficacy alone, high rate of discontinuation in inferior treatment groups (Rofecoxib groups in this case) implies larger treatment difference existed than the difference found in the analysis of average change from baseline, LOCF and completer analysis. However, a reasonable conclusion should be drawn based on comprehensively reviewing all the facts listed above.

Table I-11: Discontinuation due to lack of efficacy.

	Study 34 Number/Total (%)			Study 35 Number/Total (%)		
	Week 8	Week 12	Month 6	Week 8	Week 12	Month 6
Rofecoxib 12.5	9/231 (3.9)	11/231 (4.8)	22/231 (9.5)	21/259 (8.1)	24/259 (9.3)	34/259(13.1)
Rofecoxib 25	7/232 (3.0)	12/232 (5.2)	19/232 (8.2)	24/257 (9.3)	37/257(14.4)	50/257(19.5)
Diclofenac	5/230 (2.2)	9/230 (3.9)	13/230(5.6)	10/268 (3.7)	18/268 (6.7)	27/268(10.1)

Table I-12: Discontinuation due to clinical and laboratory adverse reaction.

	Study 34 Number/Total (%)			Study 35 Number/Total (%)		
	Week 8	Week 12	Month 6	Week 8	Week 12	Month 6
Rofecoxib 12.5	6/231 (2.6)	9/231 (3.9)	18/231 (7.8)	14/259 (5.4)	22/259 (8.5)	30/259(11.6)
Rofecoxib 25	10/232 (4.3)	13/232 (5.6)	15/232 (6.5)	17/257 (6.6)	19/257(7.4)	27/257(10.5)
Diclofenac	13/230 (5.7)	22/230 (9.6)	32/230(13.9)	16/268 (6.0)	23/268 (8.6)	49/268(18.3)

II. GI Studies:

In order to evaluate the hypothesis that Rofecoxib was a COX-2 specific inhibitor that was not associated with GI adverse reactions – perforations, ulcers and bleeds (PUBs) in the treatment of patients with OA, four primary study reports were presented to address the issue. Studies 44 and 45 were two identically designed endoscopy studies on OA patients using endoscopy identified gastricdoudenal ulcers as surrogates of PUBs. The combined result of the two endoscopy studies was reported in study 44c to compare endoscopy identified ulcer between Rofecoxib and placebo. Study 69 was designed to monitor upper GI PUBs in 8 Phase III/II OA studies.

Studies 44, 45 and 44C:

Both 44 and 45 were multicenter, double-blind, randomized, parallel-group, active- and placebo-controlled studies to evaluate the effect of Rofecoxib 25, 50 mg once daily, Ibuprofen 800 mg 3 times daily, or placebo on the incidence of gastric duodenal ulcer following up to 24 weeks of treatment in patients with osteoarthritis. In these 2 studies, patients with a history of gastric or duodenal ulcer or upper gastrointestinal bleeding were allocated to one stratum while those without this history were allocated to a second stratum. At Week 16, 95% of patients in the placebo group and 5% of patients in the other treatment groups were discontinued in a blinded fashion; the rest of the patients were to complete 24 weeks of treatment. Patients were scheduled to undergo

esophagogastroduodenoscopy at baseline and following at Week 6, 12, 24 or at early discontinuation.

The primary objective of the two studies was to determine the comparative incidence of gastric and/or duodenal ulcer (≥3mm) following administration over 12 weeks of treatment. The endpoint was defined as endoscopy identified gastric and/or duodenal ulcers (≥3mm) in either stomach or duodenum over 12 weeks of treatment. The primary statistical analysis was life table analysis to compare time to the first ulcer. Two sided 90% CI was used in the evaluation.

Study 44c was designed to merge the results of studies 44 and 45 for the purpose of evaluating the incidence of gastric and/or duodenal ulcer between Rofecoxib and placebo. For this assessment, a prespecified clinical comparability bound of 4% was established (The upper limit of CI of the treatment difference between Rofecoxib and placebo in 12-week cumulative ulcer rate must fall below 4 percentage points).

Sponsor's analyses and conclusions:

There were comparable demographic information among treatment groups within each study and comparable information in patient accounting except there were low discontinuation rate due to AE and high discontinuation rate due to lack of efficacy in placebo treatment group.

