Medical Officer's Advisory Committee GI Briefing Document
Division of Anti-Inflammatory, Analgesic and Ophthalmologic Drug Products: HFD-550

NDA 21,042 S 007
21,052
Sponsor: Merck
Name of drug: Rofecoxib (Vioxx ™)
Dose: 50 mg

Subject of Consult: Review of Vioxx Gastrointestinal Outcomes Research Study (VIGOR)

Materials reviewed: Protocol and Completed Study Report 88C
Submission date: June 29, 2000

Reviewer: Lawrence Goldkind M.D.

Table of contents

	Page
Background	2
Clinical studies	5
Study 88C	5
Study design and reviewer comments	5
Conclusions regarding study design	10
Results	11
Demographics	11
GI outcome results	14
Subgroup analyses	21
Overall conclusions	28
Appendix 1: Protocol provisions	29

Background

Vioxx (V) was approved in 1999 for the treatment of osteoarthritis (OA), acute pain and primary dysmenorrhea. The approved dose was 12.5 and 25mg/day for OA and 50 mg/day for acute pain and dysmenorrhea. This product is a highly selective inhibitor of cyclooxygenase-2 (COX-2). The drive to develop highly selective cyclooxygenase (COX) inhibitors was based on the hopes that the safety profile would be improved compared to less selective COX inhibitors. Upper gastrointestinal ulcers complicated by pain, bleeding and perforation are a labeled complication of NSAIDs. Of the two isoforms, COX-1, a constituitively-generated enzyme has been considered critical to the maintenance of the upper gastrointestinal mucosal integrity. Physiological mechanisms that are linked to "maintenance" effects of COX-1 generated prostaglandins include gastric mucous production, bicarbonate secretion and mucosal blood flow. Inhibition of this enzyme has been linked to the gastrointestinal toxicity of NSAIDs. COX-2 is upregulated in inflammatory conditions. Since the identification of the second isoform of COX, it has been hoped that selective inhibition of this isoform would effectively treat inflammatory conditions and pain with less gastrointestinal toxicity. The original NDA included extensive safety data related to upper gastrointestinal ulceration that are reflected in the product label. V was associated with fewer endoscopically defined (as opposed to symptomatically defined) ulcers compared to ibuprofen. The studies reviewed to date have not differentiated V from all NSAIDs studied in terms of gastrointestinal symptoms and clinically meaningful ulcers. Some symptoms appear to be more commonly associated with V compared to the other NSAIDs studied while some were more common in specific comparators.

Comparative safety claims are susceptible to bias selectively defining the events of interest without incorporating other potentially important toxicities. Comparative study of symptoms and clinically relevant outcomes must be linked to dose and specific comparator. Comparative study of safety and subsequent safety claims are intrinsically different than the well ploughed area of drug efficacy. Efficacy is typically established for a particular beneficial effect. Study can therefore be based on prespecified definitions, objectives, instruments of measurement and statistical analysis. Safety by comparison is multifaceted and therefore less easily studied and quantified. Specific safety claims other than those associated with ultimate endpoints such as death or permanent disability are difficult to study in an unbiased way that includes the concept of overall safety.

Upper gastrointestinal toxicity has been identified as a major health risk associated with the use of NSAIDs. Some estimates of the number of deaths due to the complications of gastrointestinal bleeding and perforation attributed to these products as a class are in the range of 10-20,000 per year in the United States. Based on these estimates, NSAIDs contain a generic warning of GI risk. Thus, gastrointestinal toxicity appears to be an appropriate specific safety issue for study. COX-2 selective inhibitors hold the promise of having less GI toxicity than less selective agents. Just as relative specificity of COX isoenzyme inhibition exists, so does the possibility of **relative** specificity of GI safety. Available information about the toxicity of NSAIDs suggests that each NSAID most likely has a somewhat unique profile. The study of relative safety has been limited by the difficulties inherent in safety studies compounded by the difficulties in comparative

studies of many agents, at different doses, over long periods of time, using different endpoints in heterogeneous populations. The presence of generic products further discourages large expensive comparative studies.

The most daunting challenge in the study of GI safety is that the most important outcomes of bleeding, obstruction and perforation are rare events, estimated to occur in less than several percent of patients on chronic NSAIDs per year. (The estimates of perforations, ulcers and bleeding that appear in the GI warning section of NSAID labels include ulcers associated with pain alone without the more serious complications). Therefore, large studies are required.

Once the morbid outcomes of bleeding, obstruction and perforation are excluded, it becomes difficult to define an appropriate safety comparison for NSAIDs. The majority of ulcers are painless and up to 30% of patients on NSAIDs experience abdominal pain. The correlation between UGI symptoms and mucosal damage is weak. Gastric adaptation to the effects of NSAIDs is well documented. This produces new difficult questions. Is abdominal pain less or more significant than other GI symptoms such as diarrhea, nausea or vomiting? Are such symptoms more relevant than other toxicities such as renal or hepatic damage?

The original NDA database suggested that V did **not** differentiate from the comparators studied in terms of symptoms as it did for endoscopic ulcers. Based on the absence of evidence for a distinctly different safety profile in terms of clinically meaningful outcomes, the current product label includes the same warnings regarding gastrointestinal toxicity that less selective NSAIDs have. Based on the theoretical advantages of COX selectivity discussed previously and the endoscopic data that appears in the product label, V has been proposed by some as “safer” than previously approved NSAIDs . Although it is tempting to accept the development of asymptomatic ulcers as a meaningful endpoint and a surrogate for clinically relevant outcomes, there is inadequate evidence to date to accept this as fact. The clinical outcome trial entitled, “MUCOSA” published in 1995 in conjunction with other studies of endoscopically defined ulcers associated with the use of NSAIDs and misoprostol are suggestive of a correlation. This study did not have prespecified outcomes and a statistical plan to allow for conclusions. Furthermore, MUCOSA cannot be extrapolated to all other potentially “gastroprotective” drugs.

Therefore, adequate evidence of a uniquely improved GI safety profile for V based on asymptomatic endoscopically defined ulcers was not established in the original NDA.

Databases are inadequate at the time of marketing to fully define the safety profile of a new drug. This is particularly true of new molecular entities and drug classes. (Some authors contend that COX-2 selective agents represent a new class. The World Health Organization has placed such agents in a separate class than traditional NSAIDs that are less selective.) The wide acceptance as evidenced by the many millions of prescriptions in the first year of marketing reflects acceptance of V as a safer alternative to traditional NSAIDs. However, clinically relevant safety endpoints are rare and may be missed in a database of even several thousand subjects. Authors outside the FDA have voiced

concern over this issue as well. The following excerpt from a lead editorial in the September 2000 Rheumatology journal highlights this issue.

> " While it is still true that Cox-1 is expressed constitutionally in most cells and Cox-2 is induced in sites of inflammation and other pathology, recent careful work has clarified several physiological situations in which Cox-2 inhibitors in the clinic are understood only partly at present…
>
> The driving force behind the rapid and forceful cooperation between basic science and drug development was concern about the serious toxicities of conventional NSAIDs and aspirin, not least the increased fatalities resulting from gastrointestinal bleeding and ulcer perforation. Those who are skeptical about extrapolation form databases such as ARAMIS are referred to a Finnish study that identified 30 fatalities from the use of NSAIDs in that country in a single year. Cox-2 is up-regulated in the inflamed joint, and the hypothesis was that selective inhibition of the inducible Cox-2 isoenzyme would offer therapeutic efficacy without this severe toxicity. Endoscopic data from clinical trials support this hypothesis, ***but information about the risk of serious events, i.e. bleeding and perforation is still not at hand. New insights into the biologic function of Cox-2 should caution us from the uncritical use of Cox-2 inhibitors. There is a convincing evidence from published trials that celecoxib is equivalent but not superior to conventional NSAIDs in the symptomatic control of osteoarthritis and rheumatoid arthritis. However, long-term safety data can be established only with time and, as with all new types of drugs, we should be vigilant in recognizing possible new types of problems. The questions that must still be addressed concern the ultimate consequences of selective inhibition of Cox-2 and its biological functions"*** [1]
>
> (bolding and italics added for emphasis by reviewer)

Another author in a review article in the New England Journal of Medicine stated that:

" In spite of enthusiasm for these promising new agents NSAIDs, some questions remain regarding their highly selective inhibition of cyclo-oxygenase-2. For example, cyclo-oxygenase-2 might generate endogenous prostanoids that are biologically active….

..although the highly selective cyclo-oxygenase-2 inhibitors offer considerable promise in the treatment of inflammatory arthritidies, careful surveillance will be important to determine their ultimate benefit and safety profile." [2]

The Division and the sponsor have agreed that evidence was needed regarding clinically meaningful upper gastrointestinal events as well as a large controlled database for overall safety assessment. While, upper gastrointestinal tract injury was the primary and prespecified endpoint, the sponsor and the Division shared the concerns noted by the above reference #1.

The VIGOR trial was conducted to establish a safety profile based on a large database to allow for meaningful study of clinically relevant outcomes.

Clinical Studies

088C: A double blind, Stratified, Parallel-group study to assess the incidence of PUBs during chronic treatment with V or naproxen in patients with Rheumatoid Arthritis

The original protocol and relevant amendments appear in appendix 1.

Studies 088 and 089 were two identical arms of a single international study that were intended to for combined analysis (088C). All comments apply to both protocols.

Objectives:

Primary

1. To determine the incidence rate of PUBs in patients with rheumatoid arthritis (RA) taking 50 mg V daily compared to patients taking naproxen, 1000mg daily

2. To study the safety and tolerability of V in patients with rheumatoid arthritis

Secondary

1. To assess the incidence of confirmed and unconfirmed PUBs in patients with RA taking 50 mg V daily compared to patients taking naproxen 1000 mg daily

2. To assess the incidence of complicated PUBs in patients with RA taking 50 mg MK-0966 daily compared to patients taking naproxen 1000 mg daily

3. To compare the efficacy of treatment of RA with V or naproxen as evaluated by the patient and investigator global assessment of disease activity and the discontinuation due to lack of efficacy

Reviewer's Comments related to objectives

Endpoint: PUB

This endpoint is a composite. The current GI warning on NSAID labels uses the term "PUB (perforation, ulcer, bleed)" to describe the GI events widely described in the medical literature at the time of the development of this section of NSAID labels. The medical literature at the time of development of this NSAID warning template was not standardized as to the definition of a PUB. This acronym in fact defines a symptomatic ulcer with or without serious morbidity. Such a term does define a clinically relevant endpoint. It represents ulcers identified during an evaluation of patients experiencing symptoms serious enough to warrant a physician intervention. Such an event must by definition be relevant to the patient. There are several difficulties with this endpoint as the primary endpoint of study in a controlled trial.

