CENTER FOR DRUG EVALUATION AND RESEARCH

APPLICATION NUMBER: 21-042/S-007, S-008, S-010, S-012, S-013, S-014 and 21-052/S-004, S-005, S-006, S-007, S-008, S-009

STATISTICAL REVIEW(S)

STATISTICAL REVIEW AND EVALUATION

NDA #:	21,042/S-12
Applicant:	Merck Research Lab.
Name of Drug:	Vioxx (rofecoxib tablets)
Indication:	Treatment of rheumatoid arthritis
Route of Administration:	Orally once daily
Documents Reviewed:	Clinical/Statistical section of the NDA electronic submission dated 02/28/01 and re-analyses submitted on 5/16/01 and 6/22/01
Medical Officer:	Joel Schiffenbauer, M.D., HFD-550
Statistical Reviewer:	Shiowjen Lee, Ph.D., HFD-725

I. Introduction and Background

Vioxx (rofecoxib) Tablets 12.5 mg and 25 mg have been approved and marketed in the US since 1999 for relief of the signs and symptoms of osteoarthritis (OA), the management of acute pain, and the treatment of primary dysmenorrhea. The Sponsor's current submission, an NDA supplement, claims that rofecoxib is efficacious and safe in the treatment of rheumatoid arthritis (RA).

Eight clinical studies, which included two Phase III pivotal trials (protocols #096 and #097), one dose-ranging study (protocols #68P1, 68P2 and 68X) and — supportive studies, were submitted in clinical/statistical section of the NDA. The — supportive studies included [redacted], and one study on the incidence of gastroduodenal ulcers in patients with RA (protocol #98).

On the filing date of the NDA, only data up to 14 weeks of study therapy were included and analyzed for Studies #096 and #097. The efficacy evaluation of rofecoxib is mainly based on the data obtained from Part I (i.e. the 1st 12 weeks). This statistical review will focus on the efficacy evaluation of the two Phase III pivotal studies and the dose-ranging trial. Without ambiguity, trials will be referred to as Study 96, Study 97, and Study 68, throughout the review.

The layout of this review is described as follows. Section II summarizes and discusses the efficacy of rofecoxib in the treatment of RA indication. In Study 96, two doses of rofecoxib, 12.5 and 25 mg, were tested and their efficacy was compared with those of placebo and naproxen. For Study 97, two doses of rofecoxib, 25 and 50 mg were compared with placebo and naproxen. Section III summarizes and discusses the dose-ranging study. The summary and conclusion of the efficacy of rofecoxib are included in Section IV.

II. Efficacy Review

Study Design of #96 and #97:
The studies were designed as 2-part, double-blind, parallel-group, 52-week and multicenter. Study 96 included 94 US sites and 6 foreign sites; while all 91 centers in Study 97 were outside the US. Part I of the studies was a 12-week, active-comparator and placebo-controlled period to evaluate the efficacy of rofecoxib 12.5- and 25-mg for treatment of RA in Study 96, and rofecoxib 25-, and 50-mg in Study 97. The efficacy assessment was measured at the weeks 2, 4, 8, and 12 of study therapy. Part II of the studies was a 40-week, active-comparator-controlled period. The treatment assignment in Part II for each study is described below.

For Study 96,
- The subjects who received placebo in Part I were re-assigned, in approximately equal proportions, to 25-mg rofecoxib and naproxen in Part II.
- Subjects who received 12.5-mg rofecoxib in Part I received 25-mg rofecoxib in Part II
- Half of the subjects who received 25-mg rofecoxib in Part I received 50-mg rofecoxib in Part II, the other half continued on 25-mg rofecoxib
- Subjects who received naproxen in Part I continued on the same treatment in Part II.

For Study 97,
- Subjects in placebo arm in Part I were re-assigned, in approximately equal proportions, to 25-mg rofecoxib and naproxen in Part II
- Half of the subjects who received 25-mg rofecoxib in Part I received 50-mg rofecoxib in Part II, the other half continued on the same treatment
- Subjects who received 50-mg rofecoxib or naproxen in Part I continued the same treatment in Part II.

The treatment re-assignment process resulted in the three treatments in Part II – rofecoxib 25-, 50-mg and naproxen.

The treatment assignment for parts I and II described above was based on a computer-generated allocation schedule. The enrolled subjects were randomized into six treatment sequences (part I/part II) for each study. Treatment allocation was stratified on the basis of corticosteroid usage (i.e. user or non-user). The six treatment allocation sequences used in the randomization for Study 96 were placebo/25 mg, placebo/naproxen, 25 mg/25 mg, 25 mg/50 mg 12.5 mg/25 mg, and naproxen/naproxen; while the six sequences in Study 97 were placebo/25 mg, placebo/naproxen, 25 mg/25 mg, 25 mg/50 mg, 50 mg/50 mg and naproxen/naproxen. It should be noted that the blinding was maintained for Parts I and II based on the randomization scheme.

The objectives of the two pivotal studies stated in the protocols were:
1. to demonstrate superior clinical efficacy for rofecoxib 25-mg daily, compared with placebo, in the treatment of RA over a 12-week period,
2. to demonstrate safety and tolerability for rofecoxib 25- and 50-mg daily over a 1-year treatment period in RA patients,
3. to explore the efficacy of rofecoxib 12.5-mg daily for the treatment of RA,

4. to explore the efficacy response to fixed-dose escalation from12.5 to 25 and from 25 to 50 mg of rofecoxib daily,
5. to assess the clinical efficacy of naproxen 500-mg twice daily over a 12-week period, and
6. to assess the maintenance of therapeutic effects for rofecoxib 25- and 50-mg daily, and naproxen 500-mg twice daily, over a 1-year period.

Population Analyzed:
The Sponsor's modified intent-to-treat (MITT) population included all randomized subjects who had a baseline and at least one post-baseline data. Their primary analysis was based on such population. This review will refer this population as MITT instead of ITT as referenced by the Sponsor in the submission.

The Per-Protocol (PP) analysis was also included in the Sponsor's NDA for the primary efficacy endpoints. The PP population excluded patients and/or data points with clinically important protocol deviations based on pre-specified criteria.

Efficacy Endpoints:
The therapeutic effectiveness of rofecoxib was evaluated based on four primary efficacy endpoints and several other endpoints in the Sponsor's submission. The primary efficacy endpoints are:

- tender joint counts (total of 68 counts)
- swollen joint counts (total of 66 counts)
- patient's global assessment of disease activity
- investigator's global assessment of disease activity

The patient's global assessment of disease activity was measured on a 0 – 100 mm visual analog scale (VAS) and the investigator's global assessment was based on a 0 – 4 point Likert scale.

Several secondary efficacy endpoints were proposed. They are:

- proportion of subjects achieving American College of Rheumatology 20% (ACR20) response and completing Part I
- patient's global assessment of pain (0 – 100 mm VAS)
- modified Health Assessment Questionnaire (HAQ, 0 – 3 point Likert scale).

Other efficacy endpoints included

- patient's global assessment of response to therapy (0 – 4 grading scale)
- investigator's global assessment of response to therapy (0 – 4 grading scale)
- discontinuation due to lack of efficacy
- duration of morning stiffness (in minutes)
- acetaminophen (for rescue) tablet count
- serum C-reactive protein, and
- short Form-36 Health Survey (SF-36)

The ACR20 responder index is a composite of improvement to demonstrate therapeutic response in the treatment of RA. Please see clinical review for detailed definition.

Statistical Analysis Plan Proposed by the Sponsor:
Sponsor's primary analysis was the time-weighted average change from baseline for each of the four primary efficacy endpoints. The primary analysis was based on analysis of covariance (ANCOVA), adjusting for corticosteroid use, and baseline.

Additionally, patient's last observed on-treatment value on the primary efficacy endpoints was also analyzed using the same ANCOVA model. The proportions of patients who met the ACR20 criteria were analyzed based on Cochran-Mantel-Haenszel test adjusting for corticosteroid use. Fisher's exact test was used to analyze the proportions of patients' drop out due to lack of efficacy.

Superiority and Comparability Criteria:
The primary comparison was rofecoxib against placebo; while the comparison between rofecoxib and naproxen was the secondary. Both p-value and 95% confidence intervals for two treatment mean difference were provided. The criterion for superiority evaluation was that p-value < 0.05. However, no comparability criteria were given for the similarity evaluation between rofecoxib and naproxen.

Multiplicity Issues:
Multiple controls: Sponsor applied closed testing procedure in the comparisons. For Study 96, the 25-mg rofecoxib must show significance with respect to placebo. Otherwise, there would be no further comparison between rofecoxib and naproxen. For Study 97, unless 25- and 50-mg rofecoxib are superior to placebo, no further consideration on the comparison between rofecoxib and naproxen would be made.

Multiple dose groups: There were two doses of rofecoxib within each study (i.e. 12.5 and 25 mg in Study 96; and 25 and 50 mg in Study 97). The Sponsor's Data Analysis Plan (DAP) implied that 25- AND 50-mg were their primary goal for efficacy claim of the test drug.

Multiple endpoints: The Sponsor stated that they followed the FDA RA Guidance, which mentioned the 3 out of 4 criterion as an alternative option. Due to the multiplicity issue based on such rule, the Sponsor specified that 3 primary endpoints (i.e. tender joint counts, patient's global assessment of disease severity, and investigator's global assessment of disease) must show superiority to placebo. Therefore, no multiplicity adjustment should be made.

Reviewer's Comments:
1. It should be noted that on the filing date of the NDA, efficacy data up to only 14 weeks were submitted and analyzed. Therefore, this review could only comment Sponsor's objective (1), (3), (4) and (5).

2. Sponsor's primary analysis in the submission was the time-weighted average change from baseline for each of the four primary efficacy endpoints on the modified ITT population, as specified in the protocol amendments (dated 6/26/00 and 6/23/00 for studies 96 and 97, respectively). The population included all randomized subjects who had baseline and at least one post-baseline data. The Agency had requested analyses for the following two populations:

- Population includes all randomized subjects regardless of having any post-baseline data
- Population includes all randomized subjects who took at least one dose of drug, regardless of having any post-baseline data.

The above two populations resulted in the identical analyses since all randomized subjects took at least one dose of drug. This review refers the two populations as ITT. It should be noted that the difference between the modified ITT and ITT populations is small (see patient disposition section for details). Therefore, the efficacy results in terms of the time-weighted average change from baseline based on modified ITT and ITT populations are expected to be similar.

For handling missing data, no scheme is needed for the Sponsor's MITT analyses in terms of the time-weighted average change, as subjects in the analysis must had at least one post-baseline data. Sponsor's re-analysis (i.e. ITT analyses) imputed subjects having missing data with zero change. The impact of such imputation will be commented in the efficacy results section. Additionally, the Sponsor also submitted the last-observed on-treatment data analysis, which is similar to the last observation carried forward analysis. To evaluate the robustness of the studies, the comparisons among different analyses will be commented in the efficacy results section.

3. Both Phase III pivotal studies were designed as multicenter. However, the primary analysis based on ANCOVA did not include center as one of the main effects. This issue will be addressed in the section of efficacy results.

4. The Sponsor claimed (in their labeling) that rofecoxib 25-, 50-mg have similar effect to naproxen in treating RA indication. However, no clinically important comparability bound was pre-defined in the protocols for the comparison despite the 95% confidence intervals for pairwise difference between treatments covering 0 and p-value $\geq$ 0.05. It should be noted that the conclusion of p-value $\geq$ 0.05 and 95% confidence intervals including 0 does not imply that two treatments are approximately similar. The assessment should be based on 95% confidence intervals along with the use of pre-defined comparability bounds.

5. Four primary efficacy endpoints were tested in the Sponsor's submission. The Agency had recommended ACR20 as the primary efficacy endpoint at the End-of-Phase 2 meeting (4/30/1998). However, the Sponsor indicated that the FDA RA Guidance mentioned the 3 out of 4 criterion as an alternative option. Due to the multiplicity issue based on such rule, the Sponsor specified in the protocols that 3 primary endpoints (i.e. tender joint counts, patient's global assessment of disease severity, and investigator's global assessment of disease) must show significance with respect to placebo. This will be discussed in details in the section of efficacy results.

Efficacy Results:

1. Patient Disposition and Baseline Characteristics

To evaluate the comparability between treatments, Table 1 presents the patient disposition for Studies 96 and 97. Generally, the discontinuation rate from part I was comparable between treatment groups within each study though placebo group may have higher rates than others.

Table AII.1 of the Appendix summarizes the number of subjects (and percentage) included in MITT and PP populations for each primary endpoint. The difference between ITT and MITT populations is small and treatment groups were comparable within the MITT and PP populations. For the comparison between treatments with respect to baseline demographics and characteristics, no outstanding discrepancies between treatments are identified in the NDA submission.

Table 1. Patient Disposition: Study 96 and Study 97

Study 96		Placebo	Rofecoxib 12.5 mg	Rofecoxib 25 mg	Naproxen 1000 mg	Total
	Entered	301	148	311	149	909
	Continuing study (at end of Part I – week12)	201 (66.8%)	110 (74.3%)	245 (78.8%)	118 (79.2%)	674 (74.1%)
	Discontinued (from Part I)	100 (33.2%)	38 (25.7%)	66 (21.2%)	31 (20.8%)	235 (25.9%)
	Clinical adverse event	10 (3.3%)	5 (3.4%)	16 (5.1%)	7 (4.7%)	38 (4.2%)
	Lab. Adverse event	0	0	1 (0.3%)	0	1 (0.1%)
	Lack of efficacy	80 (26.6%)	26 (17.6%)	33 (10.6%)	18 (12.1%)	157 (17.3%)
	Lost to follow-up	1 (0.3%)	2 (1.4%)	0	0	3 (0.3%)
	Protocol deviation	5 (1.7%)	3 (2.0%)	7 (2.3%)	2 (1.3%)	17 (1.9%)
	Others	4 (1.3%)	2 (1.4%)	9 (2.9%)	4 (2.7%)	19 (2.1%)
Study 97		**Placebo**	**Rofecoxib 25 mg**	**Rofecoxib 50 mg**	**Naproxen 1000 mg**	**Total**
	Entered	299	315	297	147	1058
	Continuing study (at end of Part I – week12)	237 (79.3%)	281 (89.2%)	250 (84.2%)	126 (85.7%)	894 (84.5%)
	Discontinued (from Part I)	62 (20.7%)	34 (10.8%)	47 (15.8%)	21 (14.3%)	164 (15.5%)
	Clinical adverse event	14 (4.7%)	12 (3.8%)	24 (8.1%)	12 (8.2%)	62 (5.9%)
	Lab. Adverse event	0	0	1 (0.3%)	0	1 (0.1%)
	Lack of efficacy	39 (13.0%)	16 (5.1%)	13 (4.4%)	5 (3.4%)	73 (6.9%)
	Lost to follow-up	0	0	2 (0.7%)	0	2 (0.2%)
	Protocol deviation	5 (1.7%)	2 (0.6%)	4 (1.3%)	2 (1.4%)	13 (1.2%)
	Others	4 (1.3%)	4 (1.3%)	3 (1.0%)	2 (1.4%)	13 (1.2%)

Source: Sponsor's electronic NDA submission (page 19 in study 96 and page 19 in study 97).