The results after 12 weeks of treatment were summarized in Table II-1. The study results demonstrated that treatment with Ibuprofen in OA patients was associated with high rate of endoscopy identified gastricdoudenual ulcer.

The comparison between each Rofecoxib treatment groups and placebo gave inconsistent results between studies. The rate of ulcer (9.92%) at week 12 was unusually high in Study 44 in placebo treatment group, which was higher than that in both Rofecoxib dose groups. While in Study 45, the rate of ulcer in placebo treatment group was 5.10%, which was lower than Rofecoxib two dose groups. Obviously, there was study by treatment interaction.

Despite of the inconsistent results observed in Studies 44 and 45, the sponsor presented a combined analysis for the comparison between each Rofecoxib treatment and placebo. However, the interpretation became difficult because of the apparent treatment by study interaction.

Interestingly enough, the sponsor and the reviewer had two different results on the test of study by treatment interaction. The sponsor had p-value 0.26, while the reviewer had p-values < 0.1. The differences between the results were discussed in the reviewer's comments.

Table II-1: Sponsor's analysis:

Cumulative Ulcer rate (95%CI)	Study 44 (%)	Study 45 (%)	Study 44c (%)
Placebo	9.92 (4.12, 15.73)	5.10 (0.75, 9.46)	7.34 (3.78, 10.91)
Rofecoxib 25 mg	4.10 (1.12, 7.07)	5.29 (1.92, 8.66)	4.69 (2.45, 6.94)
Rofecoxib 50 mg	7.31 (3.31, 11.30)	8.82 (4.55, 13.09)	8.07 (5.15, 11.00)
Ibuprofen	27.69 (20.43, 34.95)	29.18 (22.15, 36.20)	28.47 (23.43, 33.52)
Difference (90% CI)			
Rofecoxib 25 mg – Placebo	-5.83 (-11.30, -0.35)	0.19 (-4.44, 4.81)	-2.32 (-5.85, 1.21)
Rofecoxib 50 mg – Placebo	-2.62 (-8.53, 3.30)	3.72 (-1.40, 8.84)	1.01 (-2.87, 4.88)
Rofecoxib 25 mg – Ibuprofen	-23.59 (-30.17, -17.01)	-23.89 (-30.43, -17.35)	-23.74 (-28.38, -19.10)
Rofecoxib 50 mg – Ibuprofen	-20.38 (-27.33, -13.43)	-20.35 (-27.25, -13.45)	-20.37 (-25.26, -15.47)

Reviewer's comments and analyses:

Studies 44 and 45 had two control groups. The comparison between Rofecoxib and Ibuprofen was one family, while the comparison between Rofecoxib and placebo was another family. Each family controlled its own alpha level at 0.05. Since there was a very large difference in ulcer rates between Ibuprofen and Rofecoxib, the results should be very robust. Therefore, I only focus the discussion on the comparison between Rofecoxib and placebo in this section, although some of the comments and analyses can also be applied to the comparison of Rofecoxib with Ibuprofen.

1. The study design was problematic in evaluating equivalence comparison using time to the first endoscopy identified ulcer. The endoscopy was only undertaken at week 0, 6, 12 and at early discontinuation, which was not frequent enough to capture ulcers occurred in between, or introduce interval censoring. The endoscopy done on early discontinuation could also introduce bias because of the increased surveillance. Such a study design could diminish treatment difference if there was any, which made it easier to conclude that there was no difference (or equivalence) between treatment groups.

2. There were some patients who discontinued the study due to erosions identified in their GI system. Two patients who discontinued and had number of erosions greater than 10 in Study 44 and 13 patients in Study 45. Number of erosions greater than 10 can be considered as an ulcer (refer to Dr. Goldkind's review), and will be identified as an ulcer if followed up further. Therefore patients discontinued due to erosion were informative censoring to endoscopy identified ulcers. Ignoring such informative censoring, the ulcer rates estimated by survival analysis were biased. To correct the bias, patients who had erosion and discontinued were counted as having ulcers.

3. In evaluating NDA, two sided 95% CI interval is usually applied in equivalence comparison. This is consistent to one sided test at alpha level equal to 0.025. The analyses corrected by the erosion drop out was presented in Table II-2 with two sided 95%CI for both Studies 44 and 45. As it can been seen, the sponsor's results were not robust. The upper limits of 95% CI exceeded 4% for Rofecoxib 50 mg in both studies.