A. *Many patients on NSAIDs including V experience UGI symptoms that are consistent with ulcer symptoms that in fact are not related to ulcers. Up to 50% of patients on NSAIDs experience dyspepsia. Up to 15% discontinue therapy due to such symptoms.[2] Only a fraction of these patients have ulcers on UGI endoscopy. Thus, there are a significant number of patients who will have gastroduodenal ulcers on endoscopy without causal association with symptoms. The rate of such events would be even higher in a clinical trial where protocol driven ascertainment or bias within the clinical trial setting identifies ulcers that would not be identified in clinical practice. In clinical practice patients without alarm symptoms on NSAIDs are generally taken off presumed offending medication without any further sequelae. Therefore the use of the endpoint PUB in a clinical trial introduces a somewhat artificial entity that does not have the degree of clinical relevance that is inherent in the more clearly defined endpoint "POB" (perforation, obstruction or bleed).The sponsor has identified "complicated PUB" (the equivalent of POB) as a secondary endpoint.*

B. *Symptomatic ulcers, whether clinically or protocol derived do not represent the same severity of endpoint as POB. Only a small fraction of ulcers are thought to result in a clinically serious outcome. In the original NDA database for V the vast majority of ulcers identified were protocol derived and not related to any symptoms. A composite outcome should contain endpoints with similar clinical importance. The correlation between symptomatic ulcers and complicated ulcers is too weak to consider the two in the same endpoint of a prospective study. In fact, clinically silent ulcers that present with a complication rather than pain symptoms are felt to be potentially more dangerous since they do not provide warning of any underlying pathology. Therefore symptomatic ulcers may be uniquely inappropriate for inclusion in a composite endpoint of a serious outcome. The current NSAID warning used the endpoint "PUB" due to the limitations of the available data at the time of conception. This endpoint is not inherently the most appropriate composite endpoint to be studied prospectively. Furthermore, symptomatic ulcers are so much more*

common that POBs that the outcome would be primarily determined by the symptomatic ulcer results and therefore are most accurately defined as such. Separate analyses of POBs and symptomatic ulcers allow for a more meaningful and accurate understanding of the data. The sponsor has designated POBs as a well-controlled and well-defined secondary endpoint of study. In fact, the inclusion of both PUB and POB as important endpoints will allow for a further understanding of these composite endpoints in relation to one another for future trial design as well as scientific understanding of the correlation of parameters of UGI safety related to NSAIDs.

Endpoint: "All bleeds"

The criteria for this endpoint is found in the protocol reproduced in the appendix under section 5.5.1.6. This endpoint does not establish a new wll- documented meaningful endpoint that adds to the PUB and POB endpoints. If the intended endpoint was all UGI bleeding, a subset of the well-documented PUBs could be used. If the intent was to compare all significant GI bleeding (below the duodenum) in view of the lack of platelet effect and possibly diminished small bowel and colonic toxicity of a COX-2 selective agent a more rigorous approach may have been employed.

a. The adverse event terms used for screening should have been prespecified

b. Witnessed bleeding or occult positive stool or some prespecified degree of fall in hemoglobin should have been pre-defined

The criteria chosen may include self-reported dark colored stool/diarrhea or scant hemorroidal bleeding. The scenario of undocumented reports of melena or LGI bleeding has been seen frequently by this reviewer upon review of case report forms from clinical trials, particularly when GI outcomes are of interest.

Dose selection

The dose choice of 50 mg for V is the labeled dose for acute pain and twice the labeled dose for OA. As ***V is not approved for the treatment of RA****, it is unclear what the relationship of the tested dose will have to this population. The dose of naproxen is within the commonly used range for the treatment of OA and RA. While the NSAID comparators have been in use for years and have well-established dose ranges in practice, V is a relatively new molecular entity and has a less well established efficacy and dose ranging profile. A labeled safety advantage related to UGI events may suggest to consumers that there is room to "push" the dose of a drug with proposed analgesic as well as anti-inflammatory properties. This phenomenon of "dose creep" is particularly relevant in the treatment of pain when currently available therapies leave most patients with some residual pain (absence of total pain relief). The widely held expectation that new COX-2 selective agents will have little to no potential for UGI toxicity requires a robust proof of*

principle. ***Comparative safety claims therefore would be most meaningful for a high dose V. If general safety concerns prevent such doses the robustness of safety comparisons is less substantial and extrapolations from GI specific to general safety profiles could prove dangerous in practice.***

Selection of comparator

The selection of naproxen 1000mg is a reasonable choice. This is a widely used NSAID for pain, OA and RA. Generalizability is limited with the use of one comparator. As there is a range of GI toxicity within the NSAID class (albeit imperfectly characterized), a panel of several drugs across the spectrum would offer stronger support for a different class in terms of GI toxicity.

Disease model

The choice of disease is unusual, as V is not approved for the treatment of RA. Therefore the relevance of the results may be limited based on the patient population and dose selected. RA has been described as a higher risk condition compared to OA or the general population for GI toxicity with NSAID use. The largest and best-controlled data for the comparison of OA and RA appears in the CLASS trial published in 2000. This study suggested little difference in risk between these conditions in ambulatory patients. Therefore, safety ***comparisons*** *should be generalizable from RA to other rheumatological conditions. Absolute safety profile may be different for other clinical conditions, particularly those with higher morbidity overall.*

Hypothesis generating objectives

The large size of the trial will allow for assessment of other less common toxicities such as renal and cardiovascular effects. The inclusion in the final protocol of an adjudication process for cardiovascular effects and the collection of extensive laboratory information will provide a unique database for the safety of V as well as naproxen. Statistical comparisons however, will be difficult for relatively small differences in the face of the multiplicity of potential comparisons and the inherent post-hoc nature of such comparisons.

Reviewer's comments related to study design

The study size was based on a 90% power of identifying a 50% reduction in the incidence of PUBs at a 0.05% alpha level assuming a 2.25% annual incidence of PUBs in the naproxen group.

The study was well designed with adequate detail provided for randomization, double-blinding, and appropriately timed follow-up. An optimal study of chronic drug safety involves long term follow-up. The treatment period for this study was defined as the period until the last randomized patient had been observed for 6 months or 95 events had accrued. An amendment later extended the number of events to 120.

The absence of a screening endoscopy in a study population recently on NSAIDs may allow for the inclusion and therefore incorrect attribution of some ulcers, particularly early in the study. This design however is appropriate for an optimal risk assessment generalizable to clinical settings.

Exclusion criteria

1. *Subjects with a history of any illness or significant abnormalities on prestudy clinical evaluation that, in the opinion of the investigator, contraindicates a 1-2 year course of therapy with an NSAID were excluded. Subjects with significant active angina pectoris, congestive heart failure, suboptimally controlled hypertension or recent stroke or TIA were excluded as well. Subjects with a history of MI or coronary artery bypass graft surgery within prior year were explicitly excluded. Patients with health problems associated with morbid obesity were also excluded.* ***These exclusions are reasonable but significantly limit the generalizability of results. Thus, overall safety conclusions regarding the safety in sicker patients and particularly those with cardiovascular and renal disease will be substantially limited.***

2. ***Patients on any aspirin, including low doses were excluded. This exclusion prevents any confounding of PUB results that may be attributable to the effects of aspirin. However, the use of aspirin is so common in current preventive medicine practice that the generalizability of results will be substantially limited by the exclusion of patients that have the broad range of conditions that warrant low dose aspirin.***

Conclusion regarding study design

1. ***The endpoint of POB is of greatest significance as a true clinically serious outcome measure. The symptomatic ulcer however, represents a clinically meaningful additional endpoint.***

2. ***The dose of 50 mg is appropriate if a safety advantage based on the COX-2 hypothesis is to be tested. Ultimate use of this dose may be common based on the phenomenon of dose creep seen with analgesics as well as perceptions of a safety advantage over less selective COX inhibitors***

3. ***Although not approved for use in RA, the comparative safety data may be generalizable to other populations at similar risk for NSAID toxicity.***

4. ***The exclusion of subjects with significant active cardiovascular disease represents a serious limitation of the current study. Patients with RA may be at higher risk of these conditions. Safety conclusions will be limited to those without these conditions.***

5. ***The exclusion of even low dose aspirin users seriously limits the generalizability of this study to an important segment of the population.***

Results

Demographics

Sponsor tables 16,17 and 19 display demographic data.

The groups were evenly divided for relevant factors. The demographic composite indicates that:

1. *80% of subjects were between the ages of 40 and 65*

2. *80% of subjects were female*

3. *20-25% of subjects were ARA III and 2% were ARA IV*

4. *Less than 8% of subjects had a history of symptomatic ulcers and approximately 2% had a history of UGI perforation, obstruction or bleed.*

5. *Over 50% of subjects were taking steroids or DMARDs at the time of entry*

Table 16

Baseline Patient Characteristics by Treatment Group (Continuous Variables)

Treatment Group	N	Mean (SD)		Median	Range
Age (Years)					
Rofecoxib	4047	58.0	(9.5)	58.0	34.0 to 88.0
Naproxen	4029	58.2	(9.6)	58.0	37.0 to 89.0
Total	8076	58.1	(9.5)	58.0	34.0 to 89.0
Weight (kg)					
Rofecoxib	4045	72.2	(17.7)	69.5	31.0 to 193.2
Naproxen	4027	71.9	(17.0)	69.7	35.0 to 150.6
Total	8072	72.1	(17.3)	69.6	31.0 to 193.2
Height (cm)					
Rofecoxib	4026	161.8	(10.2)	161.0	115.0 to 203.2
Naproxen	4010	161.8	(10.0)	161.0	126.0 to 195.6
Total	8036	161.8	(10.1)	161.0	115.0 to 203.2
Duration of Rheumatoid Arthritis (Years)					
Rofecoxib	4043	10.9	(9.6)	8.0	0.0 to 69.0
Naproxen	4024	10.7	(9.4)	8.0	0.0 to 61.0
Total	8067	10.8	(9.5)	8.0	0.0 to 69.0

Data Source: [4.6; 4.10]

Table 17

Baseline Patient Characteristics by Treatment Group (Categorical Variables)

Baseline Demographics	Rofecoxib (N=4047)		Naproxen (N=4029)		Total (N=8076)	
	n	(%)	n	(%)	n	(%)
Gender						
Female	3223	(79.6)	3215	(79.8)	6438	(79.7)
Male	824	(20.4)	814	(20.2)	1638	(20.3)
Ethnic Group						
White	2761	(68.2)	2750	(68.3)	5511	(68.2)
Black	207	(5.1)	202	(5.0)	409	(5.1)
Asian	101	(2.5)	85	(2.1)	186	(2.3)
Hispanic	501	(12.4)	516	(12.8)	1017	(12.6)
Multi-racial	464	(11.5)	466	(11.6)	930	(11.5)
Other	13	(0.3)	10	(0.2)	23	(0.3)
Study Region						
U.S.	1748	(43.2)	1750	(43.4)	3498	(43.3)
Multinational	2299	(56.8)	2279	(56.6)	4578	(56.7)
Age Group						
<40	10	(0.2)	11	(0.3)	21	(0.3)
40-54	1521	(37.6)	1527	(37.9)	3048	(37.7)
55-64	1519	(37.5)	1421	(35.3)	2940	(36.4)
65-74	800	(19.8)	857	(21.3)	1657	(20.5)
75+	197	(4.9)	213	(5.3)	410	(5.1)
Prior History of PUBs						
Yes	314	(7.8)	316	(7.8)	630	(7.8)
No	3733	(92.2)	3713	(92.2)	7446	(92.2)
ARA Status [1.1.12]						
I	881	(21.8)	830	(20.6)	1711	(21.2)
II	2160	(53.4)	2199	(54.6)	4359	(54.0)
III	928	(22.9)	932	(23.1)	1860	(23.0)
IV	78	(1.9)	68	(1.7)	146	(1.8)

Table 17 (Cont.)