2. Primary Efficacy Endpoints

Superiority Comparisons:
In studies 96 and 97, rofecoxib 25- and 50-mg demonstrated statistically significantly greater improvement than placebo over 12-week treatment as assessed by the time-weighted average change for the four primary endpoints. Tables 2 and 3 present the efficacy results of the two studies.

The results of Tables 2 and 3 are summarized below:

- The time-weighted average analyses based on MITT and ITT populations are consistent as expected.
- In Study 96, rofecoxib 25-mg and naproxen demonstrate significantly greater improvement than placebo, but rofecoxib 12.5-mg does not show the superiority effect over placebo in tender and swollen joint counts (i.e. p-value ≥ 0.091 for tender joint counts; and p-value ≥ 0.120 for swollen joint counts).

- Rofecoxib 25- and 50-mg are superior to placebo in Study 97. However, naproxen is not shown to be significantly better than placebo in swollen joint counts (p-value ≥ 0.094).

Table 2: Primary Efficacy Results from Study 96

Primary Endpoints	Pairwise Treatment Difference in LSMean (95% Confidence Interval) p-value for the comparison of two treatments		
	12.5 mg vs. placebo	25 mg vs. placebo	Naproxen vs. placebo
MITT Analyses			
Tender joint counts	-1.50 (-3.37, 0.36) 0.114	-2.73 (-4.23, -1.23) < 0.001	-3.09 (-4.94, -1.24) 0.001
Swollen joint counts	-0.91 (-2.12, 0.30) 0.142	-1.22 (-2.19, -0.24) 0.014	-1.73 (-2.93, -0.53) 0.005
Patient's global	-5.33 (-9.34, -1.32) 0.009	-7.18 (-10.4, -3.95) < 0.001	-10.4 (-14.4, -6.45) < 0.001
Investigator's global	-0.17 (-0.33, -0.01) 0.041	-0.32 (-0.45, -0.19) < 0.001	-0.27 (-0.43, -0.11) 0.001
ITT Analyses			
Tender joint counts	-1.60 (-3.46, 0.25) 0.091	-2.91 (-4.41, -1.42) < 0.001	-3.37 (-5.22, -1.51) < 0.001
Swollen joint counts	-0.95 (-2.15, 0.25) 0.120	-1.31 (-2.28, -0.34) 0.008	-1.87 (-3.07, -0.67) 0.002
Patient's global	-5.17 (-9.14, -1.20) 0.011	-7.36 (-10.57, -4.16) < 0.001	-10.95 (-14.9, -6.99) < 0.001
Investigator's global	-0.17 (-0.33, -0.01) 0.041	-0.33 (-0.46, -0.20) < 0.001	-0.29 (-0.45, -0.13) < 0.001

Note: Decreasing values indicate improvement.
Source: Sponsor's electronic NDA submission (pages 114-120 in study 96) and re-analysis dated 5/16/01.

Table 3: Primary Efficacy Results from Study 97

Primary Endpoints	Pairwise Treatment Difference in LSMean (95% Confidence Interval) p-value for the comparison of two treatments		
	25 mg vs. placebo	50 mg vs. placebo	Naproxen vs. placebo
MITT Analyses			
Tender joint counts	-2.96 (-4.56, -1.36) < 0.001	-4.00 (-5.63, -2.38) < 0.001	-3.95 (-5.95, -1.96) < 0.001
Swollen joint counts	-1.24 (-2.23, -0.26) 0.014	-1.16 (-2.16, -0.16) 0.023	-1.00 (-2.23, 0.24) 0.113
Patient's global	-6.95 (-10.06, -3.85) < 0.001	-9.77 (-12.93, -6.62) < 0.001	-9.59 (-13.47, -5.70) < 0.001
Investigator's global	-0.33 (-0.45, -0.21) < 0.001	-0.37 (-0.49, -0.24) < 0.001	-0.43 (-0.59, -0.28) < 0.001
ITT Analyses			
Tender joint counts	-3.13 (-4.73, -1.54) < 0.001	-4.08 (-5.70, -2.46) < 0.001	-4.03 (-6.02, -2.04) < 0.001
Swollen joint counts	-1.34 (-2.32, -0.35) 0.008	-1.21 (-2.20, -0.21) 0.018	-1.05 (-2.27, 0.18) 0.094
Patient's global	-7.23 (-10.33, -4.12) < 0.001	-9.93 (-13.07, -6.78) < 0.001	-9.54 (-13.41, -5.66) < 0.001
Investigator's global	-0.34 (-0.46, -0.22) < 0.001	-0.36 (-0.49, -0.24) < 0.001	-0.44 (-0.59, -0.29) < 0.001

Note: Decreasing values indicate improvement.
Source: Sponsor's electronic NDA submission (pages 105-111 in study 97) and re-analysis dated 5/16/01.

Discussion:
It should be noted that the Sponsor's MITT results and re-analyses (i.e. ITT analyses in the review) were based on time-weighted average change from baseline. Two issues are arisen:

1. Are the results robust? How does the analysis based on time-weighted average change differ from other types of analysis? For instance the mean change from baseline to Week 12.
2. Sponsor's re-analyses imputed the missing value of change as 0. What is the impact of this imputation method on the analysis?

The following comments discuss Issue 1:

- The PP analyses based on time-weighted average change from baseline were generally consistent with those based on MITT and ITT populations for both studies (see Table AII.2 of the Appendix for summary). No significant discrepancies were identified.
- Sponsor's last observed on-treatment data analysis yielded generally consistent conclusions except the comparisons of rofecoxib 12.5-mg vs. placebo in patient's and investigator's global assessment of disease activity, and naproxen vs. placebo in investigator's global assessment in Study 96. A marginal significance of rofecoxib 25- and 50-mg with respect to placebo is noted in swollen joint counts for studies 96 and 97, respectively (i.e. p-value = 0.059 and 0.056 versus 0.014 and 0.023). The results of the last observed on-treatment analyses are summarized in Table AII.3 of the Appendix.

  The MITT analyses based on the time-weighted average change showed significant superiority of rofecoxib 12.5-mg over placebo (i.e. p-value = 0.009 and 0.041 for patient's and investigator's global assessment, respectively, in Table 2). These are in contrast to the results based on the last observed on-treatment data analyses (i.e. p-value = 0.118 and 0.284, respectively, in Table AII.3). Compared with placebo, naproxen demonstrated a significant superiority effect based on MITT time-weighted average analysis (i.e. p-value = 0.001 in Table 2), however, exhibited a non-significant result based on the last observed on-treatment data analysis (i.e. p-value = 0.209, Table AII.3).

  To further understand these discrepancies, the changes from baseline based on the time-weighted average and the last-observed on-treatment data in the four primary endpoints were summarized in Table AII.4 of the Appendix. As can be observed, the standard error term (i.e. S.E.) for time-weighted average is smaller than that of the last-observed on-treatment data. For patient and investigator's global assessments in study 96, placebo group has better improvement based on the last-observed on-treatment data. On the contrary, the improvement in rofecoxib 12.5 mg and naproxen groups is worsened. The different conclusion for the comparisons is attributed to these factors. On the other hand, all treatment groups had better improvement based on the last-observed on-treatment data in study 97 and the improvement amount was similar. Therefore, there was not much difference in the conclusion between the time-weighted average and the last-observed on-treatment data analyses in study 97.

  Above all, it reveals that time-weighted average analysis has smaller standard errors, which may detect treatment difference easier than the last-observed on-treatment data analysis.

Having said that, the comparisons of rofecoxib 25- and 50-mg vs. placebo are generally robust.

Issue 2 is discussed below:

- Sponsor's re-analyses imputed the change from baseline as 0 for subjects who were not included in the MITT population. Eleven, 5, 4 and 0 patients (i.e. 3.7%, 3.4%, 1.3% and 0) were distributed in placebo, rofecoxib 12.5 mg, rofecoxib 25 mg, and naproxen groups for Study 96, and 10, 2, 6 and 2 patients (i.e. 3.3%, 0.6%, 2%, and 1.4%) were in placebo, rofecoxib 25 mg, rofecoxib 50 mg, and naproxen groups for Study 97. Apparently, the zero-imputation method is in favor of the test drug (i.e. due to the fact that placebo group had a higher rate of subjects who were not included in the MITT analysis). Analysis based on the change from baseline to Week 12 and imputing the missing values by the average of the response resulted in similar conclusions to the last-observed on-treatment data analysis. The results are summarized in Table AII.5 of the Appendix.

Based on the above discussions, it can be concluded that

- efficacy results are generally robust for the comparisons of rofecoxib 25-mg and 50-mg against placebo.
- Rofecoxib tablets 25- and 50-mg demonstrate significantly greater improvement than placebo in treating RA indication over 12-week therapy.
- However, rofecoxib 12.5-mg is ***not*** proven to be efficacious in study 96, as it failed tender and swollen joint counts based on time-weighted average analysis, and failed all 4 primary endpoints based on the last-observed on-treatment data analysis.
- Naproxen is ***not*** shown efficacious in investigator's global assessment based on the last-observed on-treatment data analysis for study 96, and failed swollen joint counts in both time-weighted average and the last-observed on-treatment data analysis in study 97.

One may argue the marginal significance of rofecoxib 25- and 50-mg with respect to placebo in swollen joint counts based on the last-observed on-treatment data analysis and change from baseline to Week 12 using average response for imputation. However, it should be noted that the Sponsor had specified in their protocol amendments that all four primary endpoints except the swollen joint counts must show superiority over placebo. Therefore, no multiplicity adjustment would be needed as they have demonstrated the superiority of rofecoxib 25- and 50-mg over placebo in tender joint counts, patient's and investigator's global assessments.

Comparison with Naproxen:
There were no pre-defined comparability criteria for the comparison between rofecoxib and naproxen despite all 95% confidence intervals for the pairwise treatment difference including 0 and p-value ≥ 0.05. The results are summarized in Table AII.6 of the Appendix. It should be noted that the similarity evaluation of two drugs relies on the setup of comparability bounds. Therefore, the evaluation of similar therapeutic effect between rofecoxib (i.e. 25- and 50-mg) and naproxen in the Sponsor's NDA submission should be of less value.

3. Secondary Efficacy Endpoints

Superiority Comparisons:
Studies 96 and 97 showed that rofecoxib 25- and 50-mg demonstrated greater improvement than placebo over 12-week treatment as assessed by the three secondary efficacy endpoints. In addition, rofecoxib 12.5 mg showed superiority efficacy effect over placebo. The efficacy results of the two studies are summarized in Tables 4 and 5.

Table 4: Secondary Efficacy Results from Study 96

	Pairwise Treatment Difference (95% Confidence Interval) p-value for the comparison of two treatments		
Primary Endpoints	**12.5 mg vs. placebo**	**25 mg vs. placebo**	**Naproxen vs. placebo**
MITT Analyses			
ACR20 responder and completer *	12.16% (2.59%, 21.73%) 0.017	21.14% (13.52%, 28.77%) < 0.001	22.72% (13.15%, 32.28%) < 0.001
Patient's global assess. of pain #	-4.35 (-8.35, -0.35) 0.033	-8.27 (-11.47, -5.06) < 0.001	-9.06 (-13.00, -5.12) < 0.001
HAQ change #	-0.10 (-0.19, -0.01) 0.026	-0.18 (-0.25, -0.11) < 0.001	-0.20 (-0.29, -0.11) < 0.001

Note: * Proportion of subjects achieving ACR20 criteria. Larger % represents improvement
Decreasing values indicate improvement.
Source: Sponsor's electronic NDA submission (pages 122-126 in study 96).

Table 5: Secondary Efficacy Results from Study 97

	Pairwise Treatment Difference (95% Confidence Interval) p-value for the comparison of two treatments		
Primary Endpoints	**25 mg vs. placebo**	**50 mg vs. placebo**	**Naproxen vs. placebo**
MITT Analyses			
ACR20 responder and completer *	9.50% (1.64%, 17.37%) 0.018	12.20% (4.22%, 20.19%) 0.003	11.72% (1.87%, 21.57%) 0.017
Patient's global assess. of pain #	-8.20 (-11.40, -5.01) < 0.001	-10.40 (-13.64, -7.16) < 0.001	-11.91 (-15.90, -7.92) < 0.001
HAQ change #	-0.19 (-0.26, -0.12) < 0.001	-0.20 (-0.27, -0.13) < 0.001	-0.18 (-0.26, -0.09) < 0.001

Note: * Proportion of subjects achieving ACR20 criteria. Larger % represents improvement
Decreasing values indicate improvement.
Source: Sponsor's electronic NDA submission (pages 113-117 in study 97).

Table 6: Analyses of ACR20 Responders Based on Worst Scenario Case

Study 96	**placebo**	**Rofecoxib 12.5**	**Rofecoxib 25**	**naproxen**
% of ACR20 responder and completer	94/301 (31.2%)	62/148 (41.9%)	160/311 (51.4%)	79/149 (53.0%)
p-value (w.r.t. placebo)	N/A	0.037	< 0.001	< 0.001
Study 97	**placebo**	**Rofecoxib 25**	**Rofecoxib 50**	**naproxen**
% of ACR20 responder and completer	123/299 (41.1%)	157/315 (49.8%)	155/297 (52.2%)	76/147 (51.7%)
p-value (w.r.t. placebo)	N/A	0.030	0.007	0.031

Source: Reviewer's analysis based on the data sets Sponsor submitted.

Discussion:
For the calculation of ACR20 response in the Sponsor's modified ITT analyses, 4, 2, 0 and 0 patients in placebo, rofecoxib 12.5 mg, rofecoxib 25 mg and naproxen groups in Study 96, and 4,

0, 2 and 1 patients in placebo, rofecoxib 25 mg, rofecoxib 50 mg and naproxen in Study 97 were not included for analyses. In the Sponsor's re-analyses (i.e. ITT analyses per the Agency's request), subjects with missing ACR20 data were treated as failures. Obviously, such re-analyses may be in favor of rofecoxib. The "worse scenario case" analyses are done by imputing patients as successes in placebo groups and failures in active treatments. The results are consistent and are presented in Table 6.

Comparison with Naproxen:
As discussed earlier in the section of primary endpoints, no pre-defined comparability criteria were given for the similarity comparisons even though all 95% confidence interval of treatment difference included 0 and p-value ≥ 0.05. The results are summarized in Table AII.7 of the Appendix. The conclusion of similar therapeutic effect between rofecoxib and naproxen in the Sponsor's NDA submission should be of less value.

4. Other Efficacy Endpoints

In general, the results of other efficacy endpoints for rofecoxib 25- and 50-mg are consistent with those of the primary and secondary efficacy endpoints. Results are summarized below:

- Rofecoxib 25- and 50 mg demonstrated statistically significantly greater improvement than placebo in
  - patient's global assessment of response to therapy (0 to 4 grading scale)
  - investigator's global assessment of response to therapy (0 to 4 grading scale)
  - duration of morning stiffness (in minutes)
  - acetaminophen (for rescue) tablet count, and
  - short Form-36 Health Survey (SF-36 Physical)
- Naproxen 500 mg twice daily also show significantly improvement relative to placebo in these efficacy endpoints

5. Discontinuation and Missing Values

Patient dropout and missing values could have impact on the efficacy evaluation. Therefore, it is worthwhile to discuss such impact, if any.