Baseline Patient Characteristics by Treatment Group (Categorical Variables)

Baseline Demographics	Rofecoxib (N=4047)		Naproxen (N=4029)		Total (N=8076)	
	n	(%)	n	(%)	n	(%)
Treatment for Rheumatoid Arthritis at Study Entry						
Corticosteroids	2260	(55.8)	2263	(56.2)	4523	(56.0)
Methotrexate	2263	(55.9)	2269	(56.3)	4532	(56.1)
Other DMARDs†	1847	(45.6)	1826	(45.3)	3673	(45.5)
History of Cardiac Disease						
Yes	1884	(46.6)	1838	(45.6)	3722	(46.1)
No	2163	(53.4)	2191	(54.4)	4354	(53.9)
Smoking Status						
Unknown	1	(0.0)	0	(0.0)	1	(0.0)
Never Smoked	2128	(52.6)	2150	(53.4)	4278	(53.0)
Ex-Smoker	1128	(27.9)	1100	(27.3)	2228	(27.6)
Current Smoker	790	(19.5)	779	(19.3)	1569	(19.4)
Number Cigarettes/24 Hours						
<11/day	404	(51.1)	409	(52.5)	813	(51.8)
11 to 20/day	271	(34.3)	252	(32.3)	523	(33.3)
>20/day	115	(14.6)	118	(15.1)	233	(14.9)
Number of Alcoholic Drinks Per Week						
Unknown	0	(0.0)	1	(0.0)	1	(0.0)
None	2994	(74.0)	2984	(74.1)	5978	(74.0)
1 to 4	866	(21.4)	864	(21.4)	1730	(21.4)
5 to 7	101	(2.5)	88	(2.2)	189	(2.3)
8 to 10	54	(1.3)	56	(1.4)	110	(1.4)
>10	32	(0.8)	36	(0.9)	68	(0.8)

Table 19

Number (%) of Patients With Secondary Diagnoses of Clinical Upper GI Events (PUBs) (Gastroduodenal Perforation, Obstruction, Ulcer, and Upper Gastrointestinal Bleed)

	Rofecoxib (N=4047)	Naproxen (N=4029)	Total (N=8076)
	n (%)	n (%)	n (%)
Prior History of Clinical Upper GI Event (PUB) (Total)	**314 (7.8)**	**316 (7.8)**	**630 (7.8)**
Gastroduodenal Ulcer	**287 (7.1)**	**289 (7.2)**	**576 (7.1)**
Gastric Ulcer	163 (4.0)	157 (3.9)	320 (4.0)
Duodenal Ulcer	104 (2.6)	110 (2.7)	214 (2.7)
Peptic Ulcer†	29 (0.7)	37 (0.9)	66 (0.8)
Associated Perforation	18 (0.4)	17 (0.4)	35 (0.4)
Associated Obstruction	2 (0.1)	6 (0.2)	8 (0.1)
Upper GI Bleed	**91 (2.3)**	**93 (2.3)**	**184 (2.3)**

† Exact location unknown.
Patients may appear in more than 1 row but are counted only once in each relevant row.

Data Source: [4.15]

GI outcome results

Database audit*:*

Approximately 50% of endpoint packages were reviewed. Adjudication appeared to be consistently applied.

Endpoint results:

Table 12 displays the extent of exposure in the study. Figures 4 and 5 also indicate the exposure data as subjects at risk for each 2-month interval.

Sponsor figures 3,4 and 5 tables 22, 23, 24, 26, 31 and 32 display the most relevant GI results.

These results suggest that the endpoint PUB reasonably predicts the ***relative risk*** *of true complicated events (POB). The* ***actual event rate*** *tables indicate that rates of POBs and PUBs are quite different and should not be intermingled when discussing the* ***absolute*** *risks of serious/lifethreatening events versus symptomatic adverse events. The cumulative risk of PUBs was 1.8 and 3.9 for the V and naproxen respectively. The POB rates were 0.4 and 0.9 respectively for a study with a mean exposure of 8 months****. The rate for PUBs in both groups is quite similar to the range for the 3-6 months and one-year exposures that appears in the current GI warning section of NSAID labels (1% at 3-6 months and 2 to 4 percent at one year).***

The current study suggests that potentially life-threatening events (POBs) make up a fraction of the total UGI events associated with these products. GI safety must be assessed within the overall safety profile of a drug. As discussed in the background section, labeling a selective GI safety advantage in the absence of a commensurate or improved overall safety profile compared to other products in the same class may give a false impression to consumers. Thus safety profiles must be carefully analyzed based on events of comparable severity and seriousness. The reader is referred to the general safety review by Dr. Villalba.

The relative rates for the first month appear comparable for the PUB and POB endpoints. The event rates begin to diverge after 1 month. This may be due to the small number of events however, short term use does not appear to be associated with an advantage in UGI safety in the V group in this study. Results for NSAID naive subjects may be different in short-term use.

The current study is consistent with prior studies that suggest gastric ulcers (as opposed to duodenal ulcers) represent the vast majority of UGI events.

Table 12

Time in Study†

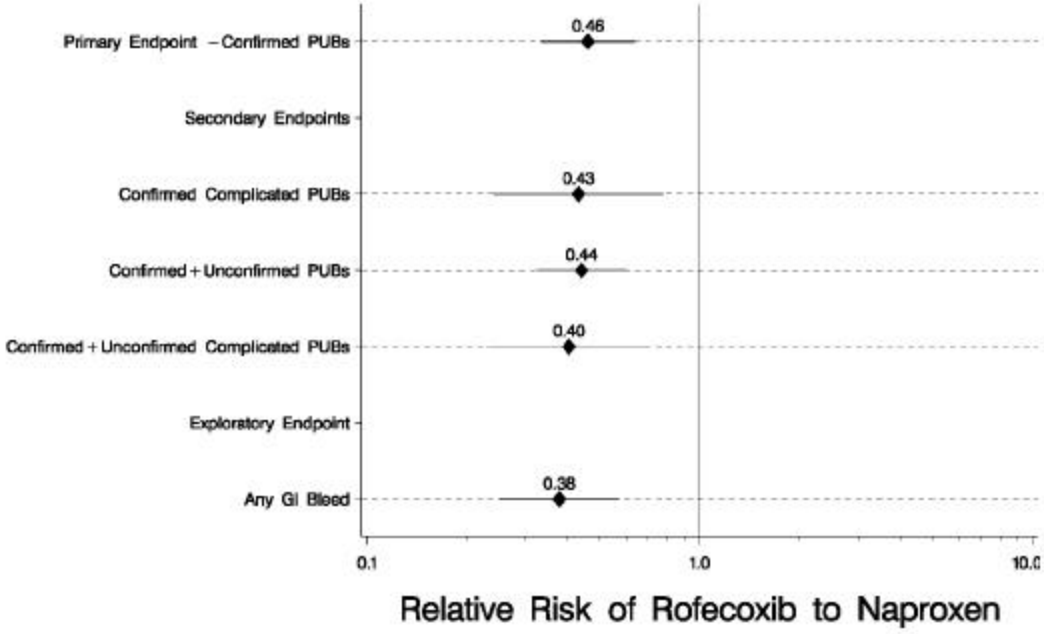
Cohort	Treatment Group	Duration of Follow-Up (Months)					
		N	Mean	SD	Median	Range	Inter-Quartile Range
Overall	Rofecoxib	4047	8.0	3.1	9.0	0.5 to 13.0	7.5 to 10.1
	Naproxen	4029	8.0	3.1	9.0	0.5 to 12.7	7.6 to 10.1
	Total	8076	8.0	3.1	9.0	0.5 to 13.0	7.6 to 10.1
U.S.	Rofecoxib	1748	7.5	3.6	8.5	0.5 to 13.0	4.4 to 10.3
	Naproxen	1750	7.5	3.5	8.5	0.5 to 12.7	4.4 to 10.3
	Total	3498	7.5	3.6	8.5	0.5 to 13.0	4.4 to 10.3
Multi-national	Rofecoxib	2299	8.4	2.7	9.2	0.5 to 12.3	8.0 to 10.0
	Naproxen	2279	8.4	2.6	9.2	0.5 to 12.2	8.1 to 10.0
	Total	4578	8.4	2.7	9.2	0.5 to 12.3	8.0 to 10.0

† Up to 14 days past discontinuation.

Data Source: [4.8]

Figure 3

Relative Risk of Rofecoxib to Naproxen With 95% CI† Primary, Secondary, and Exploratory GI Endpoints

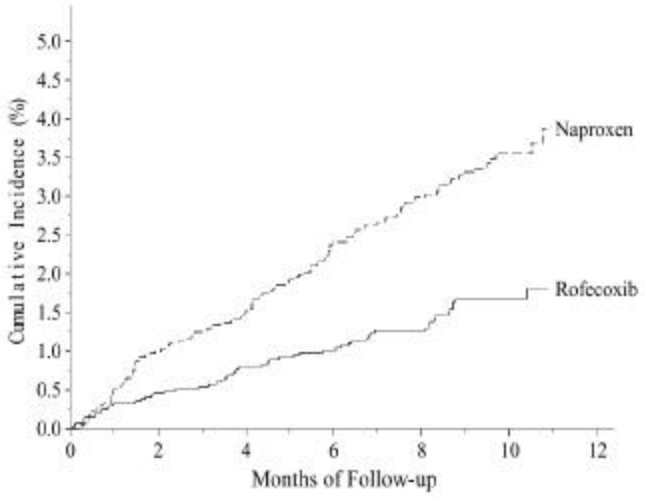


†95.44% CI for Primary Endpoint (adjusted for interim analysis)

Data Source: [4.3; 4.15]

Figure 4

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

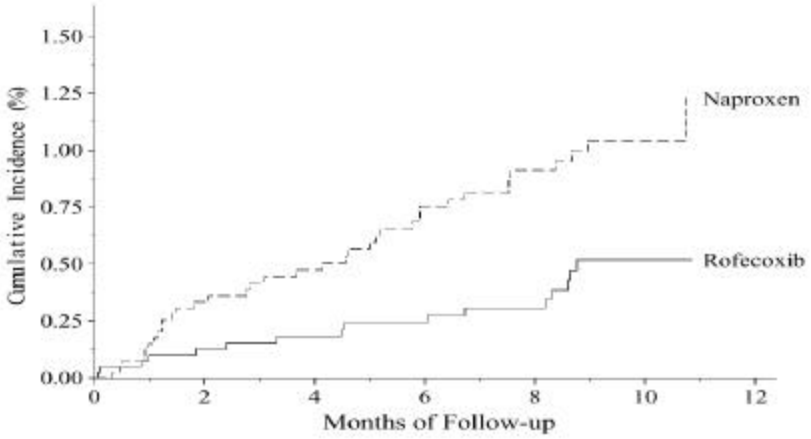


# at Risk							
Rofecoxib	n=4047	3641	3402	3180	2806	1073	533
Naproxen	n=4029	3644	3389	3163	2796	1071	513

Data Source: [4.8; 4.15]

Figure 5

Confirmed Complicated PUBs
Time-to-Event Plot (All-Patients-Randomized)



# at Risk							
Rofecoxib	n=4047	3644	3407	3181	2806	1072	533
Naproxen	n=4029	3646	3394	3170	2800	1074	514

Data Source: [4.8; 4.15]

Table 22

Summary of Gastrointestinal Safety Endpoints

Endpoint	Treatment Group	N	Number of Patients With Events	PYR[†]	Rates[‡]	Proportionality Assumption p-Value	Relative Risk[§]		
							Estimate	95% CI[‖]	p-Value
Primary-confirmed PUBs	Rofecoxib	4047	56	2697	2.08	0.378	0.46	(0.33, 0.64)	<0.001
	Naproxen	4029	121	2694	4.49				
Secondary Endpoints									
Confirmed, complicated PUBs	Rofecoxib	4047	16	2699	0.59	0.648	0.43	(0.24, 0.78)	0.005
	Naproxen	4029	37	2698	1.37				
Confirmed and unconfirmed PUBs	Rofecoxib	4047	58	2697	2.15	0.471	0.44	(0.32, 0.60)	<0.001
	Naproxen	4029	132	2693	4.90				
Confirmed and unconfirmed complicated PUBs	Rofecoxib	4047	17	2699	0.63	0.739	0.40	(0.23, 0.71)	0.002
	Naproxen	4029	42	2697	1.56				
Exploratory Endpoint									
Any GI bleed	Rofecoxib	4047	31	2698	1.15	0.707	0.38	(0.25, 0.57)	<0.001
	Naproxen	4029	82	2694	3.04				

[†] Patient-years at risk.
[‡] Per 100 PYR.
[§] Relative risk of rofecoxib with respect to naproxen from Cox model stratified by prior history of PUBs (and study region for confirmed and unconfirmed PUBs).
[‖] 95.44% CI for primary endpoint.

Data Source: [4.3; 4.6; 4.15]

Table 23

Analysis of Confirmed PUBs
All-Patients-Randomized

	Rofecoxib (N=4047)	Naproxen (N=4029)
Patients with events	56	121
Patient-years at risk	2697	2694
Rate†	2.08	4.49
Cumulative incidence‡	1.80	3.87
Relative Risk§		
Estimate	0.46	
95.44% CI	(0.33, 0.64)	
p-Value	<0.001	

† Per 100 patient-years at risk.
‡ At end of study but while at least 500 patients are at risk in each treatment group. Note: Cumulative incidence is from the survival analysis, it may not equal (number patients with events/N) x 100.
§ Of rofecoxib to naproxen from Cox model stratified by prior history of PUBs.
Proportional hazard assumption was met: p-value=0.378.
Treatment-by-PUB history not significant: p-value=0.874.

Data Source: [4.15]

Table 24

Number (%) of Types of Confirmed PUBs (Primary Endpoints)

Primary Endpoint	Rofecoxib (N=4047) n (%)	Naproxen (N=4029) n (%)
Confirmed PUBs	**56 (1.38)**	**121 (3.00)**
Gastroduodenal perforations	3 (0.07)	4 (0.10)
Gastric ulcers	28 (0.69)	81 (2.01)
Duodenal ulcers	27 (0.67)	39 (0.97)
Gastric outlet obstructions	1 (0.02)	0 (0.00)
Upper GI bleeds	14 (0.35)	35 (0.87)

Patients may be counted in more than 1 row, but are only counted once within a row.

Data Source: [4.15]

Table 26

Number (%) of Types of Confirmed, Complicated PUBs

Secondary Endpoint	Rofecoxib (N=4047) n (%)	Naproxen (N=4029) n (%)
Confirmed Complicated PUBs	**16 (0.40)**	**37 (0.92)**
Gastroduodenal perforations	3 (0.07)	4 (0.10)
Gastric ulcers	1 (0.02)	6 (0.15)
Duodenal ulcer	3 (0.07)	5 (0.12)
Gastric outlet obstructions	1 (0.02)	0 (0.00)
Upper GI bleeds	12 (0.30)	32 (0.79)

Patients may be counted in more than 1 row, but are only counted once within a row.

Data Source: [4.15]

Table 31

Analysis of Any GI Bleed
All-Patients-Randomized

	Rofecoxib (N=4047)	Naproxen (N=4029)
Patients with events	31	82
Patient-years at risk	2698	2694
Rate†	1.15	3.04
Cumulative incidence‡	1.00	2.59
Relative Risk§		
Estimate	0.38	
95% CI	(0.25, 0.57)	
p-Value	<0.001	

† Per 100 patient-years at risk.

‡ At end of study but while at least 500 patients are at risk in each treatment group. Note: Cumulative incidence is from the survival analysis, it may not equal (number patients with events/N) x 100.

§ Of rofecoxib to naproxen from Cox model stratified by prior history of PUBs.

Proportional hazard assumption was met: p-value=0.707.

Treatment-by-PUB history not significant: p-value=0.244.

Data Source: [4.3; 4.15]

Table 32

Number (%) of Types of GI Bleeds

Exploratory Endpoint	Rofecoxib (N=4047) n (%)	Naproxen (N=4029) n (%)
Any GI Bleed	**31 (0.77)**	**82 (2.04)**
Upper GI bleed	21 (0.52)	58 (1.44)
Lower GI bleed†	11 (0.27)	24 (0.60)

† Lower GI bleeds include all GI bleeds that were not of esophageal, gastric or duodenal origin.
Patients may be counted in more than 1 row, but are only counted once within a row.

Data Source: [4.3]

Subgroup analysis

Figure 11 and tables 12.3.1, 12.3.2, 12.3.3, 12.3.5, 12.3.6, 12.3.7, 12.3.8, 12.3.9, 12.3.10 and 12.3.11 display the result of subgroup analyses. The trend for lower relative risk is maintained in the subgroups displayed.

Findings of interest:

1. The absolute risk of PUBs appeared to be meaningfully higher in the subpopulation **not** taking NSAIDs at baseline compared to the group not recently on NSAIDs. These data are displayed in table 12.3.10. This finding is consistent with the concept of "gastric adaptation" that has been identified in short-term endoscopic studies in the past. The current data represent validation of this concept as a clinically relevant phenomenon. This finding is consistent with the concept of falling risk with continued exposure of NSAIDs. The current database does not show a meaningful fall in risk over time. The CLASS study recently reviewed did show a fall in the risk over time in the risk of PUBs and POBs in the traditional NSAID groups (ibuprofen and diclofenac) but not in the celecoxib group. These two large databases do not offer a consistent picture of risk over time for NSAIDs, selective or less selective.

2. Subjects with a prior history of PUB have a four-fold increase in risk of PUBs in the current study. The relative risk associated with the use of V compared to naproxen is maintained in this subpopulation (0.44). The **absolute** risk in the V group with a history of a PUB was nearly double the rate in the naproxen group that did **not** have a history of PUB. The relative risk of using naproxen in a patient with a prior history of

a PUB is over 15 fold. This finding would argue strongly against the use of naproxen in patients with a history of a PUB. This is generally accepted in clinical practice. However, the absolute UGI safety of V for patients with a history of PUB appears to be less than that of naproxen in the general population. While affording a GI safety advantage in this high risk group compared to naproxen, caution should be advised to the use of V in this population based on absolute risk rates for GI events.

3. Patients over the age of 65 have 2.5-fold higher risk than younger patients. The relative risk associated with the use of V compared to naproxen is maintained in this subpopulation (0.41). The absolute risk however in the V group **over** the age of 65 was nearly the same as the rate in the naproxen group that was **under** the age of 65.

4. The use of steroids appears to be risk factor that does not affect risk associated with the use of V but that does increase the risk associated with the use of naproxen. The strength of this finding in a post-hoc analysis is unclear.

5. A history of ASCVD had little effect on the relative risk associated with the use of naproxen. A history of ASCVD appeared to have a "protective" effect for subjects in the V group. A biologically plausible reason for this is finding is not apparent. The meaning of this finding is unclear.

The subanalysis based on "aspirin use-FDA Rules" that appears in table 12.3.9 is of unclear value. The protocol itself spoke to the issue of aspirin indicated subjects. The investigators were to use their judgement in excluding subjects that had an indication for prophylactic aspirin use. In addition, subjects with a history of significant active angina, history of MI or coronary artery bypass graft surgery within the prior year or recent stroke or TIA were explicitly excluded. It would appear that clinical decision making by investigators as to the candidacy of an individual for prophylactic aspirin would have greater validity than a post hoc chart-review based assessment.

Table 12.3-1

Confirmed PUBs
Analysis of Treatment by Prior History of PUBs Interaction
All-Patients-Randomized

	No Prior History of PUBs		Prior History of PUBs	
	Rofecoxib	Naproxen	Rofecoxib	Naproxen
N	3733	3713	314	316
Patients with events	43	92	13	29
Patient-years at risk	2504	2505	194	189
Rate†	1.72	3.67	6.72	15.33
Relative Risk‡				
Estimate	0.47		0.44	
95% CI	(0.33, 0.67)		(0.23, 0.85)	
Model Effects: Prior History vs. No Prior History (Combined Treatments)§				
Estimate	4.05			
95% CI	(2.87, 5.73)			
Model Effects: Treatment by Subgroup Interaction‡				
p-Value	0.874			

† Per 100 patient-years at risk.
‡ Cox model includes treatment, subgroup, and treatment by subgroup interaction.
§ Cox model includes treatment and subgroup main effect.

Data Source: [4.15]

Table 12.3-2

Confirmed PUBs
Analysis of Treatment by Study Region Interaction
All-Patients-Randomized

	U.S.		Non-U.S.	
	Rofecoxib	Naproxen	Rofecoxib	Naproxen
N	1748	1750	2299	2279
Patients with events	18	42	38	79
Patient-years at risk	1096	1093	1601	1601
Rate†	1.64	3.84	2.37	4.93
Relative Risk‡				
Estimate	0.43		0.48	
95% CI	(0.25, 0.74)		(0.33, 0.71)	
Model Effects: Non-U.S. vs. U.S. (Combined Treatments)§				
Estimate	1.35			
95% CI	(0.98, 1.84)			
Model Effects: Treatment by Subgroup Interaction‡				
p-Value	0.731			

† Per 100 patient-years at risk.
‡ Cox model includes treatment, subgroup, and treatment by subgroup interaction.
§ Cox model includes treatment and subgroup main effect.

Data Source: [4.6; 4.15]

Table 12.3-3

Confirmed PUBs
Analysis of Treatment by Age Group Interaction
All-Patients-Randomized

	Non-Elderly (<65 Years)		Elderly (≥65 Years)	
	Rofecoxib	Naproxen	Rofecoxib	Naproxen
N	3050	2959	997	1070
Patients with events	34	64	22	57
Patient-years at risk	2076	2034	622	660
Rate†	1.64	3.15	3.54	8.63
Relative Risk‡				
Estimate	0.52		0.41	
95% CI	(0.34, 0.79)		(0.25, 0.67)	
Model Effects: Elderly vs. Non-Elderly (Combined Treatments)§				
Estimate	2.53			
95% CI	(1.88, 3.40)			
Model Effects: Treatment by Subgroup Interaction‡				
p-Value	0.466			

† Per 100 patient-years at risk.
‡ Cox model includes treatment, subgroup, and treatment by subgroup interaction.
§ Cox model includes treatment and subgroup main effect.

Data Source: [4.6; 4.15]

Table 12.3-5

Confirmed PUBs
Analysis of Treatment by Gender Interaction
All-Patients-Randomized

	Female		Male	
	Rofecoxib	Naproxen	Rofecoxib	Naproxen
N	3223	3215	824	814
Patients with events	45	98	11	23
Patient-years at risk	2149	2139	549	555
Rate†	2.09	4.58	2.01	4.14
Relative Risk‡				
Estimate	0.46		0.48	
95% CI	(0.32, 0.65)		(0.24, 0.99)	
Model Effects: Male vs. Female (Combined Treatments)§				
Estimate	0.92			
95% CI	(0.63, 1.34)			
Model Effects: Treatment by Subgroup Interaction‡				
p-Value	0.892			

† Per 100 patient-years at risk.
‡ Cox model includes treatment, subgroup, and treatment by subgroup interaction.
§ Cox model includes treatment and subgroup main effect.