In studies 96 and 97, the discontinuation rate due to lack of efficacy for placebo groups is statistically significantly higher than other treatment groups (p-value ≤ 0.044). However, the treatment groups are comparable within each study with respect to the discontinuation due to adverse events and laboratory adverse events (See Table 1). Even though rofecoxib and naproxen groups may have numerically higher discontinuation rates due to adverse events, the difference is not significant. Thus, the imbalance of dropout due to lack of efficacy is apparent.

To see the efficacy impact over various reasons of discontinuation, the results of average change from baseline for the four primary endpoints are summarized in Table AII.8 of the Appendix. Generally, the change from baseline for subjects who discontinued due to various reasons is quite consistent with the MITT analyses. Subjects in placebo group who discontinued had numerically higher improvement than other active treatment groups in some efficacy endpoints. On the other hand, subjects in rofecoxib groups who discontinued due to adverse events had

generally better treatment effect than placebo. Therefore, the impact on the overall analysis is not pronounced combining these factors.

Additionally, the pattern of missing values by assessment week for the four primary endpoints is presented in Table AII.9 of the Appendix. In general, the rates at each treatment were similar across endpoints within each study. Placebo group had higher missing data rates than other treatments for all assessment weeks. In study 96, rofecoxib 25 mg had numerically higher rates of missing values than naproxen at weeks 4 and 8. On the contrary, naproxen had higher rates than rofecoxib in study 97. Despite the different results, the rates of missing data at Week 12 between rofecoxib and naproxen were not significantly different within each study.

6. Center and Low-Dose Corticosteroid Use Effects

Due to the fact that most centers enrolled a very small number of subjects for each treatment, where some treatment groups have zero enrollment, the homogeneity assessment of efficacy between centers was difficult. Therefore, the center effect was not included in ANCOVA analysis for the four primary endpoints in the NDA submission. To uncover any possible different response due to center effect, the treatment response for each center was examined. Generally, Rofecoxib 25-, 50-mg daily and Naproxen 500-mg twice daily demonstrated consistently greater improvement than placebo. A number of centers showed opposite outcomes in the 4 primary endpoints as compared to placebo. However, the definite conclusion cannot be drawn due to relatively small sample sizes within individual centers.

The factor of low-dose corticosteroid use was included in the Sponsor's ANCOVA analysis as one of the main effects. It is worth to mention its impact on the efficacy results. Rofecoxib 25- and 50-mg, and naproxen demonstrated superiority over placebo numerically regardless of the corticosteroid use. In study 96, the effect of corticosteroid use in ANCOVA analysis for patient's global and investigator's global assessment was significant. In reviewing the efficacy results, the non-corticosteroid users had better improvement than users. On the other hand, the difference between corticosteroid users and non-users for the primary efficacy endpoints in study 97 is non-significant.

Other subgroup analyses suggest that the treatment effects of rofecoxib 25- and 50-mg were generally similar. No significant and clinically meaningful disparity conclusion among subgroups could be drawn.

7. Dose Escalation from Part I to II

Patient's responses to change of therapy from Part I to II were assessed using the average of the last 2 observations in Part I and the average of the first 2 assessments in Part II. Patients who received placebo at Part I and re-assigned to rofecoxib 25-mg or Naproxen appear to show greater improvement numerically. The responses for patients who did not change therapy from Part I to II were either to maintain the effect or to improve. The results were summarized in Table AII.10(a) and AII.10(b) of the Appendix for studies 96 and 97, respectively.

From statistical point of view, the validity of dose escalation analyses, however, is not clear despite the Sponsor's conclusion of better improvement numerically or non-significant dose escalation effect for rofecoxib. The reasons are:

(1) the time interval for the last 2 observations in Part I is different from that for the 1st 2 assessment at Part II (i.e. Week 8 – Week 12 vs. Week 12 – Week 14);
(2) the treatment effect based on the last 2 observations from Part I was conditioned on the same treatment from baseline; while the 1st two assessments at Part II were obtained within a relatively short time based on different dose. Drug effect may take a certain length of time to show efficacy, whether positive or negative.

The medical reviewer should address the issue of dose escalation effect taken into the consideration of safety assessment.

8. Efficacy Results over 14-Week of Treatment

The efficacy results for patients who were under the same treatment in Parts I and II over 14 weeks generally show consistent results with those over 12-week (See Tables AII.6 and AII.7 of the Appendix). The efficacy results over 14-week for studies 96 and 97 are summarized in Table AII.11 of the Appendix. In study 96, the responses between rofecoxib 25-mg and Naproxen are not significantly different though Naproxen is numerically better than rofecoxib 25-mg in some endpoints. In study 97, Naproxen is numerically better than rofecoxib 25-mg in all four primary endpoints and has similar responses to those of rofecoxib 50-mg. However, the differences between Naproxen and Rofecoxib are not significant.

III. Dose-Ranging Study

Study 68 is a double-blind, randomized, multicenter, parallel-group, 52-week study to assess the safety and tolerability, and to further define the clinically effective dose range of rofecoxib in patients with RA. It consists of three parts – P068P1, P068P2, and P068X. Part P068X is an extension of parts P068P1 and P068P2. The object is to assess the long term safety and tolerability of rofecoxib 25- and 50-mg. On the filing date of the NDA, part P068X is still on-going. Only the interim analyses of safety at the cut-off date of March 31, 2000 were presented in the submission. In part P068P1, three doses of rofecoxib were tested against placebo over an 8-week treatment period. They were — 25- and 50-mg. At the completion of part P068P1, patients in — ng rofecoxib and placebo were re-assigned to rofecoxib 25- and 50-mg, and naproxen. Therefore, treatment groups of rofecoxib 25-, 50-mg and naproxen were in study part P068P2. The objective of part P068P2 was primarily the safety assessment. The medical reviewer should comment any safety issues related to rofecoxib. Consequently, this review will focus on part P068P1 for which the primary objective was efficacy.

In study part P068P1, rofecoxib 25- and 50-mg arms showed significantly greater improvement over placebo and rofecoxib — ng groups in tender joint counts, patient's global, and investigator's global assessment of disease activity for 8-week treatment therapy. However, no significant superiority effect over placebo was demonstrated in swollen joint counts. Compared with placebo and rofecoxib — ng groups, rofecoxib 25- and 50-mg arms also had significantly superiority effect in patients who achieved ACR20 response, patient's global assessment of pain

and health assessment questionnaire. No significant difference between rofecoxib 25- and 50 mg groups or between rofecoxib — ng and placebo was noted. Analyses of other endpoints showed generally consistent treatment effects. The results of the primary and the secondary endpoints from study part P068P1 are summarized in Table 7.

Discussion:
In contrast to the two pivotal 12-week trials, Sponsor's dose-ranging study was an 8-week therapy trial. Despite the treatment duration difference, the treatment effects were generally maintained at a constant level across 2- to 8-week period. The change between 8- to 12-week among treatment groups in the pivotal studies showed also parallel and approximately constant level. Thus, from efficacy standpoint, it is plausible to extrapolate the efficacy results of rofecoxib 25- and 50-mg over 8-week treatment to 12-week treatment. The optimal dose of rofecoxib in treating RA patients should take into account safety concern in addition to the efficacy component.

Table 7: Efficacy Results from Study Part P068P1

Endpoint	LS mean change from baseline			
	placebo	— ng	25 mg	50 mg
Tender joint counts	-13.28	-11.86	-15.90	-14.96
Swollen joint counts	-7.09	-6.02	-6.89	-7.36
Patient global assessment (0-100 mm VAS)	-18.84	-20.50	-29.24	-28.88
Investigator global assessment (0-4 Likert scale)	-0.90	-0.89	-1.25	-1.20
ACR20 responder *	53/167 (31.74%)	53/158 (33.54%)	74/169 (43.79%)	80/161 (49.69%)
Global assessment of pain (0-100 mm VAS)	-16.39	-17.24	-25.64	-27.07
HAQ (0-3 Likert scale)	-0.17	-0.16	-0.34	-0.33
	LS mean difference between rofecoxib and placebo (95% CI)			
	25 mg vs. placebo	25 mg vs — mg	50g vs. placebo	50g vs — mg
Tender joint counts	-2.62 (-4.76, -0.48)	-4.04 (-6.21, -1.87)	-1.68 (-3.83, -0.47)	-3.10 (-5.27, -0.92)
Swollen joint counts	0.20 (-1.17, 1.57)	-0.87 (-2.26, 0.51)	-0.27 (-1.65, 1.10)	-1.34 (-2.73, 0.04)
Patient global assessment	-10.40 (-14.71, -6.08)	-8.74 (-13.13, -4.35)	-10.04 (-14.41, -5.66)	-8.38 (-12.81, -3.94)
Investigator global assessment	-0.35 (-0.52, -0.18)	-0.36 (-0.53, -0.19)	-0.29 (-0.47, -0.12)	-0.30 (-0.48, -0.13)
ACR20 responder *	12.05 (1.77, 22.34)	10.24 (-0.25, 20.74)	17.95 (7.49, 28.42)	16.15 (5.48, 26.82)
Global assessment of pain	-9.25 (-13.47, -5.03)	-8.41 (-12.71, -4.10)	-10.68 (-14.97, -6.39)	-9.83 (-14.20, -5.47)
HAQ	-0.17 (-0.26, -0.08)	-0.18 (-0.27, -0.09)	-0.16 (-0.25, -0.07)	-0.17 (-0.26, -0.08)

Note: * Proportion of subjects achieving ACR20 criteria. Larger % represents improvement
Other endpoints: decreasing values indicate improvement.
Source: Sponsor's electronic NDA submission (pages 94-112 in Study P068P1)

IV. Summary and Conclusion

Efficacy: **(Studies 96 and 97)**
The Sponsor in this submission presented results for two pivotal studies (Study 96 and Study 97) in support of the efficacy and safety claim of Vioxx tablets 25- and 50-mg for the treatment of rheumatoid arthritis (RA) indication. The tablets were administered once daily for 12 weeks in these trials. Efficacy results from these studies can be summarized as follows.

- For Study 96, Rofecoxib 25-mg is significantly more effective than placebo in the primary and secondary efficacy endpoints. However, rofecoxib 12.5-mg is ***not*** proven to be efficacious, as it failed tender and swollen joint counts based on the time-weighted average analysis, and failed all 4 primary endpoints based on the last-observed on-treatment data analysis.

- For Study 97, Rofecoxib 25- and 50-mg is significantly more effective than placebo in the primary and the secondary efficacy endpoints.

- The similarity comparisons between Rofecoxib and Naproxen ***cannot*** be drawn based on the Sponsor's submission despite the 95% confidence intervals for the difference between treatments included 0 and p-value $\geq$ 0.05. Pre-specified comparability bounds are required to draw the conclusion of similar therapeutic effect between treatments.

Dose-Escalation: **(Studies 96 and 97)**
From statistical point of view, the validity of dose escalation assessment based on the last two observations in Part I and the first two data in Part II is not clear as:
- the time interval for the last 2 observations in Part I is different from that for the 1st 2 assessment at Part II (i.e. Week 8 – Week 12 vs. Week 12 – Week 14);
- the treatment effect based on the last 2 observations from Part I was conditioned on the same treatment from baseline; while the 1st two assessments at Part II were obtained within a relatively short time based on different dose. Drug effect may take a certain length of time to show efficacy, whether positive or negative.

Dose-Ranging Study: **(Study 68)**
From efficacy standpoint, it is plausible to test the doses of rofecoxib 25- and 50-mg in the Phase III trials, as:
- dose ranging study (P068P1) showed that rofecoxib 25- and 50-mg is superior to placebo in all primary efficacy endpoints except the swollen joint counts over an 8-week treatment period.
- Despite the treatment duration difference (i.e. 12-week pivotal trials vs. 8-week dose ranging study), the drug effects were generally similar between pivotal and dose-ranging trials. The treatment effects were maintained at a constant level across 2- to 8-week period. The effect between 8- to 12-week among groups in the pivotal studies showed generally parallel and approximately constant level.

The optimal dose of rofecoxib for RA indication should take into account safety assessment in addition to the efficacy component.

Shiowjen Lee, Ph.D.
Mathematical Statistician, Biometrics III

Concur: Stan Lin, Ph.D.
Team Leader, Biometrics III

cc:
HFD-550/Div. File
HFD-550/PM/Gould
HFD-550/MO/ML Villalba
HFD-550/MO/J Schiffenbauer
HFD-700/Anello
HFD-725/Huque
HFD-725/S Lin
HFD-725/Lee

This review contains 26 pages (16 pages of text and 10 pages of Appendix).

APPENDIX

Table AII.1: Modified ITT and PP Populations

Endpoints	Treatment	Subjects randomized (ITT)	Modified ITT	PP
	STUDY 96			
Tender joint counts (total 68 joints)	Placebo	301	294 (97.7%)	266 (88.4%)
	12.5 mg	148	146 (98.6%)	139 (93.9%)
	25 mg	311	309 (99.4%)	283 (91.0%)
	naproxen	149	149 (100%)	147 (98.7%)
Swollen joint counts (total 66 joints)	Placebo	301	294 (97.7%)	266 (88.4%)
	12.5 mg	148	146 (98.6%)	139 (93.9%)
	25 mg	311	309 (99.4%)	283 (91.0%)
	naproxen	149	149 (100%)	147 (98.7%)
Patient global assessment of disease activity	Placebo	301	293 (97.3%)	264 (87.7%)
	12.5 mg	148	144 (97.3%)	137 (92.6%)
	25 mg	311	307 (98.7%)	281 (90.4%)
	naproxen	149	149 (100%)	147 (98.7%)
Investigator global assessment of disease activity	Placebo	301	294 (97.7%)	265 (88.0%)
	12.5 mg	148	145 (98.0%)	138 (93.2%)
	25 mg	311	308 (99.0%)	282 (90.7%)
	naproxen	149	149 (100%)	147 (98.7%)
	STUDY 97			
Tender joint counts (total 68 joints)	Placebo	299	294 (98.3%)	275 (92.0%)
	25 mg	315	315 (100%)	297 (94.3%)
	50 mg	297	295 (99.3%)	285 (96.0%)
	naproxen	147	146 (99.3%)	133 (90.5%)
Swollen joint counts (total 66 joints)	Placebo	299	294 (98.3%)	275 (92.0%)
	25 mg	315	315 (100%)	297 (94.3%)
	50 mg	297	295 (99.3%)	285 (96.0%)
	naproxen	147	146 (99.3%)	133 (90.5%)
Patient global assessment of disease activity	Placebo	299	294 (98.3%)	274 (91.6%)
	25 mg	315	314 (99.7%)	296 (94.0%)
	50 mg	297	295 (99.3%)	285 (96.0%)
	naproxen	147	145 (98.6%)	132 (89.8%)
Investigator global assessment of disease activity	Placebo	299	291 (97.3%)	272 (91.0%)
	25 mg	315	314 (99.7%)	296 (94.0%)
	50 mg	297	291 (98.0%)	281 (94.6%)
	naproxen	147	145 (98.6%)	132 (89.8%)

Source: Sponsor's electronic NDA submission (page 84 in study 96 and page 81 in study 97).