Data Source: [4.6; 4.15]

Table 12.3-6

Confirmed PUBs
Analysis of Treatment by Baseline Steroid Use Interaction
All-Patients-Randomized

	No Baseline Steroid Use		Baseline Steroid Use	
	Rofecoxib	Naproxen	Rofecoxib	Naproxen
N	1803	1776	2244	2253
Patients with events	24	35	32	86
Patient-years at risk	1184	1178	1513	1516
Rate†	2.03	2.97	2.11	5.67
Relative Risk‡				
Estimate	0.68		0.37	
95% CI	(0.41, 1.15)		(0.25, 0.56)	
Model Effects: Baseline Steroid Use vs. No Baseline Steroid Use (Combined Treatments)§				
Estimate	1.56			
95% CI	(1.14, 2.14)			
Model Effects: Treatment by Subgroup Interaction‡				
p-Value	0.073			

† Per 100 patient-years at risk.
‡ Cox model includes treatment, subgroup, and treatment by subgroup interaction.
§ Cox model includes treatment and subgroup main effect.

Data Source: [4.5; 4.15]

Table 12.3-7

Confirmed PUBs
Analysis of Treatment by H. Pylori‖ Interaction
All-Patients-Randomized

	Negative H. Pylori		Positive H. Pylori	
	Rofecoxib	Naproxen	Rofecoxib	Naproxen
N	2244	2260	1740	1712
Patients with events	21	67	34	54
Patient-years at risk	1470	1486	1186	1170
Rate†	1.43	4.51	2.87	4.62
Relative Risk‡				
Estimate	0.32		0.62	
95% CI	(0.19, 0.52)		(0.40, 0.95)	
Model Effects: Positive vs. Negative H. Pylori (Combined Treatments)§				
Estimate	1.27			
95% CI	(0.94, 1.70)			
Model Effects: Treatment by Subgroup Interaction‡				
p-Value	0.043			

† Per 100 patient-years at risk.
‡ Cox model includes treatment, subgroup, and treatment by subgroup interaction.
§ Cox model includes treatment and subgroup main effect.
‖ Serology was measured by the HM-CAP method. Values ≤100 were considered negative.

Data Source: [4.9; 4.15]

Table 12.3-8

Confirmed PUBs
Analysis of Treatment by ASCVD History
All-Patients-Randomized

	No History ASCVD		ASCVD History	
	Rofecoxib	Naproxen	Rofecoxib	Naproxen
N	3809	3813	238	216
Patients with events	55	114	1	7
Patient-years at risk	2550	2555	148	139
Rate†	2.16	4.46	0.68	5.04
Relative Risk‡				
Estimate	0.48		0.13	
95% CI	(0.35, 0.67)		(0.02, 1.09)	
Model Effects: ASCVD History vs. No ASCVD History (Combined Treatments)§				
Estimate	0.85			
95% CI	(0.42, 1.73)			
Model Effects: Treatment by Subgroup Interaction‡				
p-Value	0.235			

† Per 100 patient-years at risk.
‡ Cox model includes treatment, subgroup, and treatment by subgroup interaction.
§ Cox model includes treatment and subgroup main effect.

Data Source: [4.10; 4.15]

Table 12.3-9

Confirmed PUBs
Analysis of Treatment by Indication for Aspirin Use—FDA Rules
All-Patients-Randomized

	Not Indicated for Cardioprotective Aspirin—FDA Rules		Indicated for Cardioprotective Aspirin—FDA Rules	
	Rofecoxib	Naproxen	Rofecoxib	Naproxen
N	3877	3878	170	151
Patients with events	55	116	1	5
Patient-years at risk	2592	2592	106	102
Rate†	2.12	4.47	0.95	4.91
Relative Risk‡				
Estimate	0.47		0.19	
95% CI	(0.34, 0.65)		(0.02, 1.64)	
Model Effects: Indicated for Cardioprotective Aspirin vs. Not (Combined Treatments)§				
Estimate	0.88			
95% CI	(0.39, 1.99)			
Model Effects: Treatment by Subgroup Interaction‡				
p-Value	0.412			

† Per 100 patient-years at risk.
‡ Cox model includes treatment, subgroup, and treatment by subgroup interaction.
§ Cox model includes treatment and subgroup main effect.

Data Source: [4.10; 4.15]

Table 12.3-10

Confirmed PUBs
Analysis of Treatment by Baseline NSAID Use Interaction
All-Patients-Randomized

	No Baseline NSAID Use		Baseline NSAID Use	
	Rofecoxib	Naproxen	Rofecoxib	Naproxen
N	703	688	3344	3341
Patients with events	14	33	42	88
Patient-years at risk	455	435	2242	2260
Rate†	3.07	7.59	1.87	3.89
Relative Risk‡				
Estimate	0.41		0.48	
95% CI	(0.22, 0.76)		(0.33, 0.69)	
Model Effects: Baseline NSAID Use vs. No Baseline NSAID Use (Combined Treatments)§				
Estimate	0.54			
95% CI	(0.39, 0.76)			
Model Effects: Treatment by Subgroup Interaction‡				
p-Value	0.645			

† Per 100 patient-years at risk.
‡ Cox model includes treatment, subgroup, and treatment by subgroup interaction.
§ Cox model includes treatment and subgroup main effect.

Data Source: [4.5; 4.15]

Table 12.3-11

Confirmed Complicated PUBs
Analysis of Treatment by Study Region Interaction
All-Patients-Randomized

	U.S.		Non-U.S.	
	Rofecoxib	Naproxen	Rofecoxib	Naproxen
N	1748	1750	2299	2279
Patients with events	10	21	6	16
Patient-years at risk	1096	1094	1603	1604
Rate†	0.91	1.92	0.37	1.00
Relative Risk‡				
Estimate	0.48		0.38	
95% CI	(0.22, 1.01)		(0.15, 0.96)	
Model Effects: Baseline NSAID Use vs. No Baseline NSAID Use (Combined Treatments)§				
Estimate	0.49			
95% CI	(0.28, 0.84)			
Model Effects: Treatment by Subgroup Interaction‡				
p-Value	0.700			

† Per 100 patient-years at risk.
‡ Cox model includes treatment, subgroup, and treatment by subgroup interaction.
§ Cox model includes treatment and subgroup main effect.

Data Source: [4.6; 4.15]

Overall conclusions:

1. ***The sponsor has demonstrated a statistically significant reduction associated with the use of V compared to naproxen at the endpoints of symptomatic, bleeding obstructing and perforating UGI ulcers (PUBs) as well as serious events including bleeding, obstruction and perforation (POBs).***

2. ***Absolute risk assessment requires separating out PUBs and POBs to allow for a meaningful analysis of overall risk based on seriousness of risk. A comparison of overall safety requires assessment of the entire database of adverse events. The medical officer's review by Dr. Villalba addresses this issue.***

3. ***The relative risk reduction associated with the use of V compared to naproxen is maintained in all important subgroups. The absolute risk associated with V in high risk subjects (elderly, prior PUB history, steroid use) remains in the range suggested in the current GI warning template. Patients with poor overall health status were to a great extent excluded based on the composite of multiple exclusion criteria in the current study. The relative and absolute UGI safety cannot be extrapolated to this population. Prescribing physicians as well as patients should be aware of these facts in the context of any proposed labeling change based on the current study.***

4. ***The substantial differences in absolute rates of PUBs in different subgroups based on concomitant disease, medication, study location and age highlight the difficulties and risks in extrapolating absolute risk rates beyond the population studied. A study with even slight differences in inclusion criteria, exclusion criteria and population enrolled would likely result in substantially different absolute rates, although the relative rates to comparators may be maintained. There will be a strong incentive to cross-compare to other large outcome studies or controlled databases. Such an approach is to be discouraged and may be highly misleading.***

Appendix 1

Relevant protocol provisions and amendments

4. STUDY HYPOTHESES AND OBJECTIVES

4.1 Primary Hypotheses

1. The risk of confirmed PUBs (gastroduodenal perforations, symptomatic ulcers, gastric outlet obstructions, or upper-GI bleeds) during the treatment period will be reduced in the group of patients with rheumatoid arthritis taking 50 mg rofecoxib daily, compared to the group of patients with rheumatoid arthritis taking naproxen 1000 mg daily [3.2].

2. Rofecoxib administered at a dose of 50 mg daily will be safe and well tolerated.

4.2 Secondary Hypothesis

1. The risk of confirmed and unconfirmed PUBs during the treatment period will be reduced in the group of patients with rheumatoid arthritis taking 50 mg rofecoxib daily compared to the group of patients with rheumatoid arthritis taking naproxen 1000 mg daily.

4.2 Secondary Hypothesis (Cont.)

2. The risk of confirmed complicated PUBs during the treatment period will be reduced in the group of patients with rheumatoid arthritis taking 50 mg rofecoxib daily compared to the group of patients with rheumatoid arthritis taking naproxen 1000 mg daily.

4.3 Primary Objectives

1. To determine the relative risk of confirmed PUBs in patients with rheumatoid arthritis taking 50 mg rofecoxib daily compared to patients taking naproxen, 1000 mg daily.

2. To study the safety and tolerability of rofecoxib in patients with rheumatoid arthritis.

4.4 Secondary Objectives

1. To assess the relative risk of confirmed and unconfirmed PUBs in patients with rheumatoid arthritis taking 50 mg rofecoxib daily compared to patients taking naproxen 1000 mg daily.

2. To assess the relative risk of complicated PUBs in patients with rheumatoid arthritis taking 50 mg rofecoxib daily compared to patients taking naproxen 1000 mg daily.

3. To assess the efficacy of treatment of rheumatoid arthritis with rofecoxib or naproxen as evaluated by the Patient and Investigator Global Assessment of Disease Activity and the discontinuation rates due to lack of efficacy.

4.5 Exploratory Objectives

1. To determine relative risk of occurrence of bleeding from any location in the GI tract in patients taking 50 mg rofecoxib daily, compared to patients taking naproxen 1000 mg daily.

2. To assess the efficacy of treatment of rheumatoid arthritis with 50 mg rofecoxib daily compared to patients taking naproxen 1000 mg daily as evaluated by the modified Health Assessment Questionnaire.

5. INVESTIGATIONAL PLAN

5.1 Overall Study Design and Plan

This active-comparator-controlled, parallel-group, stratified, double-blind, multicenter study was conducted under in-house blinding procedures to further evaluate the long-term safety of rofecoxib compared with naproxen. Patients with rheumatoid arthritis who met entry criteria were randomized to rofecoxib 50 mg once a day, or naproxen 500 mg twice a day. The primary endpoint of this study was the occurrence of PUBs and a key secondary endpoint was the occurrence of complicated PUBs (see Section 5.5.1.5 for detailed endpoint definitions). All events identified by the investigators as potential PUBs were adjudicated by a blinded Case Review Committee that had final say on the classification of all such events as described in detail below in Section 5.5.1.4. The study was to terminate after a minimum of 120 PUBs and 40 complicated PUBs had been confirmed by the Case Review Committee and the study had run at least 6 months from the date of the last patient randomized, whichever came last.

5.2 Detailed Description of Study Design

Patients with a history of RA who were thought to require treatment with NSAIDs for at least 1 year were permitted to enter the study. After a minimum of a 3-day washout of NSAIDs, patients with RA who met the entry criteria were randomized to rofecoxib 50 mg once daily or naproxen 500 mg twice a day. There were no "flare" criteria for entry into this study. Allocation was stratified according to whether the patient had a prior history of a PUB due to the increased risk of experiencing a significant upper GI event associated with such a history. At the screening visit, patients were instructed that they were permitted to take acetaminophen or other analgesics (except for NSAIDs or aspirin) for rescue medication. Choice of rescue therapy was at the discretion of the investigator. Intra-articular, intramuscular, and oral steroids were permitted during the study. Topical creams or lotions containing NSAIDs or salicylates were not permitted during the study. Disease-Modifying Anti-Rheumatic Drugs (DMARDs) were allowed during the study. These could be initiated or the dose may have been changed at the discretion of the investigator during the course of the study. (Treatment with cyclosporin was not allowed during the study.)

The duration of the study was expected to be 1 to 1.5 years based on the time needed to observe at least 120 PUBs and 40 complicated PUBs.

5.2 Detailed Description of Study Design (Cont.)

Clinic visits occurred at screening, randomization, and Weeks 6, 17, 35, and 52. Thereafter, patients would have been seen approximately every 4 months (Weeks 69, 87, and 104) had the study continued, until the termination of the study. At the termination of the study, patients were called in for an end-of-study visit. Patients were asked to remain off NSAIDs for 14 days after the end-of-study visit. Every effort was made to bring all patients back for this visit.

Efforts were made to keep in contact with the patients between clinic visits. Patients were contacted by phone at Week 10 and then every 4 months thereafter (Weeks 26 and 43). The primary endpoint for this study was the occurrence of confirmed PUBs. The key secondary endpoint was confirmed complicated PUBs. Other clinical endpoints included: confirmed and unconfirmed PUBs, confirmed and unconfirmed complicated PUBs, bleeding from any location in the GI tract; discontinuation due to lack of efficacy, Patient Global Assessment of Disease Activity, Investigator Global Assessment of Disease Activity, and Modified Health Assessment Questionnaire. Safety was monitored by repeat clinical and laboratory assessments throughout treatment and adverse events monitored throughout treatment for 14 days following completion of treatment or last dose of study therapy for discontinued patients.

Patients were encouraged to remain in the study for the full duration of the study. However, if patients were unwilling to continue study medication, they were asked to return for a discontinuation visit within 48 hours.

Patients were asked to remain off NSAIDs for 14 days after the discontinuation visit. In addition, they were contacted by telephone 14 days after the last dose of medication to ascertain whether any adverse events had occurred. Lastly, they were telephoned 45 days after the last dose of study medication and at the conclusion of the study (i.e., when they were to have concluded the study had they continued) to ascertain whether a PUB had occurred. Every effort was made to remain in contact with these patients.

Indiscriminate use of low-dose H_2 blockers was to be avoided. Initiation of high-dose H_2 blockers (defined as any dose higher than ranitidine 75 mg twice a day [or 150 mg once daily], famotidine 10 mg twice a day [or 20 mg once daily], cimetidine 200 mg twice a day [or 400 mg once daily], nizatidine 75 mg twice a day [or 150 mg once daily]), proton-pump inhibitors, sucralfate, or misoprostol during the study required discontinuation of the patient from the study. In addition, the use of low-dose aspirin was prohibited in the study, since even low-dose aspirin can affect gastric mucosal integrity. Furthermore, patients who required the use of low-dose aspirin as cardioprotective prophylaxis were excluded from study entry.

5.2 Detailed Description of Study Design (Cont.)

Clinical suspicion of GI bleeding or other possible GI complications were investigated by the appropriate clinical procedures. Any gastric or duodenal ulcers, obstructions, or perforations detected during the workup of GI symptoms or suspected bleeding were classified in accordance with guidelines provided in the Endpoint Classification Document [3.2]. These patients were discontinued from the study. In addition to the standard 14-day follow-up phone call, patients with a possible study endpoint (PUB) were contacted 6 weeks after the occurrence of the event to collect health care resource utilization information (completion of HCCR form).

Events that were determined by the investigator to be possible study endpoints were reported to headquarters, and the investigator was asked to assemble an endpoint package. Endpoint packages were sent to the Case Review Committee for final adjudication.

5.2.1 NSAID Washout

After prestudy laboratory tests were verified to be within defined limits by the central laboratory and the investigator, patients that fulfilled all entry criteria and had signed an informed consent were contacted by telephone. NSAID users were instructed to discontinue their current NSAID regimen.

After an initial evaluation, patients were given 3 stool hemoccult cards. Patients were instructed to return to the clinical research center for repeat evaluation and randomization 3 days after discontinuing their NSAIDs. At the randomization visit and prior to randomization, the 3 stool hemoccult cards were collected and developed. If necessary, the randomization visit occurred 3 to 14 days following Visit 1.

5.3 Selection of the Study Population

5.3.1 Inclusion Criteria

1. Patient was male or female, was at least 50 years of age or was 40 to 49 years of age and was taking chronic oral corticosteroids, had a clinical diagnosis of rheumatoid arthritis, and was judged by the investigator to require chronic NSAID therapy for at least 1 year.

2. Female patients must have demonstrated a serum beta human chorionic gonadotropin (β-hCG) level consistent with a nongravid state at the prestudy visit and must have agreed to remain abstinent, use oral birth control pills or single-barrier contraception (such as: partner using condom or patient using diaphragm, contraceptive sponge, or intrauterine [IUD]) beginning at least 7 days prior to treatment and continuing at least 14 days after the end-of-study visit or a discontinuation visit. Women who were postmenopausal or status posthysterectomy or tubal ligation were exempt from this requirement. (Postmenopausal was defined as no menses for the previous 1 year. If cessation of menses was within 18 months, follicle-stimulating hormone [FSH] must have been documented as elevated into the postmenopausal range prestudy.)

3. Except rheumatoid arthritis, the patient was judged to be in general reasonable health, based on medical history, physical examination, and laboratory screening tests, enabling him or her to complete the trial without anticipated serious comorbid event.

4. Patient was able to understand and complete the study questionnaires.

5. Patient understood the study procedures and agreed to participate in the study by giving written informed consent.

5.3.2 Exclusion Criteria

1. Patient had a history of the following:

   Other inflammatory arthritis (e.g., systemic lupus, spondyloarthropathy, polymyalgia rheumatica). Note: Patients with a history of gout were allowed to

5.3 Selection of the Study Population (Cont.)

enroll into the study; however, additional NSAID therapy was not allowed for treatment of exacerbations during the course of the study. Patients with rheumatoid arthritis and secondary Sjögrens disease or fibromyalgia were permitted to enter the study.

2. The patient was, in the opinion of the investigator, mentally or legally incapacitated such that informed consent could not be obtained or the patient could not read or comprehend written material.

3. The patient had a history of any illness or had significant abnormalities on prestudy clinical or laboratory evaluation that, in the opinion of the investigator, contraindicated a 1- to 2-year course of therapy with a NSAID such as naproxen.

   Note: Patients with low hemoglobin values [3.2; 3.8] must have had a history of chronic anemia or at least 2 stable baseline values which were repeated at least 1 week apart. An algorithm for assessing out-of-range laboratory values was provided [3.2; 3.8].

4. The patient had a documented history of ulcer or upper GI bleeding within the recent past which was thought by the investigator to mandate that NSAID therapy be given with concurrent proton-pump inhibitors, misoprostol or other medications not allowed per study protocol.

5. Patient had a history of gastric, biliary, or small intestinal surgery that caused malabsorption.

6. The patient had evidence of impaired renal function, defined as estimated creatinine clearance ≤30 mL/min. (Creatinine clearance estimated as follows—Men: [140-age] x weight [kg]/[serum creatinine [mg/dL] x 72]; Women: [0.85] [140-age] weight [kg]/[serum creatinine [mg/dL] x 72).

7. The patient had angina or congestive heart failure with symptoms that occurred at rest or with minimal activity. (Note: patients with a history of myocardial infarction or coronary arterial bypass grafting more than 1 year prior to study start may have participated if they did not require any concomitant medication excluded in this protocol. However, if a patient developed unstable angina or a myocardial infarction during the study they must have discontinued from the study.)

8. The patient had uncontrolled hypertension (diastolic blood pressure >95 mm Hg, or systolic blood pressure >165 mm Hg).

5.3 Selection of the Study Population (Cont.)

9. The patient had a history of stroke or transient ischemic attack (TIA) within the previous 2 years. (Note: If a patient had a TIA or stroke during the study they must have been discontinued from the study.)

10. The patient had active hepatitis/hepatic disease.

11. Patient had a history of neoplastic disease (exceptions: (a) patients with adequately treated basal cell carcinoma or carcinoma in situ of the cervix, and (b) patients with other malignancies that had been successfully treated ≥5 years prior to screening, where in the judgment of both the investigator and treating physician, appropriate follow-up had revealed no evidence of recurrence from the time of treatment through the time of screening). Patients with a history of leukemia, lymphoma, or myeloproliferative disease were ineligible for the study regardless of the time since treatment.

12. Patient was currently a user (including "recreational use") of any illicit drugs, or had a history of drug or alcohol abuse within the past 5 years.

13. Patient was allergic to or had hypersensitivity (e.g., bronchoconstriction in association with nasal polyps) to aspirin, naproxen, and other NSAIDs. (Note: Patients with a history of a potential idiosyncratic allergic reaction [e.g., rash] to a single NSAID in the past but who tolerated at least 2 other NSAIDs without hypersensitivity reactions may have participated).

14. Patient had morbid obesity and demonstrated significant health problems stemming from their obesity.

15. Patient had a positive result for the fecal occult blood screening test at baseline.

16. Patient had a history of esophageal or gastric surgery. (Patients with a history of simple closure of a perforation greater than 3 months prior to the beginning of the study were allowed to be enrolled. In addition, patients with a history of a simple hiatal hernia repair may have been enrolled.)

17. Patient had a history of inflammatory bowel disease.

18. Patient had a history of a bleeding diathesis.

5.3 Selection of the Study Population (Cont.)

19. The patient was excluded from participation in the study if the following medications had been used:

    - Misoprostol or sucralfate within the past 1 month.

    - Recent sustained use (any period longer than 4 consecutive days during the month prior to study start) of H_2 blockers (e.g., cimetidine, ranitidine, famotidine, nizatidine), or a proton-pump inhibitor (e.g., omeprazole, lansoprazole) at prescription doses, or doses indicated for treatment of active gastroduodenal ulcers. (Note: Use of antacids or over-the-counter doses of ranitidine (75 mg twice a day or 150 mg once daily), famotidine [10 mg twice a day or 20 mg once daily], cimetidine [200 mg twice a day or 400 mg once daily], and nizatidine [75 mg twice a day or 150 mg once daily] prior to randomization were not grounds for exclusion.)

    - Ongoing cyclosporin A treatment.

20. Patients taking aspirin, even low-dose (325 mg or less, daily or every other day) or other antiplatelet agents (e.g., ticlopidine) may not have enrolled in the study. Patients were not to stop taking low-dose aspirin or ticlopidine in order to participate. Exceptions: Patients taking aspirin solely as treatment of their rheumatoid arthritis may have participated in this study after discontinuation of their aspirin during the washout period. Patients were also excluded if use of antiplatelet agents (other than NSAIDs) within the following year was anticipated.

21. Patients were excluded from the study if the following concomitant medications were required: warfarin, or heparin (or low molecular weight heparin). Patients taking digoxin or lithium were not excluded from the study; however, baseline serum drug levels should have been drawn at Visit 1 and should have been monitored in the first few weeks of the study since naproxen and other NSAIDs can increase blood levels of these medications.