Table AII.2: Time-Weighted Average Change from Baseline: Per-Protocol (PP) Analyses

Primary Endpoints	Pairwise Treatment Difference in LSMean (95% Confidence Interval) p-value for the comparison of two treatments		
STUDY 96	**12.5 mg vs. placebo**	**25 mg vs. placebo**	**Naproxen vs. placebo**
Tender joint counts	-1.64 (-3.55, 0.28) 0.093	-2.52 (-4.08, -0.96) 0.002	-3.20 (-5.08, -1.32) < 0.001
Swollen joint counts	-0.93 (-2.17, 0.31) 0.142	-0.91 (-1.92, 0.10) 0.077	-1.54 (-2.76, -0.32) 0.013
Patient's global	-6.06 (-10.10, -2.01) 0.003	-7.57 (-10.86, -4.27) < 0.001	-11.53 (-15.48, -7.58) < 0.001
Investigator's global	-0.17 (-0.33, -0.00) 0.048	-0.26 (-0.39, -0.12) < 0.001	-0.28 (-0.44, -0.12) < 0.001
STUDY 97	**25 mg vs. placebo**	**50 mg vs. placebo**	**Naproxen vs. placebo**
Tender joint counts	-2.57 (-4.19, -0.94) 0.002	-4.22 (-5.86, -2.57) < 0.001	-3.46 (-5.51, -1.41) < 0.001
Swollen joint counts	-1.37 (-2.39, -0.36) 0.008	-1.31 (-2.34, -0.28) 0.012	-0.77 (-2.05, 0.51) 0.238
Patient's global	-6.57 (-9.74, -3.39) < 0.001	-9.93 (-13.14, -6.72) < 0.001	-10.48 (-14.50, -6.46) < 0.001
Investigator's global	-0.30 (-0.42, -0.17) < 0.001	-0.36 (-0.49, -0.24) < 0.001	-0.38 (-0.54, -0.23) < 0.001

Note: Decreasing values indicate improvement.
Source: Sponsor's electronic NDA submission (pages 1908-1911 in study 96 and pages 2276-2279 in stud 97).

Table AII.3: Change from Baseline: Last Observed On-Treatment Data Analyses

Primary Endpoints	Pairwise Treatment Difference in LSMean (95% Confidence Interval) p-value for the comparison of two treatments		
STUDY 96	**12.5 mg vs. placebo**	**25 mg vs. placebo**	**Naproxen vs. placebo**
Tender joint counts	-1.11 (-3.43, 1.20) 0.346	-2.59 (-4.45, -0.73) 0.006	-2.80 (-5.10, -0.51) 0.017
Swollen joint counts	-1.04 (-2.55, 0.47) 0.175	-1.17 (-2.38, 0.05) 0.059	-1.57 (-3.07, -0.07) 0.040
Patient's global	-3.92 (-8.84, 1.00) 0.118	-6.62 (-10.58, -2.66) 0.001	-5.56 (-10.43, -0.69) 0.025
Investigator's global	-0.11 (-0.32, 0.09) 0.284	-0.28 (-0.44, -0.11) 0.001	-0.13 (-0.34, 0.07) 0.209
STUDY 97	**25 mg vs. placebo**	**50 mg vs. placebo**	**Naproxen vs. placebo**
Tender joint counts	-3.23 (-5.15, -1.31) < 0.001	-3.60 (-5.55, -1.65) < 0.001	-3.97 (-6.36, -1.57) 0.001
Swollen joint counts	-1.35 (-2.50, -0.19) 0.022	-1.14 (-2.32, 0.03) 0.056	-1.00 (-2.44, 0.44) 0.174
Patient's global	-7.46 (-11.31, -3.62) < 0.001	-9.41 (-13.31, -5.51) < 0.001	-10.04 (-14.84, -5.23) < 0.001
Investigator's global	-0.37 (-0.52, -0.22) < 0.001	-0.36 (-0.52, -0.21) < 0.001	-0.47 (-0.66, -0.28) < 0.001

Note: Decreasing values indicate improvement.
Source: Sponsor's electronic NDA submission (pages 1926-1929 in study 96 and pages 2294-2297 in study 97)

Table AII.4: Difference in Change from Baseline between Time-Weighted Average and the Last-Observed On-Treatment Data

Treatment group	Tender Joint Counts		Swollen Joint Counts	
	Time-weighted average (S.E)	Last-observed on-treatment data (S.E)	Time-weighted average (S.E)	Last-observed on-treatment data (S.E)
	STUDY 96			
Placebo	-11.81 (0.65)	-12.31 (0.77)	-5.93 (0.43)	-6.01 (0.51)
12.5 mg rofecoxib	-12.77 (0.97)	-12.90 (1.20)	-6.50 (0.63)	-6.72 (0.76)
25 mg rofecoxib	-14.32 (0.61)	-14.70 (0.70)	-6.98 (0.40)	-7.00 (0.46)
Naproxen	-14.80 (0.88)	-15.07 (1.11)	-7.83 (0.58)	-7.77 (0.67)
	STUDY 97			
Placebo	-10.58 (0.73)	-11.73 (0.84)	-5.81 (0.45)	-6.32 (0.50)
25 mg rofecoxib	-13.27 (0.62)	-14.66 (0.74)	-6.78 (0.41)	-7.37 (0.47)
50 mg rofecoxib	-14.41 (0.62)	-15.13 (0.76)	-6.83 (0.40)	-7.31 (0.48)
Naproxen	-14.54 (0.82)	-15.71 (0.94)	-6.84 (0.56)	-7.36 (0.62)
Treatment group	**Patient's global assessment of disease activity**		**Investigator's global assessment of disease activity**	
	Time-weighted average (S.E)	Last-observed on-treatment data (S.E)	Time-weighted average (S.E)	Last-observed on-treatment data (S.E)
	STUDY 96			
Placebo	-21.55 (1.22)	-22.76 (1.45)	-0.85 (0.056)	-0.88 (0.068)
12.5 mg rofecoxib	-26.95 (1.81)	-26.81 (2.26)	-1.02 (0.077)	-0.99 (0.095)
25 mg rofecoxib	-27.54 (1.20)	-28.18 (1.45)	-1.19 (0.054)	-1.18 (0.064)
Naproxen	-31.75 (1.79)	-28.16 (2.17)	-1.09 (0.071)	-0.99 (0.088)
	STUDY 97			
Placebo	-22.55 (1.28)	-23.69 (1.49)	-0.68 (0.057)	-0.70 (0.067)
25 mg rofecoxib	-28.65 (1.16)	-30.30 (1.43)	-0.98 (0.051)	-1.04 (0.061)
50 mg rofecoxib	-32.44 (1.20)	-33.21 (1.46)	-1.00 (0.050)	-1.01 (0.060)
Naproxen	-31.12 (1.63)	-32.72 (2.03)	-1.11 (0.072)	-1.17 (0.081)

Note: Decreasing values indicate improvement.
S.E. = standard error of the change
Source: Sponsor's electronic NDA (pages 114-120, 1926-1929 in study 96 and pages 105-111, 2294-2297 in study 97).

Table AII.5: Change from Baseline to Week 12 based on ITT Population – Imputation by the Average Response

Primary Endpoints	Pairwise Treatment Difference in LSMean (95% Confidence Interval) p-value for the comparison of two treatments		
STUDY 96	**12.5 mg vs. placebo**	**25 mg vs. placebo**	**Naproxen vs. placebo**
Tender joint counts	-1.12 (-3.40, 1.15) 0.334	-2.58 (-4.41, -0.75) 0.006	-2.63 (-4.90, -0.36) 0.023
Swollen joint counts	-0.93 (-2.41, 0.56) 0.223	-1.05 (-2.25, 0.15) 0.086	-1.41 (-2.89, 0.074) 0.063
Patient's global	-4.00 (-8.86, 0.86) 0.107	-6.60 (-10.50, -2.69) 0.001	-5.50 (-10.33, -0.68) 0.026
Investigator's global	-0.11 (-0.32, 0.09) 0.272	-0.28 (-0.44, -0.12) < 0.001	-0.12 (-0.32, 0.085) 0.256
STUDY 97	**25 mg vs. placebo**	**50 mg vs. placebo**	**Naproxen vs. placebo**
Tender joint counts	-3.33 (-5.23, -1.42) < 0.001	-3.59 (-5.53, -1.66) < 0.001	-4.00 (-6.38, -1.63) 0.001
Swollen joint counts	-1.39 (-2.53, -0.24) 0.018	-1.16 (-2.33, 0+) 0.050	-1.07 (-2.50, 0.36) 0.145
Patient's global	-7.59 (-11.40, -3.78) < 0.001	-9.54 (-13.40, -5.68) < 0.001	-10.11 (-14.87, -5.35) < 0.001
Investigator's global	-0.37 (-0.52, -0.22) < 0.001	-0.36 (-0.51, -0.21) < 0.001	-0.47 (-0.66, -0.28) < 0.001

Note: Decreasing values indicate improvement.
Source: Reviewer's analysis based on the data sets Sponsor submitted.

APPEARS THIS WAY
ON ORIGINAL

Table AII.6: Change from Baseline: Comparisons of Rofecoxib vs. Naproxen in the Primary Endpoint (MITT Analyses)

Primary Endpoint	Pairwise Treatment Difference in LSMean (95% Confidence Interval) p-value for the comparison of two treatments		
	STUDY 96	STUDY 97	
	25 mg vs. naproxen	25 mg vs. naproxen	50 mg vs. naproxen
Tender joint counts	0.36 (-1.47, 2.20) 0.699	0.99 (-0.98, 2.97) 0.324	-0.05 (-2.05, 1.94) 0.960
Swollen joint counts	0.51 (-0.68, 1.70) 0.400	-0.25 (-1.47, 0.97) 0.690	-0.16 (-1.40, 1.07) 0.794
Patient's global	3.24 (-0.70, 7.17) 0.107	2.63 (-1.21, 6.47) 0.179	-0.19 (-4.07, 3.70) 0.924
Investigator's global	-0.05 (-0.21, 0.11) 0.543	0.11 (-0.04, 0.26) 0.166	0.07 (-0.08, 0.22) 0.379

Note: Decreasing values indicate improvement.
Source: Sponsor's electronic NDA submission (pages 114-120 in study 96 and pages 105-111 in study 97).

Table AII.7: Change from Baseline: Comparisons of Rofecoxib vs. Naproxen in the Secondary Endpoints (MITT Analyses)

Secondary Endpoint	Pairwise Treatment Difference in LSMean (95% Confidence Interval) p-value for the comparison of two treatments		
	STUDY 96	STUDY 97	
	25 mg vs. naproxen	25 mg vs. naproxen	50 mg vs. naproxen
ACR20 responder and completer *	-1.57% (-11.32%, 8.18%) 0.771	-2.21% (-12.02%, 7.59%) 0.661	0.49% (-9.42%, 10.39%) 0.886
Patient's global assess. of pain #	0.79 (-3.11, 4.70) 0.690	3.71 (-0.25, 7.66) 0.066	1.51 (-2.48, 5.50) 0.458
HAQ change #	0.02 (-0.07, 0.10) 0.699	-0.01 (-0.10, 0.07) 0.792	-0.02 (-0.11, 0.07) 0.643

Note: * Proportion of subjects achieving ACR20 criteria. Larger % represents improvement
Decreasing values indicate improvement.
Source: Sponsor's electronic NDA submission (pages 122-126 in study 96 and pages 113-117 in study 97).

Table AII.8: Average Change from Baseline (# of subjects) for Primary Endpoints by Dropout Reasons: Studies 96 and 97

			Discontinuation		
Treatment groups	Modified ITT analyses	Completers	Due to lack of efficacy	Due to AE	Due to other reasons
	STUDY 96				
Tender joint counts					
Placebo	-11.52 (294)	-13.55 (200)	-7.27 (77)	-10.07 (10)	-8.41 (7)
Rofecixob 12.5 mg	-13.02 (146)	-15.11 (110)	-6.01 (26)	-10.53 (5)	-2.89 (5)
Rofecoxib 25 mg	-14.25 (309)	-15.34 (245)	-5.34 (33)	-15.91 (17)	-14.36 (14)
Naproxen	-14.61 (149)	-15.86 (118)	-5.60 (18)	-12.25 (7)	-18.08 (6)
Swollen joint counts					
Placebo	-5.82 (294)	-6.99 (200)	-3.27 (77)	-7.69 (10)	-8.54 (7)
Rofecixob 12.5 mg	-6.73 (146)	-7.73 (110)	-3.02 (26)	-4.61 (5)	-5.14 (5)
Rofecoxib 25 mg	-7.04 (309)	-7.39 (245)	-2.72 (33)	-7.73 (17)	-6.79 (14)
Naproxen	-7.55 (149)	-7.92 (118)	-5.56 (18)	-9.02 (7)	-5.50 (6)
Patient's global					
Placebo	-20.61 (293)	-24.42 (200)	-9.96 (76)	-27.18 (10)	-24.61 (7)
Rofecixob 12.5 mg	-25.94 (144)	-31.55 (108)	-5.32 (26)	-20.76 (5)	-12.34 (5)
Rofecoxib 25 mg	-27.79 (307)	-31.85 (244)	-4.96 (32)	-18.73 (17)	-21.82 (14)
Naproxen	-31.02 (149)	-34.00 (118)	-15.22 (18)	-25.94 (7)	-33.37 (6)
Investigator's global					
Placebo	-0.84 (294)	-1.11 (200)	-0.20 (76)	-0.79 (10)	-0.67 (7)
Rofecixob 12.5 mg	-1.01 (145)	-1.22 (109)	-0.23 (26)	-0.93 (5)	-0.54 (5)
Rofecoxib 25 mg	-1.15 (308)	-1.28 (245)	-0.07 (32)	-1.23 (17)	-1.11 (14)
Naproxen	-1.10 (149)	-1.25 (118)	-0.24 (18)	-1.02 (7)	-0.94 (6)
	STUDY 97				
Tender joint counts					
Placebo	-10.42 (294)	-12.39 (237)	0.60 (39)	-4.68 (12)	-15.88 (6)
Rofecixob 25 mg	-13.38 (315)	-14.10 (281)	-3.00 (16)	-9.57 (12)	-12.52 (6)
Rofecoxib 50 mg	-14.42 (295)	-15.21 (250)	-6.11 (13)	-11.69 (25)	-14.25 (7)
Naproxen	-14.37 (146)	-15.18 (126)	-1.90 (5)	-14.36 (11)	-5.68 (4)
Swollen joint counts					
Placebo	-5.68 (294)	-6.95 (237)	0.20 (39)	-2.14 (12)	-5.95 (6)
Rofecixob 25 mg	-6.93 (315)	-7.28 (281)	-1.87 (16)	-4.56 (12)	-7.06 (6)
Rofecoxib 50 mg	-6.84 (295)	-7.02 (250)	-5.91 (13)	-5.24 (25)	-8.69 (7)
Naproxen	-6.68 (146)	-6.72 (126)	-0.52 (5)	-8.02 (11)	-6.55 (4)
Patient's global					
Placebo	-22.14 (294)	-26.43 (237)	-1.71 (38)	-13.95 (13)	-3.66 (6)
Rofecixob 25 mg	-29.09 (314)	-31.04 (280)	-3.52 (16)	-22.15 (12)	-20.81 (6)
Rofecoxib 50 mg	-31.91 (295)	-33.70 (250)	-8.76 (13)	-22.50 (25)	-41.11 (7)
Naproxen	-31.72 (145)	-34.68 (125)	-4.53 (5)	-14.42 (11)	-18.27 (4)
Investigator's global					
Placebo	-0.66 (291)	-0.83 (233)	0.15 (39)	-0.40 (13)	-0.52 (6)
Rofecixob 25 mg	-0.99 (314)	-1.05 (280)	0.02 (16)	-0.77 (12)	-0.79 (6)
Rofecoxib 50 mg	-1.03 (291)	-1.10 (246)	0.12 (13)	-0.84 (25)	-1.19 (7)
Naproxen	-1.10 (145)	-1.14 (125)	-0.45 (5)	-0.94 (11)	-0.57 (4)

Note: Decreasing values indicate improvement.
Source: Sponsor's electronic NDA (pages 1982-1983 in study 96 and pages 2344-2345 in study 97).

Table AII.9: Pattern of Missing Values over Assessment Week

Study 96: number (%)				Study 97: number (%)			
Treatment group	**Week 4**	**Week 8**	**Week 12**	**Treatment group**	**Week 4**	**Week 8**	**Week 12**
Tender joint				**Tender joint**			
Placebo	39/301 (13.0)	74/301 (24.6)	91/301 (30.2)	Placebo	24/299 (8.0)	40/299 (13.4)	52/299 (17.4)
12.5 mg	11/148 (7.4)	27/148 (18.2)	32/148 (21.6)	25 mg	10/315 (3.2)	16/315 (5.1)	31/315 (9.8)
25 mg	22/311 (7.1)	38/311 (12.2)	56/311 (18.0)	50 mg	9/297 (3.0)	21/297 (7.1)	34/297 (11.4)
Naproxen	4/149 (2.7)	12/149 (8.1)	27/149 (18.1)	Naproxen	8/147 (5.4)	15/147 (10.2)	21/147 (14.3)
Swollen joint				**Swollen joint**			
Placebo	39/301 (13.0)	74/301 (24.6)	91/301 (30.2)	Placebo	24/299 (8.0)	40/299 (13.4)	52/299 (17.4)
12.5 mg	11/148 (7.4)	27/148 (18.2)	32/148 (21.6)	25 mg	10/315 (3.2)	16/315 (5.1)	31/315 (9.8)
25 mg	22/311 (7.1)	38/311 (12.2)	56/311 (18.0)	50 mg	9/297 (3.0)	21/297 (7.1)	34/297 (11.4)
Naproxen	4/149 (2.7)	12/149 (8.1)	27/149 (18.1)	Naproxen	8/147 (5.4)	15/147 (10.2)	21/147 (14.3)
Patient's global				**Patient's global**			
Placebo	40/301 (13.3)	74/301 (24.6)	88/301 (29.2)	Placebo	26/299 (8.7)	41/299 (13.7)	52/299 (17.4)
12.5 mg	11/148 (7.4)	27/148 (18.2)	32/148 (21.6)	25 mg	10/315 (3.2)	16/315 (5.1)	31/315 (9.8)
25 mg	21/311 (6.8)	37/311 (11.9)	55/311 (17.7)	50 mg	7/297 (2.4)	21/297 (7.1)	36/297 (12.1)
Naproxen	4/149 (2.7)	12/149 (8.1)	29/149 (19.5)	Naproxen	8/147 (5.4)	14/147 (9.5)	21/147 (14.3)
Invest. global				**Invest. global**			
Placebo	41/301 (13.6)	74/301 (24.6)	91/301 (30.2)	Placebo	25/299 (8.4)	40/299 (13.4)	53/299 (17.7)
12.5 mg	11/148 (7.4)	27/148 (18.2)	32/148 (21.6)	25 mg	10/315 (3.2)	16/315 (5.1)	31/315 (9.8)
25 mg	21/311 (6.8)	37/311 (11.9)	56/311 (18.0)	50 mg	9/297 (3.0)	21/297 (7.1)	34/297 (11.4)
Naproxen	4/149 (2.7)	13/149 (8.7)	28/149 (18.8)	Naproxen	8/147 (5.4)	13/147 (8.8)	21/147 (14.3)

Source: Results were summarized from Sponsor's electronic NDA submission (page 1985 in Study 96 and page 2347 in Study 97).

Table AII.10(a): Change from Part I to II for Primary Endpoints: Study 96

Treatment (Base/Extension)	N	Part I Mean[†]	Part II Mean[‡]	Mean Change[§]	SD[‖] of Change	LS Mean[¶] Change	95% CI[#] for LS Mean[¶] Change
Tender Joint Count (Total 68)							
Placebo/25 mg	94	13.32	10.07	-3.24	6.19	-3.43	(-4.86, -2.00)
Placebo/Naproxen	104	15.19	12.13	-3.06	8.39	-2.83	(-4.19, -1.46)
25 mg/25 mg	130	11.99	11.41	-0.58	6.44	-0.75	(-2.02, 0.53)
25 mg/50 mg	114	13.22	12.12	-1.10	8.97	-0.96	(-2.32, 0.39)
12.5 mg/25 mg	110	12.51	11.28	-1.23	6.06	NA[††]	(-2.37, -0.10)[††]
Naproxen/Naproxen	118	12.66	10.69	-1.96	5.15	NA[††]	(-2.89, -1.03)[††]
Swollen Joint Count (Total 66)							
Placebo/25 mg	94	9.91	8.45	-1.46	4.25	-1.48	(-2.33, -0.64)
Placebo/Naproxen	104	10.40	8.91	-1.50	4.54	-1.42	(-2.23, -0.61)
25 mg/25 mg	130	9.60	9.81	0.20	4.27	0.11	(-0.60, 0.81)
25 mg/50 mg	114	11.07	10.44	-0.64	3.80	-0.59	(-1.34, 0.16)
12.5 mg/25 mg	110	9.77	9.53	-0.24	3.81	NA[††]	(-0.95, 0.47)[††]
Naproxen/Naproxen	118	10.04	9.29	-0.75	3.91	NA[††]	(-1.45, -0.04)[††]
Patient Global Assessment of Disease Activity (0 to 100 Visual Analog Scale)							
Placebo/25 mg	94	44.90	35.41	-9.49	15.76	-9.53	(-12.68, -6.38)
Placebo/Naproxen	104	45.44	35.86	-9.58	16.86	-9.47	(-12.48, -6.45)
25 mg/25 mg	130	38.28	37.53	-0.75	16.08	-0.55	(-3.11, 2.01)
25 mg/50 mg	114	37.49	34.10	-3.39	13.86	-3.38	(-6.09, -0.67)
12.5 mg/25 mg	110	38.18	36.05	-2.14	15.14	NA[††]	(-4.97, 0.69)[††]
Naproxen/Naproxen	118	38.04	37.26	-0.78	14.95	NA[††]	(-3.48, 1.92)[††]
Investigator Global Assessment of Disease Activity (0 to 4 Likert Scale)							
Placebo/25 mg	94	1.40	1.02	-0.38	0.72	-0.36	(-0.50, -0.22)
Placebo/Naproxen	103	1.35	1.01	-0.34	0.76	-0.34	(-0.47, -0.20)
25 mg/25 mg	130	1.25	1.32	0.08	0.75	0.05	(-0.07, 0.18)
25 mg/50 mg	114	1.30	1.24	-0.07	0.75	-0.07	(-0.20, 0.06)
12.5 mg/25 mg	110	1.28	1.23	-0.05	0.62	NA[††]	(-0.17, 0.07)[††]
Naproxen/Naproxen	118	1.29	1.21	-0.08	0.67	NA[††]	(-0.20, 0.04)[††]

† The average of last 2 assessments in Part I.
‡ The average of first 2 assessments in Part II.
§ Between Part I and Part II.
‖ Standard deviation.
¶ Least-square mean.
# Confidence interval.
†† There is no least-squares mean since this treatment sequence is not analyzed by the Analysis of Covariance model. The 95% CI is for raw mean change.

Source: Sponsor's electronic NDA submission (page 143 in Study 96).

APPEARS THIS WAY
ON ORIGINAL

Table AII.10(b): Change from Part I to II for Primary Endpoints: Study 97

Treatment (Base/Extension)	N	Part I Mean†	Part II Mean‡	Mean Change§	SD of Change	LS Mean‖ Change	95% CI¶ for LS Mean Change
Tender Joint Count (Total 68 Joints)							
Placebo/25 mg	114	16.51	12.13	-4.38	6.39	-4.59	(-5.75, -3.43)
Placebo/Naproxen	123	18.21	14.64	-3.57	7.53	-3.39	(-4.51, -2.28)
25 mg/25 mg	139	15.93	14.96	-0.97	7.23	-0.92	(-2.04, 0.19)
25 mg/50 mg	141	14.80	13.58	-1.22	6.33	-1.29	(-2.40, -0.18)
50 mg/50 mg	249	14.21	12.77	-1.44	6.34	NA#	(-2.23, -0.66)#
Naproxen/Naproxen	124	13.98	12.52	-1.46	6.29	NA#	(-2.57, -0.35)#
Swollen Joint Count (Total 66 Joints)							
Placebo/25 mg	114	8.83	7.85	-0.98	3.64	-1.01	(-1.65, -0.36)
Placebo/Naproxen	123	9.25	8.30	-0.96	3.65	-0.93	(-1.55, -0.30)
25 mg/25 mg	139	8.68	7.87	-0.81	3.75	-0.81	(-1.41, -0.21)
25 mg/50 mg	141	8.57	8.49	-0.09	3.59	-0.09	(-0.69, 0.51)
50 mg/50 mg	249	9.01	8.38	-0.63	4.24	NA#	(-1.16, -0.11)#
Naproxen/Naproxen	124	9.82	8.92	-0.90	4.40	NA#	(-1.68, -0.13)#
Patient's Global Assessment of Disease Activity (0 to 100 Visual Analog Scale)							
Placebo/25 mg	114	44.04	35.90	-8.14	15.70	-8.67	(-11.35, -5.99)
Placebo/Naproxen	123	47.43	38.01	-9.42	16.40	-8.84	(-11.43, -6.24)
25 mg/25 mg	139	42.13	39.30	-2.83	15.18	-2.72	(-5.02, -0.43)
25 mg/50 mg	140	41.08	38.81	-2.28	13.90	-2.41	(-4.70, -0.13)
50 mg/50 mg	250	40.42	40.14	-0.28	14.29	NA#	(-2.05, 1.50)#
Naproxen/Naproxen	124	37.69	36.89	-0.79	14.71	NA#	(-3.38, 1.79)#
Investigator's Global Assessment of Disease Activity (0 to 4 Likert Scale)							
Placebo/25 mg	114	1.58	1.18	-0.40	0.73	-0.42	(-0.54, -0.30)
Placebo/Naproxen	123	1.65	1.28	-0.37	0.73	-0.35	(-0.47, -0.24)
25 mg/25 mg	139	1.45	1.27	-0.18	0.56	-0.17	(-0.26, -0.07)
25 mg/50 mg	141	1.34	1.27	-0.08	0.67	-0.09	(-0.19, 0.00)
50 mg/50 mg	249	1.38	1.29	-0.09	0.69	NA#	(-0.17, -0.00)#
Naproxen/Naproxen	125	1.33	1.28	-0.05	0.64	NA#	(-0.16, 0.06)#

† The average of last 2 assessments in Part I.
‡ The average of first 2 assessments in Part II.
§ Between Part I and Part II.
‖ Least-square mean.
¶ CI = Confidence interval.
# There is no Least-square mean since this treatment sequence is not analyzed by the ANCOVA model. The 95% CI is for raw mean change.

Source: Sponsor's electronic NDA submission (page 134 in Study 97).

Table AII.11: Efficacy Results over 14-Week Treatment: Studies 96 and 97

	STUDY 96			STUDY 97			
	LS mean Change			LS mean Change			
Endpoints	25 mg	Naproxen	95% CI for Diff.	25 mg	50 mg	Naproxen	95% CI for Diff.
Primary Endpoints							
Tender joint counts	-14.54	-14.83	(-1.61, 2.19)	-12.89	-14.65	-14.66	(-0.34, 3.88) (-1.84, 1.86)
Swollen joint counts	-7.39	-7.75	(-0.91, 1.63)	-6.79	-6.92	-6.81	(-1.26, 1.28) (-1.22, 1.00)
Patient's global	-28.06	-31.22	(-1.28, 7.59)	-29.66	-32.11	-32.02	(-1.94, 6.67) (-3.87, 3.69)
Invest. global	-1.15	-1.10	(-0.21, 0.12)	-0.98	-1.03	-1.10	(-0.04, 0.28) (-0.07, 0.21)
Secondary Endpoints							
ACR20 response	86/166 (51.81%)	80/149 (53.69%)	(-14.0%, 7.8%)	82/154 (53.25%)	157/295 (53.22%)	76/146 (52.05%)	(-10.1%, 12.5%) (-8.7%, 11.1%)
Pain assessment	-24.18	-26.13	(-2.54, 6.44)	-26.76	-28.81	-30.36	(-0.85, 8.06) (-2.35, 5.45)
HAQ evaluation	-0.39	-0.41	(-0.07, 0.12)	-0.39	-0.41	-0.38	(-0.11, 0.09) (-0.11, 0.06)

Note: Decreasing values indicate improvement for all endpoints except "ACR20 response".
"95% CI for Diff." represents 95% confidence interval for difference between Rofecoxib and Naproxe
(i.e. rofecoxib – naproxen)
Source: Sponsor's electronic NDA submission (pages 1951-1957 in Study 96 and pages 2313-2319 in Study 97).

APPEARS THIS WAY
ON ORIGINAL

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

/s/

Showjien Lee
8/29/01 03:50:12 PM
BIOMETRICS

Stan Lin
8/29/01 04:36:33 PM
UNKNOWN

APPEARS THIS WAY
ON ORIGINAL

Statistical Review

NDA#	**21-042 s-007**
Name of Drug:	Rofecoxib
Applicant:	Merck Research Laboratories
Indication:	Gastrointestinal (GI) Safety Label Change
Documents Reviewed:	Statistical section submitted in June 29, 2000
Medical Reviewer:	Maria Lourdes Villalba, MD
Statistical Reviewer:	Qian Li, Sc.D.
Period of Review:	June 2000 - March 2001

I. Introduction:

Rofecoxib was originally submitted as an NDA in November 1998 and approved by the Agency in May 1999 for the relief of sign and symptom of osteo-arthritis (OA) and for the management of acute pain and dysmenorrhea. The current approved maximum dose was 25 mg daily for OA and 50 mg daily for acute pain. The purpose of this supplemental NDA submission was to provide evidence for label revision in gastrointestinal (GI) warning section for rofecoxib. A GI outcome study (Protocols 088/089) named VIGOR (Vioxx Gastrointestinal Outcomes Research study) was conducted to support the GI safety claim. The VIGOR trial was a double-blind, randomized, stratified, parallel-group study to compare the occurrence of PUBs (gastroduodenal perforations, gastroduodenal ulcers, or upper gastrointestinal bleeds) between rofecoxib 50 mg daily or naproxen 1000mg per day during chronic treatment for patients with rheumatoid arthritis (RA). This study was divided into two protocols, Protocols 088 and 089, a U.S cohort and an international cohort.