22. Other chronic medications had not been used at a stable dosage for at least 2 weeks.

23. Patient had donated a unit of blood or plasma or participated in another clinical study with an investigational agent within the last 4 weeks. The patient could not have participated in any other clinical study with an investigational agent during the course of this study.

24. Patient had previously been enrolled in a rofecoxib clinical study. Note: Patients previously enrolled in a rofecoxib study and allocated to placebo may have participated in this study. Identification of treatment allocation in prior rofecoxib studies must have been verified by the Merck Monitor.

5.4.3 Method of Assigning Patients to Treatment Groups

Patient allocation was stratified according to whether the patient had prior history of a PUB. Within each allocation range, those patients who had a history of a PUB were entered into Stratum 1 and those who did not into Stratum 2.

Within each allocation schedule, patients in Stratum 1 were assigned allocation numbers (ANs) sequentially starting with the highest number at each site and proceeding sequentially to the lower numbers. Patients in Stratum 2 were assigned ANs sequentially starting with the lowest number at each site and proceeding sequentially to higher numbers.

5.4.4 Selection of Doses in the Study

The selection of dose for this study was based both upon the compilation of Phase III safety data in osteoarthritis [1.2.3; 1.2.7; 1.2.8; 1.2.11; 1.2.14 to 1.2.17 to 1.2.20; 1.2.28; 2.1.2] and RA pilot efficacy and dose-ranging studies [1.2.1; 2.1.1]. The Phase IIb RA dose-ranging study demonstrated that 25 and 50 mg were equally efficacious in the treatment of RA. Both doses were superior to placebo and were safe and well tolerated. In Part II of Protocol 068, both 25 and 50 mg demonstrated efficacy similar to naproxen 500 mg twice daily [2.2.1; 2.2.2]. The choice of rofecoxib 50 mg ensured that the safety of the drug was studied at a dose that is anticipated to be 2 times the maximum dose for the treatment of both osteoarthritis and rheumatoid arthritis. Comparing this dose of rofecoxib to the most commonly used dose of a standard nonselective NSAID would provide the most rigorous testing of the GI safety of rofecoxib.

Naproxen is a widely prescribed NSAID that is a dual COX-1/COX-2 inhibitor approved for the treatment of rheumatoid arthritis. In most countries worldwide, the recommended dose for the treatment of rheumatoid arthritis is 500 to 1000 mg per day. However, in some countries, the dose may be increased to a maximum of 1500 mg per day when a higher level of anti-inflammatory/analgesic activity is required. The dose chosen for naproxen in this study (500 mg 2 times a day) is the most commonly used dose for the treatment of rheumatoid arthritis and was anticipated to provide similar efficacy to rofecoxib 50 mg. Therefore, the safety profile was actually biased against rofecoxib; the dose of 50 mg is anticipated to be 2 times greater than the dose indicated for RA whereas the dose of naproxen used in this study was within the recommended dose range for the drug.

Table 4

Study Flow Chart

	Scheduled Office/Clinic Visits										
			Treatment Week							End of	Discon-
Visit ID:	Screening	Random-ization	6	17 (Month 4)	35 (Month 8)	52 (Month 12)	69[†] (Month 16)	87[†] (Month 20)	104[†] (Month 24)	Study Visit	tinuation Visit
Visit Number:	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	End	D
Review entry criteria and study procedures	X	X									
Obtain informed consent[††]	X										
Document medical history	X										
Stool hemoccult	X										
Review concomitant medications	X	X	X	X	X	X	X	X	X	X	X
Discontinue prestudy nonsteroidal anti-inflammatory drugs (NSAIDs)	X[‡]										
Vital signs and weight[#]	X	X	X	X	X	X	X	X	X	X	X
Complete physical examination	X									X	X
Document interim history/monitor for study endpoints/adverse experiences		X	X	X	X	X	X	X	X	X	X
Develop stool hemoccult cards		X									
Retrieve used study medication				X	X	X	X	X	X	X	X
Complete blood count (CBC); serum chemistry	X[§,¶]		X			X			X	X[§]	X[§]
Hemoglobin, hematocrit				X	X		X	X			
Serum for β-human chorionic gonadotropin (β-hCG)[‖]	X										
Urine for β-hCG[‖]			X	X	X	X	X	X	X	X	X
Urinalysis	X		X							X	X

Table 4 (Cont.)

Study Flow Chart

	Scheduled Office/Clinic Visits										
			Treatment Week							End of	Discon-
Visit ID:	Prestudy	Random-ization	6	17 (Month 4)	35 (Month 8)	52 (Month 12)	69† (Month 16)	87† (Month 20)	104† (Month 24)	Study Visit	tinuation Visit
Visit Number:	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	End	D
Serum for archive; plasma for archive		X		X		X		X		X	X
Electrocardiogram (EKG)	X										
Dispense study medication		X		X	X	X	X	X	X		
Count study medication tablets				X	X	X	X	X	X	X	X
Modified Health Assessment Questionnaire		X	X							X	X
Patient global assessment of disease activity	X	X	X	X		X		X	X	X	X
Investigator global assessment of disease activity		X	X	X		X		X	X	X	X
Distribute investigational study identification card and appointment reminder cards	X	X	X	X	X	X	X	X	X		

† If study still ongoing.
‡ Patients who were on NSAIDs were to be called after laboratory tests had been evaluated and were told to stop NSAIDs and to return for Visit 2 after a 3-day washout period.
§ Serum chemistry obtained following 8-hour fast.
‖ For premenopausal woman only.
¶ Baseline levels of digoxin and lithium should have been drawn for patients taking these medications.
# Obtained vital signs post 10 minutes of rest.
†† Permission should have been obtained to collect medical records and copies of endoscopy or radiographic reports should an endpoint occur.

Table 5

Phone Visits

Visit ID: Visit Number:	Pre-screening 0	Treatment Week Week 10 Month 2.5 P4	Week 26 Month 6 P5	Week 43 Month 10 P6	Week 60[†] Month 14 P7	Week 78 Month 18 P8	Week 95[†] Month 22 P9	End of Study for Enrolled and Discontinued Patients	14 Days Post Last Dose of Study Therapy	45 Days Post Discon-tinuation
Prescreen patients to determine if they meet entry criteria that can be screened by phone	X									
Schedule Screening Visit 1	X									
Encourage patient compliance and participation in study		X	X	X	X	X	X			
Review concomitant medication restrictions		X	X	X	X	X	X			
Monitor for interim study endpoints		X[‡]	X[‡]	X[‡]	X[‡]	X[‡]	X[‡]			
Monitor for adverse events, interim gastrointestinal (GI) endpoints									X[¶]	
Monitor for GI endpoints								X[§,‖]		X[‖]

[†] If study still ongoing.
[‡] Suspicion of a possible endpoint prompted an unscheduled visit and/or retrieval of additional information.
[§] Schedule end-of-study visit for those patients currently enrolled in study and on study drug. Priority given to patients who reported a possible PUB. Eight-hour fast required.
[‖] If a discontinued patient reported a possible PUB, then all documenting information was obtained.
[¶] Did not need to be performed for patients who completed the study. However, should the investigator have become aware of a serious adverse experience (SAE) or endpoint which occurred within 14 days post last dose of study therapy, these should have been reported.

Data Source: [3.2]

5.5.1.2 Documentation of Potential Endpoints

All potential endpoints occurring in the study were identified, documented and submitted for adjudication. The investigators were instructed as to the endpoint definitions and criteria for confirmation. At each investigator meeting, and in periodic newsletters, the potential signs and symptoms of upper GI endpoints and standard work-ups for these signs and symptoms were reviewed. However, it was the ultimate responsibility of the investigator to determine if a case qualified as a potential endpoint based on the specific clinical presentation. A properly completed significant GI Event Form (GICL), and a concise but complete clinical narrative of the case, both signed by the investigator, were required for a potential endpoint to be submitted for adjudication. The GICL worksheet was designed for purposes of investigator documentation of a potential endpoint. Instructions for the completion of the GICL were given to all investigators. Patients were asked to sign medical releases so that medical records could be obtained for any endpoint reported.

As part of the field procedures manual, all investigators were given instructions as to how to collect source documentation for and report potential endpoints, and store records related to them. Confidentiality of patient identifying information was maintained.

In addition, to the instructions given to the investigators, extensive efforts were made to ensure that endpoints did not inadvertently go unreported. Field monitoring and in-house data review personnel were instructed to review worksheets and the database for terms which may have been indicative of an endpoint or gastrointestinal work-up (e.g., gastric ulcer, GI bleeding, positive stool hemocults, endoscopies). Specific queries of the database were designed to look for these terms. When such terms were found, the investigator was requested to assess the event and determine if a GI endpoint had occurred. At all times, the decision to report an upper GI endpoint was made by the investigator.

All patients who discontinued early from the GI Outcomes study who did not have a GI endpoint reported were followed via telephone for the occurrence of an endpoint. To elicit this information, patients were asked about recent hospitalizations, GI work-ups and physician visits for GI-related events. These telephone contacts occurred at 14 and 45 days postdiscontinuation, and at study completion. The informed consent in the primary protocol covered the data proposed for patient follow-up. The Discontinued Patient Follow-up (DPF) Form was used to collect safety data regarding endpoints occurring after the usual 14-day postdiscontinuation follow-up. Important ancillary data, including concomitant medications, excessive alcohol use, and other relevant data were collected on the DPF as well.

5.5.1.5 Adjudication Criteria for Upper-GI Perforations, Ulcers, Obstructions, and Bleeds

Specific endpoint adjudication criteria were established and prespecified to allow the CRC to confirm the diagnosis reported by the investigator and determine whether the endpoint was clinically complicated (Table 6) . The CRC adjudicated each endpoint with respect to the confirmatory criteria first, followed by adjudication with respect to the clinically complicated criteria. Potential endpoints judged to meet the prespecified criteria by a majority of the CRC (2 of 3) were adjudicated as "confirmed." Similarly, an endpoint judged to meet the clinically complicated criteria by a majority of the CRC was adjudicated as "complicated." Thus, there were 4 classes of endpoints (confirmed and complicated, confirmed and uncomplicated, unconfirmed and complicated, and unconfirmed and uncomplicated). The CRC adjudicated an event as being "not an upper-GI event", if by majority opinion, the potential endpoint did not involve the upper-GI tract as defined (e.g., a case reported as an upper GI bleed by the investigator was determined by the committee to be a lower GI bleed based on the case documentation). In addition, the CRC may have reclassified a potential endpoint if there was sufficient evidence to do so (e.g., a pyloric channel ulcer reported as a "gastric ulcer" may have been reclassified as a "duodenal ulcer" based on the endoscopy report). All adjudications by the committee were final.

Table 6

Endpoint Adjudication Criteria

Event	Criteria for Confirmed Event	Criteria for Confirmed Complicated Events
Gastric or Duodenal Perforation due to Active Gastric Ulcer (GU) or Duodenal Ulcer (DU)	Report of gastric or duodenal perforation (excluding perforation caused by a malignant ulcer) confirmed by 1 or more of the following: 1. Endoscopy 2. Surgery 3. Unequivocal radiographic results consistent with free intraperitoneal air or extravasation of contrast media 4. Autopsy	All gastric or duodenal perforations are classified as complicated.

Table 6 (Cont.)