During the VIGOR trial, many serious cardiovascular (CV) events were observed. To address the issue of serious cardiovascular events, the sponsor organized a special section in the VIGOR study report to discuss analyses on thrombotic cardiovascular serious adverse events. In addition, clinical trial reports from Protocols 085 and 090, designed to compare the safety and efficacy of rofecoxib 12.5 mg daily vs. nebumetone 1000 mg per day in patients with OA, as well as a 6-week geriatric study (Protocol 58), were submitted to support concomitant use of low-dose aspirin with rofecoxib for cardio-protection.

[ ]

In this statistical review, analyses on GI safety profile and cardiovascular events between rofecoxib 50 mg daily and naproxen 1000 mg per day treatment groups were reviewed based on the results of the VIGOR study. The meta-analysis was also reviewed and the conclusion of the review was presented to the advisory committee in response to the sponsor's presentation on the meta-analysis. The review of the meta-analysis was attached to the end of the VIGOR trial review. This statistical review did not cover these

additional studies that allowed concomitant use of aspirin for cardiovascular evaluations, as they were short term and low dose studies, and not powered to evaluate the GI and cardiovascular safety of the combination use of rofecoxib and aspirin.

II. Study Design and Statistical Methodology:

The primary object of the VIGOR study was to determine the relative risk of confirmed PUBs in patients with RA taking rofecoxib 50 mg daily compared to patients taking naproxen 1000 mg/day. Patients of age 40 or older, with rheumatoid arthritis which required treatment with nonsteroidal anti-inflammatory drugs (NSAIDs) therapy for at least 1 year were recruited to the studies. Patients who met entry criteria of the study were randomized to rofecoxib, 50 mg daily, or naproxen 500 mg twice daily. Patient allocation was stratified with a prior history of peptic ulcer, upper GI bleeding or perforation versus those who had no prior history. Clinic visits were scheduled at screening, randomization, weeks 6, 17, 35, 52, and every 4 months thereafter until the termination of the study. At the termination, patients were called in for an end-of-study visit and patients were asked to remain off NSAIDs for 14 days. The study was planned to stop when at least 120 confirmed PUBs and a minimum of 40 confirmed complicated PUBs were observed in the study, and minimum duration of treatment was 6 month for the last randomized patient, which ever came last.

The original protocol was designed to stop the trial when 95 confirmed PUBs were observed. In response to the FDA's emphasis on confirmed complicated PUBs, the VIGOR protocol was amended to observe a minimum of 40 confirmed complicated cases as an additional condition before stopping the trial. During the trial, it was found that only 25-30% of the confirmed cases were complicated. In order to achieve this requirement to observe a minimum of 40 confirmed complicated cases, it was necessary to increase the total confirmed PUBs from 95 to 120. Since the sample size change was not due to the interim result of primary end point, penalty on alpha level was not necessary.

Reviewer's comment on study design:

Rofecoxib has not been approved for rheumatoid arthritis patients. Since RA and OA are two different disease populations, the efficacy effect of rofecoxib is expected to be different for the two patient populations. It was not clear if the two patient populations would share the same GI safety profile.

The dosage of rofecoxib used in RA patients in this VIGOR trial was twice of the maximum approved chronic dose for OA patients. It was unavailable at present what would be the effective dose for RA if rofecoxib would be approved for this indication. Therefore, it is too early to conclude what was observed in this VIGOR study represented the worst scenario of rofecoxib in actual use.

Different NSAIDs had different GI safety profile. Therefore using naproxen alone as a NSAID representative may not be appropriate for a claim against a class of drug.

However, if there was evidence to show that naproxen was the mildest in GI toxicity in the whole NSAID class, it would be appropriate for rofecoxib to gain the claim against the class of NSAIDs. However, naproxen has not been shown that it was the mildest among the NSAIDs in GI toxicity.

1. Analysis populations:

Two analysis populations were defined in this study. They were:

All-patient-randomized (APR): the population included all the randomized patients.

Per-protocol population excluded patients who were identified as substantive protocol violation. Substantive protocol violators were defined based on a set of pre-specified criteria.

2. PUBs evaluation:

At each study visit, patients were asked questions concerning the occurrence of PUBs. Suspicious of possible study end point prompted the retrieval of additional information and source documents. Between visits including phone visit, the patients were encouraged to call the study site if a PUB, GI work-up, or other serious adverse experience were occurred. The patients were asked to provide permission to obtain medical records and copies of endoscopy or radiographic reports. An initial end point report form was completed and submitted to an External Coordinating Center. Classification of PUBs was adjudicated by an independent End Point Classification Committee (See medical officer's review for classification).

Primary endpoints:

The primary study end point was defined to be confirmed PUBs by the sponsor. However, the agency placed more emphasis on confirmed and complicated PUBs. The sponsor used this endpoint as a secondary endpoint.

Secondary GI variables specified by the sponsor:

(1) Confirmed and complicated PUBs.
(2) Confirmed and unconfirmed PUBs.
(3) Confirmed and unconfirmed complicated PUBs.
(4) GI related adverse experience.
(5) Any GI bleeding.

3. Other safety evaluations:

Pre-specified safety analyses:

Other than routine safety analyses on adverse events, vital sign and laboratory parameters were tabulated. In addition to the routine safety analyses, the protocol and data analysis plan also specified the following safety parameters for detailed statistical analyses.

(1) Serious clinical adverse experiences (overall)
(2) Drug-related (possibly, probably, definitely) clinical adverse experiences (overall)
(3) Clinical adverse experiences leading to study discontinuation (overall)
(4) Discontinuations due to digestive adverse experiences including abdominal pain
(5) Discontinuations due to edema-related adverse experiences
(6) Discontinuations due to hypertension-related adverse experiences
(7) Discontinuations due to renal-related adverse experiences (clinical and/or laboratory adverse experiences)
(8) Discontinuations due to hepatic-related adverse experiences (clinical and/or laboratory adverse experiences)
(9) Congestive heart failure adverse experiences
(10) Serious laboratory adverse experiences (overall)
(11) Drug-related (possibly, probably, definitely) laboratory adverse experiences (overall)
(12) Laboratory adverse experiences leading to study discontinuation (overall).

Serious cardiovascular adverse events:

In this study, investigator identified cardiovascular events were adjudicated according to Cardiovascular Adjudication Standard Operation Procedures. The primary analysis of the events focused on confirmed thrombotic cardiovascular serious adverse events.

4. Efficacy evaluation:

Rofecoxib has not been approved for the indication of rheumatoid arthritis. Efficacy evaluation in this VIGOR study was not sufficient, as the study design was not oriented to the efficacy evaluation. Nevertheless, the following efficacy endpoints were assessed in this trial:

(1) Patient global assessment of disease activity: a patient global assessment of disease activity on a 5-point Likert scale was administrated at Visit 1.0, 2.0, 3.0, 4.0, 6.0, 8.0, 9.0, end-of-study, and discontinuation. The scale is 0=very well, 1=well, 2=fair, 3=poor, 4=very poor.
(2) Investigator global assessment of disease activity: using the same 5-point likert scale as patient global assessment of disease activity.
(3) Discontinuation due to lack of efficacy.
(4) Modified health assessment questionnaire on dressing and grooming, arising, eating, walking, hygiene, reach, grip and activities and recorded at visit 2.0, 3.0, and end-of-study.

5. Statistical Analyses:

The primary GI endpoint, pre-specified safety analysis and serious cardiovascular adverse events were analyzed based on the APR population.

For the primary end point of confirmed PUBs, Cox proportional hazard model was used to compare the relative risk between the two treatment groups. Covariates included in this model were treatment group indicator and stratum of prior history of PUBs.

For other time-to-event end points including various types of PUBs, discontinuations due to lack of efficacy, the pre-specified safety analyses variables, and cardiovascular serious adverse events, similar survival analyses were used to evaluate time to the first event during the study period. Patient's and investigator's global assessments, as well as modified HAQ (US only) were analyzed as the average change from baseline over the treatment period using an analysis of co-variance (ANCOVA) model with factors of treatment, study center, stratum, and baseline value as covariates.

One interim analysis was planned when 60 confirmed PUBs was observed, which was half information time of the total 120 confirmed PUBs. A group sequential stopping rule was used to control the overall type I error rate at 0.05. The corresponding two sided stopping boundaries were 2.753 (α_1=0.0059) and 1.982 (α_2=0.0475) based on an O'brain-Fleming type of α-spending function $\alpha(-4,t)$.

Subgroup analyses:

Prior history of a PUB (yes/no), age (<65 years/≥65 years), gender, race (caucasian/other), study region (U.S./non-U.S.), use of systemic corticosteroids at baseline and H. pylori status at baseline (positive/negative requested by the agency) were evaluated to determine whether or not the effect of rofecoxib compared to naproxen was consistent in the subgroups. For each subgroup variable listed above, a Cox regression model was used for the primary end point and included the treatment, subgroup, and treatment-by-subgroup interaction.

III. Study Results

Three hundred and one sites from United States and other nations screened 9539 patients. Eight thousand and seventy-six patients were enrolled between Jan 14,1999 to March 17, 2000. The median duration of time in the study was 9.0 months ranged from 0.5 month to 13 months. Four thousand and forty-seven patients were randomized to receive rofecoxib, and 4029 were randomized to naproxen treatment group. A total of 151 patients were excluded from the per-protocol analysis (73 and 78 patients in the rofecoxib and naproxen treatment groups, respectively). Patient accounting information was summarized in Table 1-1.

Table 1-1: Patient accounting information at the end of the study.

Patient Accounting	rofecoxib 50 mg N (%)	naproxen 1000 mg N (%)	Total N (%)
Total	4047	4029	8076
Completed	2862 (70.7)	2880 (71.5)	5742 (71.1)
Death *	0 (0.0)	1 (0.3)	1 (0.5)
GI(confirmed & unconfirmed)*	2 (0.0)	5 (0.1)	7 (0.0)
Discontinued	1185 (29.3)	1149 (28.5)	2334 (28.9)
Death	22 (0.5)	14 (0.3)	36 (0.4)
GI(confirmed & unconfirmed)	60 (1.5)	130 (3.2)	190 (2.3)
Clinical AEs	563 (13.9)	492 (12.2)	1055 (13.1)
Laboratory AEs	22 (0.5)	12 (0.3)	34 (0.4)
Lack of efficacy	256 (6.3)	263 (6.5)	519 (6.4)
Lost to follow-up	6 (0.2)	4 (0.1)	10 (0.1)
Other reasons	27 (0.7)	30 (0.7)	57 (0.7)
Patients moved	17 (0.4)	16 (0.4)	33 (0.4)
Patient withdrew consent	138 (3.4)	130 (3.2)	268 (3.3)
Protocol deviations	74 (1.8)	58 (1.4)	132 (1.6)
Total death	22 (0.5)	15 (0.3)	37 (0.5)
Total GI(confirmed & unconfirmed)	64 (1.6)	142 (3.5)	206 (2.5)

* Occurred after the completion of the trial.
Source: Sponsor's response on Jan. 11, 2001 to an request from the agency.

One thousand and one hundred thirty-one and 1032 patients in the rofecoxib and naproxen groups, respectively, discontinued the study for any reason other than the primary endpoint. The rates of discontinuation were 42.6 and 38.9 per 100 patients years for rofecoxib and naproxen, respectively. The relative risk for rofecoxib vs. naproxen was 1.10 (95% CI: 1.01, 1.19; p=0.033). This showed rofecoxib treatment group had statistically significantly more patients discontinued study than that in naproxen group for reasons other than the primary endpoint.

Thirty-seven deaths occurred in the VIGOR trial, 22 (0.5%) and 15 (0.4%) in the rofecoxib and naproxen groups, respectively.

Demographic information and baseline disease assessments of RA showed reasonable balances between treatment groups.

Reviewer's comment on discontinuations:

As the withdrawal rate was about 30% in the VIGOR study and there were only about 2% patients developed the GI end point, it was a concern if the relatively high withdrawal

rate (compared PUB event rate) could introduce potential bias in analysis results. Patients discontinued the study for reasons other than the study end point formed censoring for the end point PUBs. Some of the censoring such as withdrawal due to moving, lost to follow-up and lack of efficacy were unlikely to be informative censoring to PUBs, therefore were not the sources of bias. Protocol deviation and consent withdrawal was considered to be non-informative censoring as well since the reasons of those withdrawals were not directly associated with the end point. This was confirmed with the medical reviewer, Dr. Villalba, who had reviewed samples of those case report forms. Some of those who discontinued the study due to clinical and laboratory adverse events, especially those who discontinued due to GI related adverse events, might be informative censoring to PUBs if the adverse events were the pre-cursor of PUB. In this case, bias could occur. In the VIGOR study, there were 370 (9.2%) patients discontinued study due to adverse reaction in digestive system in naproxen treatment group and 267(6.6%) in rofecoxib group. If the bias exists, the risk of developing PUBs in naproxen treatment group could be under estimated. However, the association of the GI related adverse events to the study end point PUBs was not well understood by medical experts. Therefore, it was difficult to assess any potential bias possibly caused by discontinuation due to GI related adverse events. If the withdrawal mechanism is exactly the same in practice as that was in the VIGOR trial, there was no need to worry about the bias even if such a bias exists. However, if the withdrawal pattern is different, we may observe different risks of PUBs in post-marketing data.

1. GI events:

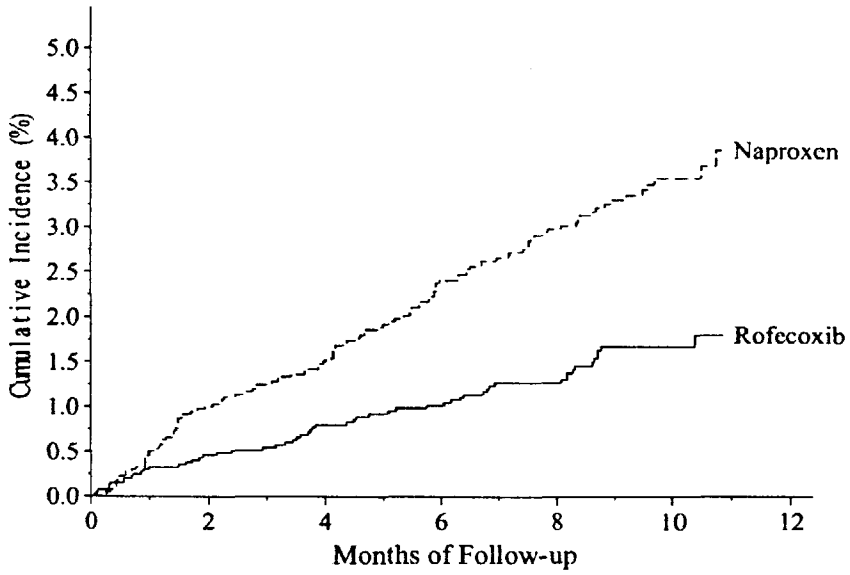
Sponsor's results of primary endpoint at the end of the study:

A total of 208 patients with potential PUB events were adjudicated. Sixteen events that occurred more than 14 days after discontinuation of study therapy were excluded from the primary analysis. Of the 16 events, six occurred in rofecoxib group and 10 in naproxen group. One hundred and ninety-one patients with PUBs were eligible for the primary analyses: 177 patients had confirmed events, 13 were unconfirmed and 1 was classified as "not an upper GI event". Of the 177 PUB events, 56 occurred in rofecoxib treatment group and 121 in naproxen group. The risk rates for the confirmed PUBs were 2.08 and 4.49 per 100 patient-years for rofecoxib and naproxen respectively. Based on Cox model with a stratification factor (prior history of PUBs) as a covariate, the relative risk of developing confirmed GI PUBs for rofecoxib treatment group vs. naproxen treatment group was 0.46 with 95% CI (0.33, 0.64) and p-value <0.001. The results of the primary analysis as well as some of results from the secondary endpoints are summarized in Table 2. Figure 1 showed the cumulative incidence curves for the confirmed PUBs of the two treatment groups.

In the per-protocol analysis, 48 rofecoxib patients and 113 naproxen patients experienced 1 or more confirmed PUBs with rates of 1.80 and 4.25, respectively, per 100 patient-years at risk. The relative risk based on the Cox model was 0.42 (95% CI: 0.30 to 0.59); p<0.001. These results were consistent with the primary analysis.

Figure 1

Primary Endpoint—Confirmed PUBs
Time-to-Event Plot (All-Patients-Randomized)



isk							
oxib	n=4047	3641	3402	3180	2806	1073	533
xen	n=4029	3644	3389	3163	2796	1071	513

Source: VIGOR clinical trial report submitted on June 29, 2000.

Sponsor's interim analysis:

Interim analysis was conducted when 66 confirmed PUBs were observed, 20 from rofecoxib treatment group and 46 from the naproxen group. The risk ratio of developing confirmed PUBs for rofecoxib vs. naproxen was 0.44 with p-value 0.002 and 95%CI (0.26, 0.74). The results of interim analysis were consistent with the final result.

Sponor's secondary GI endpoints at the end of study:

There were 16 rofecoxib patients and 37 naproxen patients that experienced 1 or more confirmed, complicated PUBs with rates of 0.59 and 1.37, respectively, per 100 patient-years at risk. The relative risk from the Cox model stratified by prior history of PUBs was 0.43 (95% CI: 0.24 to 0.78) and p=0.005.

There were 58 rofecoxib patients and 132 naproxen patients that experienced 1 or more confirmed and unconfirmed PUBs with rates of 2.15 and 4.90, respectively, per 100 patient-years at risk. The relative risk from the Cox model stratified by prior history of PUBs and study region was 0.44 (95% CI: 0.32 to 0.60) and p<0.001.

There were 17 rofecoxib patients and 42 naproxen patients that experienced 1 or more confirmed and unconfirmed complicated PUBs with rates of 0.63 and 1.56, respectively, per 100 patient-years at risk. The relative risk from the Cox model stratified by prior history of PUBs was 0.40 (95% CI: 0.23 to 0.71) and p=0.002.

Thirty-one rofecoxib patients and 82 naproxen patients experienced 1 or more GI bleeds with rates of 1.15 and 3.04, respectively, per 100 patient-years at risk. The relative risk from the Cox model stratified by prior history of PUBs was 0.38 (95% CI: 0.25 to 0.57) and p<0.001.

Table 2: Sponsor's analyses on GI end points at the end of study.

Endpoint	Treatment	N	Events	Rates	Relative Risk		
					Estimate	95%CI	p-value
Primary-Confirmed PUBs	rofecoxib	4047	56	2.08	0.46	(0.33, 0.64)	<0.001
	naproxen	4029	121	4.49			
Secondary Endpoints							
Confirmed, complicated PUBs	rofecoxib	4047	16	0.59	0.43	(0.24, 0.78)	0.005
	naproxen	4029	37	1.37			
Confirmed and unconfirmed PUBs	rofecoxib	4047	58	2.15	0.44	(0.32,0.60)	<0.001
	naproxen	4029	132	4.90			
Confirmed & unconfirmed complicated PUBs	rofecoxib	4047	17	0.63	0.40	(0.23, 0.71)	0.002
	naproxen	4029	42	1.56			
Any GI bleeds	rofecoxib	4047	31	1.15	0.38	(0.25, 0.57)	<0.001
	naproxen	4029	82	3.04			

Source: VIGOR clinical trial report submitted on June 29, 2000.

Subgroup analyses:

In addition to the subgroup analyses specified in DAP, the agency requested some additional subgroup analyses including prior cardiovascular history and baseline NSAID usage on confirmed PUBs, as well as study region effects on confirmed complicated PUBs.

Table 3 listed some of the results from those subgroup analyses that either had statistically significant subgroup effects at level 0.05 or statistically significant subgroup by treatment interactions at level 0.10.

Reviewer's comment on subgroup analyses:

Subgroup analysis based on prior history of PUBs (yes or no) suggested that there were statistically significantly (p-value=0.0001) increased risk of developing PUBs in the subgroup that prior history of PUBs existed, as compared to the subgroup that had no prior history of PUBs. However, the risk ratios between the two treatment groups were similar in both subgroups. Similar observations were found in subgroups based on baseline NSAIDs use or age groups (<65 years old or ≥65 years old).

Table 3: Results of subgroup analyses.

Subgroups:	Treatment	N	Events	Rates	Relative Risk	
					Estimate	95%CI
Prior history of PUBs: p-value for prior history=0.0001, for interaction=0.874						
Prior history of PUBs:	rofecoxib	314	13	6.72	0.44	(0.23, 0.85)
	naproxen	316	29	15.33		
No prior history of PUBs	rofecoxib	3733	43	1.72	0.47	(0.33, 0.67)
	naproxen	3713	92	3.67		
Age: p-values for age=0.0001 , for interaction=0.466						
Non-elderly (<65 years)	rofecoxib	3050	34	1.64	0.52	(0.34, 079)
	naproxen	2959	64	3.15		
Elderly (≥65 years)	rofecoxib	997	22	3.54	0.41	(0.25, 0.67)
	naproxen	1070	57	8.63		
Baseline steroid use: p-values for baseline steroid use=0.0012 , for interaction=0.073						
No baseline steroid use	rofecoxib	1803	24	2.03	0.68	(0.41, 1.15)
	naproxen	1776	35	2.97		
Baseline steroid use	rofecoxib	2244	32	2.11	0.37	(0.25, 0.56)
	naproxen	2253	86	5.67		
H. Pylori: p-values for H.Pylori=0.8800, for interaction=0.043						
Negative H. Pylori	rofecoxib	2244	21	1.43	0.32	(0.19, 0.52)
	naproxen	2260	67	4.51		
Positive H. Pylori	rofecoxib	1740	34	2.87	0.62	(0.40, 0.95)
	naproxen	1712	54	4.62		
Baseline NSAIDs use: p-values for NSAIDs use=0.0011, for interaction=0.645						
No baseline NSAIDs use	rofecoxib	703	14	3.07	0.41	(0.22, 0.76)
	naproxen	688	33	7.59		
Baseline NSAIDs use	rofecoxib	3344	42	1.87	0.48	(0.33, 0.69)
	naproxen	3341	88	3.89		

Sources: VIGOR clinical trial report submitted on June 29, 2000. P-values for subgroup effect were added by the reviewer.

It was reasonable that patients with prior history of PUBs or older than 65 years old had higher risk for PUBs, no matter which treatment they were receiving. However, it was not clear why the patients who were not NSIADs users at baseline also had relatively higher risk compared with those who were NSAIDs user at baseline. Even the non-NSAIDs users at baseline who received rofecoxib had risk of PUBs similar to naproxen patients who were NSAIDs user at baseline. One possible reason could be that some of the patients who were not NSAIDs users at baseline might be those who could not tolerate NSAIDs before and at high risk of PUBs.

Statistically significant (p-value=0.073) treatment by baseline steroid use interaction was observed. This was due to the increased risk of developing PUBs in naproxen treatment group in the subgroup that had baseline steroid use. Similarly, statistically significant treatment by baseline H. pylori status interaction was observed (p-value=0.043). This

interaction was due to the increased risk of PUBs in rofecoxib treatment group in H. pylori positive subgroup.

Since statistically significant subgroup effects were observed in age groups (<65 years old or ≥65 years old), prior history of PUBs, baseline NSAIDs use and baseline steroid use, a proportional hazard model including all the factors as covariates was used to analyze the primary end point. The result of this analysis was similar to the primary analysis with only the stratification factor as the covariate. The treatment difference in risks of developing PUBs observed in this study was very robust.

Reviewer's comments on Study 69 and generalization of the VIGOR results:

Study 69 was submitted in the original rofecoxib NDA to support the claim of GI safety of rofecoxib. In the original NDA review, this study was discredited to support GI safety claim by the agency. It was brought up again by the sponsor in this supplemental NDA submission and presented to the advisory committee to support the generalizability of the GI safety results obtained from the VIGOR trial. The issue here was whether Study 69 can be used to support the generalization.

As it was reviewed in the original NDA submission, Study 69 consisted of about 8 phase II/III trials that had different dose levels of rofecoxib, different study duration, and different NSAID comparators. There were three 6-week studies, two 6-month studies and three studies lasted over one year. The dose ranges of rofecoxib were from 12.5 mg to 50 mg. [redacted]. The NSAIDs comparators used in these trials included nabumetone, ibuprofen, and diclofenac. [redacted]. The observed incidence rates of PUBs were 1.61 and 3.58 per 100 patient-years in combined rofecoxib group and combined NSAIDs group, respectively. The results of Study 69 in comparison to the VIGOR trial were listed in Table 4.

Table 4: Comparisons of risks for GI events between Study 69 and the VIGOR trial.

Upper GI events	Treatment	Study 69 (OA patients)		VIGOR (RA patients)	
		Incidence rate	Risk ratio	Incidence rate	Risk ratio
Confirmed	rofecoxib	1.61	0.45	2.08	0.46
	NSAIDs	3.58		4.49	
Confirmed & unconfirmed	rofecoxib	1.61	0.35	2.15	0.44
	NSAIDs	4.56		4.90	
Confirmed complicated	rofecoxib	0.42	0.52	0.59	0.43
	NSAIDs	0.81		1.37	
Confirmed & unconfirmed complicated	rofecoxib	0.42	0.37	0.63	0.40
	NSAIDs	1.14		1.56	

Source: Advisory committee meeting package submitted on Dec. 18, 2000.

As can be seen from Table 4, the risks of developing GI events for rofecoxib treatment group in Study 69 were consistently lower numerically than that the VIGOR trial. This suggested a possible dose response relationship for the risk of GI events in rofecoxib treatment. However, the difference of the rates between the two studies was small. As the

rates of GI events for combined rofecoxib treatment in Study 69 was primarily driven by the low doses of rofecoxib (12.5 and 25 mg), such small difference between the two studies suggested a shallow dose-response relationship for the risk of GI events in rofecoxib treatment group. It also suggested that the difference between study populations (RA and OA) on GI safety might be ignoble.

It also can be seen that the risks of developing GI events in NSAIDs treatment group in Study 69 were consistently lower numerically than that observed in the VIGOR trial in naproxen treatment. Since the rate was some weighted average risks based on the three NSAID comparators, this average risk rate may vary as the change of NSAIDs or the change of proportion for certain NSAIDs. This suggested that different NSAIDs may have different risks in developing GI events and some NSAIDs may have lower risk of developing GI events than naproxen. However, Study 69 did not provide more information regarding the treatment difference in GI events in comparison to other individual NSAIDs, especially those that had less GI toxicity than naproxen.

The value of Study 69 lies in that it did not contradict what was observed in the VIGOR trial. However, it should by no means be used to generalize the observations of the VIGOR trial. Despite the statistically significant superior GI safety profile observed in the two studies, the GI safety profile of rofecoxib in comparison to the whole class of NSAIDs was not clear.

2. Safety analysis:

Pre-specified safety variables:

Survival analysis using Cox proportional hazard model with treatment as the covariate was used to analyze the pre-specified adverse experiences. Results that were statistically significant at level 0.1 were listed in Table 5.

Table 5: Results of pre-specified safety analyses.

Type of Adverse Experience	Treatment	N	Events	Rates	Relative Risk		
					Estimate	95%CI	p-value
Serious clinical Aes	rofecoxib	4047	378	14.48	1.21	(1.04,1.40)	0.013
	naproxen	4029	315	11.97			
Discontinued due to GI Aes + abdominal pain	rofecoxib	4047	307	11.47	0.73	(0.63, 0.85)	<0.001
	naproxen	4029	416	15.62			
Discontinued due to edema-related Aes	rofecoxib	4047	25	0.93	1.92	(0.98,3.75)	0.057
	naproxen	4029	13	0.48			
Discontinued due to hypertension-related Aes	rofecoxib	4047	28	1.04	4.67	(1.93, 11.28)	<0.001
	naproxen	4029	6	0.22			
Discontinued due to hepatic disease Aes	rofecoxib	4047	10	0.37	3.33	(0.92, 12.11)	0.067
	naproxen	4029	3	0.11			
CHF Aes	rofecoxib	4047	19	0.70	2.11	(0.96, 4.67)	0.065
	naproxen	4029	9	0.33			

Lab Aes leading to discontinuation	rofecoxib naproxen	4047 4029	22 12	0.82 0.44	1.83	(0.91, 3.71)	0.091

Source: VIGOR clinical trial report submitted on June 29, 2000.

Reviewer's comment on safety analyses:

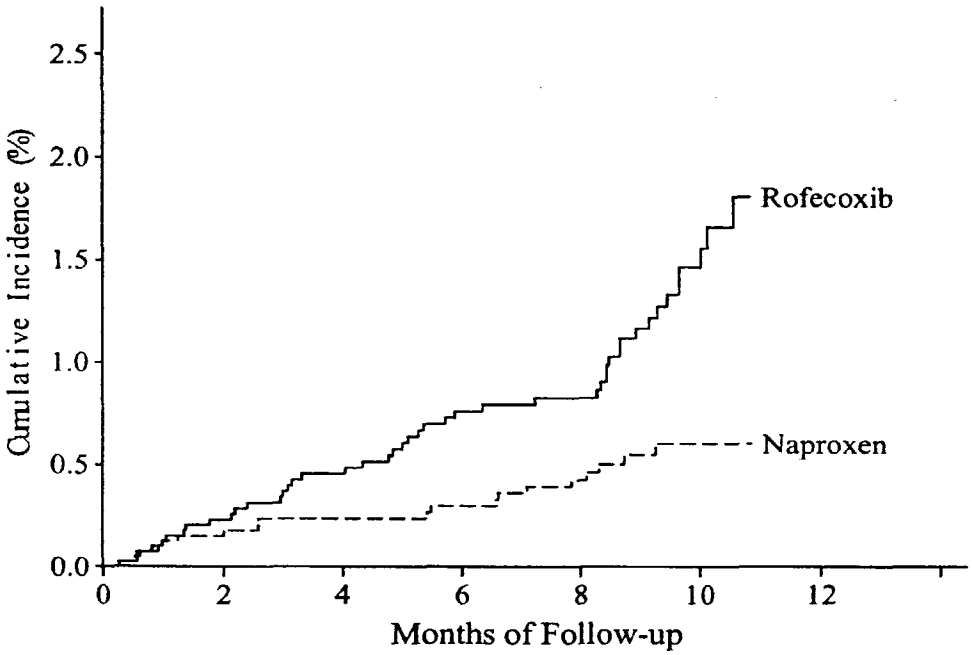
Most clinical trials were not powered to detect safety differences among treatments. It is important to identify those treatment differences so that a comprehensive understanding to treatment procedures could be obtained. Statistically, the p-values were used to identify all the possible safety differences rather than make decisions. Therefore, instead of adjusting multiple tests, statistical significance level 0.1 was used in Table 5 to identify the safety variables that showed possible treatment difference.

As can be seen from the table, rofecoxib treatment group had statistically significantly less patients ($p<0.001$) discontinued due to GI adverse events and abdominal pain than naproxen treatment group. However, compared with naproxen, more patients in rofecoxib treatment group experienced serious clinical adverse events ($p=0.013$); more patients in rofecoxib discontinued due to edema-related adverse events ($p=0.057$); more patients in rofecoxib discontinued due to hypertension-related adverse events ($p<0.001$); more patients in rofecoxib discontinued due to hepatic disease ($p=0.067$); more patients in rofecoxib experienced CHF adverse events ($p=0.065$); and more patients in rofecoxib discontinued due to lab adverse events ($p=0.091$). Based on the pre-specified analyses on safety variables, rofecoxib 50 mg daily revealed several undesirable safety concerns compared to naproxen in this VIGOR trial.

Cardiovascular events:

Ninety-eight cases (65 from rofecoxib and 33 from naproxen) of cardiovascular serious adverse events were sent for adjudication to the vascular endpoint adjudication committee. Forty-six cases from 45 rofecoxib patients and 20 cases from 19 naproxen patients were confirmed to have thrombotic cardiovascular serious adverse events. The sponsor's analyses were focused on the 66 confirmed cases from the 64 patients. The result of survival analysis on the 64 patients showed that the risk of developing a cardiovascular event in rofecoxib treatment group was 2.37 times of that in naproxen treatment group with p-value 0.0016 and 95% CI (1.39, 4.06). Figure 2 showed the cumulative incidence curves of the two treatment groups for confirmed thrombotic cardiovascular serious adverse events.

Figure 2
Confirmed Thrombotic Cardiovascular Serious Adverse Experiences
in Rheumatoid Arthritis Patients in the VIGOR Study
Time-to-Event Plot (All Patients Randomized)
Updated Application Data



sk						
oxib	n=4047	3643	3405	3177	2806	1067 531
xen	n=4029	3647	3395	3172	2798	1073 514

Source: Safety update submitted on Oct. 13, 2000.

Results from some of the supportive analyses on the thrombotic cardiovascular serious adverse events were listed below,

(1) Subgroup analysis (Aspirin indicated vs. aspirin not indicated): only 321 patients were aspirin indicated patients (170 in rofecoxib and 151 in naproxen). The risk ratio of developing serious cardiovascular events between rofecoxib and naproxen was 4.89 with p-value 0.012 and 95% CI (1.41, 16.88). The risk ratio for aspirin not indicated patients was 1.89 with p-value 0.041 and 95% CI (1.03, 3.45).
(2) Analyses of cardiovascular events in the VIGOR study using endpoint definition standard in large anti-platelet trials: for composite endpoint including cardiovascular death, MI and CVA, 35 events occurred in rofecoxib treatment group and 18 in

naproxen group. The risk ratio for such events was 1.96 for naproxen vs. rofecoxib with 95% CI (1.10, 3.45).

(3) Incidence of events judged by investigators to be potential thrombotic cardiovascular serious adverse experiences: As mentioned before, events experienced by 64 patients in rofecoxib and 32 patients in naproxen were eligible for adjudication. The risk ratio of experiencing such events was 2 for rofecoxib vs. naproxen with 95% CI (1.32, 3.03).

Reviewer's comments on cardiovascular serious adverse events:

In addition to the analyses of thrombotic cardiovascular serious adverse events, all the cardiovascular events from the adverse event data sets that were serious in investigator's opinion were compared between the two treatments. One hundred and eleven patients in rofecoxib treatment group experienced serious cardiovascular adverse events, while 50 patients in naproxen treatment group experienced such events. Survival analysis showed the risk for serious cardiovascular events in rofecoxib treatment group was 2.22 times of the risk in naproxen treatment group. The p-value obtained from survival analysis was 0.0001.

Based on the sponsor's primary analysis on confirmed thrombotic cardiovascular serious adverse events and other supportive analyses on cardiovascular serious adverse events, there was clear evidence to show that rofecoxib 50 mg daily had increased risk of developing serious cardiovascular adverse events compared to naproxen 1000 mg per day.

3. Efficacy:

The results of this study on the patient and investigator global assessments of disease status and HAQ did not show treatment difference between rofecoxib and naproxen. Analysis on discontinuation due to lack of efficacy yielded p-value 0.769.

Since the VIGOR trial was not designed to evaluate efficacy in treating RA patients, the results of the efficacy analyses could not be used to establish efficacy property of rofecoxib 50 mg daily in comparison to naproxen 1000 mg per day in RA patients.

IV. Meta-analysis on cardiovascular events:

In response to the finding on cardiovascular events in the VIGOR trial, the sponsor organized a large scale meta-analysis that consisted of more than 25 studies in different phase stages and more than 28,000 patients, including the VIGOR trial. The purposes of the meta analysis were to assess the incidence of thrombotic cardiovascular events in patients treated with rofecoxib compared to naproxen, other (non-naproxen) NSAIDs, or placebo. The end point used in the meta-analysis was the combined endpoint defined by

the Antiplatelet Trialists' Collaboration (APTC). This end point was the secondary end point defined in the VIGOR trial.

The studies included in the meta-analysis differed in many aspects. The main difference can be summarized as the follows:
(1) different dose levels of rofecoxib (from 12.5 mg to 50 mg) were pooled together;
(2) different NSAIDs comparators were used in different studies;
(3) different study duration (from 6 weeks to more than one year) were pooled together;
(4) different indications (RA, OA, Alzheimer and ——— were combined by stratified analyses
(5) some of the end points were adjudicated while some were not.

Data were analyzed based on three primary comparisons: rofecoxib vs. naproxen, rofecoxib vs. non-naproxen NSAIDs, and rofecoxib vs. placebo. The results of the comparisons were listed in Table 6. Based on the results listed in Table 6, the sponsor concluded that "the risk of sustaining a cardiovascular thrombotic event is similar on rofecoxib, placebo and nonselective NSAIDs without sustained antiplatelet activity (non-naproxen NSAIDs)".

Table 6: Results from sponsor's meta-analysis.

Indication for treatment	Rofecoxib		Comparator		Relative risk (95% CI)
	N	Cases/PYR (rate)	N	Cases/PYR (rate)	
Rofecoxib vs. naproxen					
RA	6057	46/3947 (1.17)	4859	20/3078 (0.65)	1.74 (1.02, 2.96)
OA	3026	11/675 (1.63)	3011	7/665 (1.05)	1.55 (0.60, 4.00)
Alz/ —	0	-	0	-	-
Total	9083	57/4622 (1.23)	7870	27/3742 (0.72)	1.69 (1.07, 2.69)
Rofecoxib vs. other nonselective NSAIDs					
RA	0	-	0	-	-
OA	4549	21/1934 (1.09)	2755	14/984 (1.42)	0.79 (0.40, 1.55)
Alz/ —	0	-	0	-	-
Total	4549	21/1934 (1.09)	2755	14/984 (1.42)	0.79 (0.40,1.55)
Rofecoxib vs. placebo					
RA	1622	3/337 (0.89)	989	1/201 (0.50)	1.78 (0.14, 93.70)
OA	3165	12/655 (1.83)	1215	3/232 (1.30)	1.53 (0.43, 5.44)
Alz/ —	1503	18/1197 (1.50)	1278	28/1246 (2.25)	0.68 (0.37, 1.23)
Total	6290	33/2189 (1.51)	3482	32/1678 (1.91)	0.84 (0.51, 1.38)

Source: Meta-analysis report submitted on Jan. 8, 2001.

Reviewer's comments on the meta-analysis:

The objective of the meta-analysis stated by the sponsor was ambiguous. As the issue of cardiovascular risk arose from rofecoxib 50 mg in the VIGOR trial, it is important to understand the role of rofecoxib 50 mg in relation to cardiovascular event. Therefore, two questions should be asked from this meta analysis: one was the role of rofecoxib 50 mg in relation to cardiovascular events, and the other was the role of rofecoxib lower doses such as 12.5 and 25 mg in relation to the cardiovascular events.

The meta-analysis database did not provide sufficient information about rofecoxib 50 mg at all. Of the 28,000 patients in the meta-analysis database, there were about 6,000 patients in rofecoxib 50 mg treatment. In these 6,000 patients, 4,047 patients were from the VIGOR trial. Only 1,881 patients were from trials outside the VIGOR trial. Among these patients, only about 900 patients were from the trials that had duration longer than 6 month. By sample size calculation, more than 4,000 patients are needed to observe statistical significant treatment difference at level 0.05 (two-sided) with 80% power for the treatment difference observed in the VIGOR trial.

Combining all level doses of rofecoxib made it difficult to assess the role of rofecoxib 50 mg. No convincing evidence was provided by the sponsor that there was no dose response relationship to cardiovascular events for rofecoxib. Without such evidence, combining all doses of rofecoxib together may obscure the risk associated with high dose level.

Different duration of the trials complicated the interpretation of cardiovascular events in relation to rofecoxib 50 mg if the long-term studies had only lower doses of rofecoxib. In addition, as it can be seen from the cumulative incidence curves for serious thrombotic cardiovascular events from the VIGOR trial in Figure 2, the two incidence curves did not separate until 6 weeks after receiving treatments. This suggested that trials with short duration would not be of any help in demonstrating the treatment difference.

Comparison between rofecoxib and naproxen was basically in the RA and OA patients. The majority patients in the RA indication were from the VIGOR trial. The OA indication had only lower doses of rofecoxib and short term duration trials (6 and 12 weeks). As can be seen from Table 6, there were only a fraction of patient-years and fewer events from OA indication for this comparison. Therefore, the result of this comparison was primarily driven by the VIGOR trial, and not surprisingly, the result was consistent with what was observed in the VIGOR trial.

For the comparison between rofecoxib and other NSAIDs (non-naproxen), only OA patients were used in this comparison. The majority rofecoxib patients were rofecoxib low doses patients. Notice that the group of OA patients used in this comparison was different from those used in the comparison between rofecoxib and naproxen. The risk rate for rofecoxib in this comparison was lower (1.09 per 100 patient-years) compared to that in OA patients in the other comparison with naproxen (1.63 per 100 patient-years). The conclusion driven by this comparison was from rofecoxib low dose treatment and combination of three NSAIDs, diclofenic, ibuprofen and nabumetone.

For the comparison between rofecoxib and placebo, all three indications were included. However, the result was primarily driven by the Alzheimer and ——— patients as this indication had more patient-years and events compared to RA and OA indication. The indication again did not have any patients taking rofecoxib 50 mg. Notice that those RA patients used for the comparison between rofecoxib and placebo were only a subset of the RA patients used in the comparison between rofecoxib and naproxen excluding the

patients from the VIGOR trial. It was also true for the OA patients used in this comparison. The point estimates for the relative risk ratios for OA and RA indications were in the opposite direction of the Alzheimer and ——— indication, which made the risk ratio of the stratified analysis approaching 1 and concluded similar risks.

Clearly, the conclusion based on the three comparisons in this meta-analysis can not be applied to rofecoxib 50 mg. On the other hand, the results of the meta-analysis did not provide signal of cardiovascular risk for low doses of rofecoxib, as no treatment difference was observed in the comparisons between rofecoxib and non-naproxen NSAIDs, and between rofecoxib and placebo. However, the results from this meta-analysis can not be used to show there was no cardiovascular risk for low doses of rofecoxib, as the comparisons were under powered and driven by special patient population such as Alzheimer patients.

V. Conclusion:

The VIGOR trial demonstrated robustly that rofecoxib 50 mg daily treatment statistically significantly reduced risk of developing PUBs compared to naproxen 1000 mg per day treatment in RA patients. The risk of PUBs in rofecoxib treatment group was reduced 0.46 times of that in naproxen treatment group, with 95% CI (0.33, 0.64). The risk of confirmed and complicated PUBs was also reduced 0.43 times with 95% CI (0.24, 0.78). All other secondary GI end points and secondary analyses supported the finding.

The VIGOR trial also revealed some safety concerns for the use of rofecoxib 50 mg daily. For the 12 pre-specified safety analyses, half of them showed statistically significant trend of undesirable safety aspects for rofecoxib 50 mg daily compared to naproxen 1000 mg per day. These undesirable safety aspects included serious clinical adverse events, discontinued due to edema related AEs, discontinued due to hypertension related AEs, discontinuation due to hepatic diseases, CHF AEs, Lab AEs leading to discontinuation. Analyses on confirmed thrombotic cardiovascular serious adverse events showed rofecoxib 50 mg daily had increased the risk of the event 2.38 times compared with naproxen 1000 mg per day. Analysis on serious cardiovascular adverse events judged by the investigators also showed that rofecoxib 50 mg daily doubled the risk of such events compared to naproxen 1000 mg per day. The meta-analysis did not provide any new information about the risk of cardiovascular events in relation to rofecoxib 50 mg.

The results from the VIGOR trial for GI safety can not be generalized to other NSAIDs due to the study design. As it was pointed out in this review, Study 69 did not provide evidence for such generalization at all. Also it was not appropriate to claim that the risk of sustaining a cardiovascular event was similar to placebo and other non-naproxen NSAIDs. The results generated from the meta-analysis did not provide more information about the cardiovascular risk for rofecoxib 50 mg than the VIGOR trial.

Qian Li, Sc.D
Mathematical Statistician

Concur:

APPEARS THIS WAY
ON ORIGINAL

Stan Lin , Ph.D
Team Leader

CC:
HFD-550/Division File
HFD-550/Dr. Bull
HFD-550/Dr. Goldkind
HFD-550/Dr. Villalba
HFD-550/Ms. Cook

HFD-725/Division File
HFD-725/Dr. Huque
HFD-725/Dr. Lin
HFD-725/Dr. Li

APPEARS THIS WAY
ON ORIGINAL

/s/

Qian Li
3/23/01 07:10:24 PM
BIOMETRICS

Stan Lin
3/23/01 07:15:41 PM
UNKNOWN

APPEARS THIS WAY
ON ORIGINAL