Endpoint Adjudication Criteria

Event	Criteria for Confirmed Event	Criteria for Confirmed Complicated Events
Obstruction due to Active Gastric Ulcer (GU) or Duodenal Ulcer (DU)	Postprandial nausea and vomiting lasting for at least 24 hours **AND** evidence of narrowing of the distal stomach, pylorus, or duodenum due to a nonmalignant ulcer documented by: 1. Endoscopy 2. Surgery 3. Radiography 4. Autopsy	All obstructions are classified as complicated
Development of Active Gastric Ulcer (GU) or Duodenal Ulcer (DU)	Report of GU or DU confirmed by one or more of the following: 1. Endoscopy 2. Surgery 3. Unequivocal radiological evidence of active GU or DU on upper-GI series with contrast 4. Autopsy	GU or DU associated with a confirmed upper-GI hemorrhage as defined under Development of Upper-GI Hemorrhage, criteria 1, 2, or 3.

Table 6 (Cont.)

Endpoint Adjudication Criteria

Event	Criteria for Confirmed Event	Criteria for Confirmed Complicated Events
Development of Upper-GI (esophageal, gastric, or duodenal) Hemorrhage	Report of upper-GI hemorrhage fulfilling one or more of the following: 1. Healthcare provider witnessed frank hematemesis (distinguished from blood tinged or streaked emesis), including coffee-grounds vomitus, **OR** healthcare provider-witnessed frank blood or coffee grounds by gastric aspiration or lavage (distinguished from scant coffee-grounds that clear rapidly) 2. Healthcare provider witnessed frank melena (distinguished from other dark stool, e.g., that due to bismuth salts). 3. Active upper-GI bleeding documented by endoscopy, angiography, or surgery. 4. Heme-positive stool associated with a documented upper-GI lesion judged by the healthcare provider to be the source of the bleeding **AND** associated with either of the following: a) Significant bleeding/volume loss b) Stigmata of recent bleeding (visible vessel, pigmented spot or clot on ulcer base) on endoscopy	1. Upper-GI hemorrhage associated with significant bleeding/volume loss (1).

Table 6 (Cont.)

Endpoint Adjudication Criteria

Event	Criteria for Confirmed Event	Criteria for Confirmed Complicated Events
	5. Patient reported hematemesis or melena associated with a documented upper-GI lesion judged by the healthcare provider to be the source of the bleeding **AND** associated with 1 or more of the following: a) Significant bleeding/volume loss b) Stigmata of recent bleeding (visible vessel, pigmented spot or clot on ulcer base) on endoscopy	1. Upper-GI hemorrhage associated with significant bleeding/volume loss (1).

(1) Criteria for significant bleeding/volume loss: One or more of the following (a, b, c, or d) is temporally related to the event:

a. Decrease in hemoglobin ≥2 gm/dL (or ≥6% drop in hematocrit if hemoglobin not available).
b. Evidence of orthostatic (sitting to standing, or lying to sitting) changes; one or more of: (i) pulse rate increase of >20 beats per minute (BPM), (ii) decrease in systolic blood pressure (SBP) >20 mm Hg, (iii) decrease in diastolic blood pressure (DBP) >10 mm Hg.
c. Other evidence of significantly reduced circulatory volume (e.g., significant hypotension corrected by volume replacement).
d. Transfusion of blood or packed red blood cells.

Data Source: [3.2]

5.5.1.6 Criteria for Exploratory "All GI Bleed" Analysis

One of the following predefined criteria needed to be met to be included in the analysis of clinically significant bleeds from any location in the GI tract:

a. Upper GI bleeds adjudicated by the CRC as confirmed or unconfirmed.

b. Adverse experiences suggestive of a lower GI bleed or GI bleed of unspecified location were identified from the adverse experience and serious adverse experience (SAE) forms. The adverse experience terms to be included were identified prior to unblinding. See [3.5] for a listing of included terms. To be included in this analysis, those adverse experiences must have met 1 of the following criteria:

   - Reported as a SAE;
   - Resulted in discontinuation of the patient from the study;
   - Associated with a 2-gm drop in hemoglobin from baseline within 14 days before the start date of the event and/or 30 days after.

c. Upper GI bleeds adjudicated by the CRC as "not an upper GI event" were included in this analysis if it met 1 of the following criteria:

   - Reported as a SAE.
   - Associated with a 2-gm drop in hemoglobin from baseline within 14 days before the start date of the event and or 30 days after.

5.7.1 Statistical and Analytical Plans to Address Study Objectives

5.7.1.1 Primary Objectives

The relative risk of confirmed PUBs in patients with rheumatoid arthritis taking 50 mg rofecoxib daily compared with patients taking naproxen 1000 mg daily was evaluated using the Cox proportional hazard model via SAS PROC PHREG (a procedure in SAS that does Cox proportional hazard model analyses) [3.5] with treatment as an explanatory factor and stratum of prior history of PUBs as a stratification factor.

The overall safety and tolerability of rofecoxib in patients with rheumatoid arthritis was evaluated by comparing the incidence of clinical and laboratory adverse experiences between treatment groups. A prespecified listing of potentially relevant safety parameters was examined. In addition, the percents of patients exceeding defined limits of change and mean values for clinical and laboratory safety measurements were compared among treatments.

5.7.1.2 Secondary Objectives

The relative risk of confirmed complicated PUBs in patients with rheumatoid arthritis taking 50 mg rofecoxib daily compared with patients taking naproxen 1000 mg daily was evaluated using the same method described for the primary PUB endpoint.

The relative risk of confirmed and unconfirmed PUBs in patients with rheumatoid arthritis taking 50 mg rofecoxib daily compared with patients taking naproxen 1000 mg daily was evaluated using the Cox proportional hazard model via SAS PROC PHREG [3.5] with treatment as an explanatory factor, and stratum of prior history of PUBs and study region effects (U.S. versus multinational) as stratification factors.

The relative risk of confirmed and unconfirmed complicated PUBs in patients with rheumatoid arthritis taking 50 mg rofecoxib daily compared with patients taking naproxen 1000 mg daily was evaluated using the same method described for the primary PUB endpoint.

The efficacy of treatment of rheumatoid arthritis with rofecoxib or naproxen was evaluated using 95% CI on the difference between treatment groups in average change from baseline for Patient and Investigator Global Assessment of Disease Status and by comparing the discontinuation rates due to lack of efficacy.

5.7.1.3 Exploratory Objectives

The relative risk of occurrence of bleeding from any location in the GI tract in patients taking 50 mg rofecoxib daily versus patients taking naproxen 1000 mg daily was evaluated using the same method described for the primary PUB endpoint.

5.7.2.3 Criteria to Determine a Positive Study

The trial was to be considered a positive study if a significant ($p<0.050$) reduction in the risk of confirmed PUBs in the rofecoxib 50-mg daily group compared to the naproxen 1000-mg daily group was found and if a trend ($p\leq0.20$) was found for the reduction in risk of confirmed complicated PUBs.

5.7.2.4 Power and End of Study Stopping Rule

The study was planned to stop when a minimum of 120 patients experienced confirmed PUB events, 40 patients experienced confirmed complicated PUB events, or 6 months after the last patient was randomized, whichever came last.

For the primary gastrointestinal safety hypothesis, the targeted number of patients with events of 120 provided at least 97% power ($\alpha=0.05$, 2-tailed) to detect a reduction in risk of at least 50%. This calculation accounted for 1 interim analysis described in 5.7.3.3. The targeted number of patients with complicated PUB events of 40 provided more than 80% power to show a trend ($p\leq0.20$) if the reduction in confirmed complicated PUBs due to rofecoxib was ≥50% and more than 80% power to show a statistically significant effect ($p<0.05$) if the reduction due to rofecoxib was ≥60%. The targeted sample size of 3500 patients per treatment group assumed that the upper-GI side effects of perforations, ulceration, obstructions, and bleeding would occur in 2 to 4% of RA patients treated with NSAIDs for 1 year and was chosen to provide a reasonable study duration under varying assumptions about dropout rates and patient accrual.

5.7.3 Statistical/Analytical Methods and Issues

5.7.3.1 Approaches to Analyses

5.7.3.1.1 All-Patients-Randomized Approach

The primary approach for gastrointestinal safety endpoints was based on an All-Patients-Randomized (APR) population, i.e., all patients randomized were included based on their randomized treatment assignment. The primary time frame for the analysis of the PUB data included a 14 day postdiscontinuation follow-up period. Since most of the endpoints are analyzed as time-to-first-event, no values were imputed.

5.7.3.4 Assessment of Consistency of Treatment Effects Across Subgroups

To explore whether treatment effects were consistent across different subgroups, treatment-by-factor interactions were evaluated for the primary endpoint in the All-Patients-Randomized population. The patient characteristics and baseline covariates of interest were:

- Prior history of PUBs (yes/no)
- Study region (U.S./non-U.S.)
- Age group (<65 years/≥65 years)
- Ethnic group (Caucasian/Other)
- Gender (female/male)
- Baseline use of systemic corticosteroids (yes/no)

For each subgroup variable listed above, a Cox regression model was performed for the primary endpoint and included the treatment, subgroup, and treatment-by-subgroup interaction. The interactions were tested at α=0.05 significance level. When an interaction is not significant, the main treatment effect is interpreted reasonably as the effect averaged over the different levels of the subgroup factor.

Summary statistics (cases, patient-years at risk, incidence rates, relative risk, and 95% confidence interval for relative risk) were presented within the subgroups for the primary endpoint.

5.7.3.5 Multiplicity

There was only 1 primary endpoint and one primary treatment group comparison defined for VIGOR, and the interim analysis was conducted using sequential stopping boundaries. For the primary endpoint in the final analysis, the p-values and confidence intervals quoted were adjusted to reflect the interim analyses. Secondary analyses were used to support and help interpret the primary analyses, and thus, no p-value adjustment for multiplicity was applied other than the adjustment for the interim analysis.

No p-value adjustments were applied to the numerous safety evaluations to be overly conservative with regard to missing items of interest.

5.7.3.7 Other Safety Analyses

Safety and tolerability were assessed by statistical and/or clinical review of all safety parameters, including adverse experiences, laboratory values, and vital signs, as described in this section. All patients randomized were included in the safety analyses. Formal statistical tests focused on prespecified safety concerns while estimates were provided for all other parameters. The following variables were tested (described in the following sections) in the analysis of adverse experiences: discontinuations due to Digestive adverse experiences including abdominal pain, discontinuations due to hypertension, discontinuations due to edema, discontinuations due to renal-related adverse experiences, discontinuations due to hepatic-related adverse experiences, and congestive heart failure (CHF).

Dual COX-1/COX-2 inhibitors such as naproxen inhibit platelet aggregation via suppression of serum levels of thromboxane B2 which is a product of platelet-derived COX-1. Naproxen, like aspirin and unlike other NSAIDs such as diclofenac and meloxicam, has been shown to maximally inhibit platelet aggregation throughout its dosing interval [1.2.23; 1.2.24]. In contrast,

rofecoxib, a specific inhibitor of COX-2 does not suppress serum levels of TXB_2, and therefore has been shown to have no effect on platelet aggregation [1.2.23; 1.2.24]. Low-dose cardioprotective aspirin was not allowed in this study since even low-dose aspirin can affect gastric mucosal COX-1. Therefore, there was the theoretical possibility that naproxen, through its effects on platelet aggregation, may have provided cardioprotective effects not provided by rofecoxib, resulting in a lower incidence of thromboembolic events in the naproxen treatment group. To assess this possibility, cardiovascular thrombotic or embolic serious adverse experiences (SAEs) were adjudicated by an independent committee as a part of a program-wide effort. Procedures for handling these SAEs and the analytic methods to be used were defined in separate documents that can be found in [3.2]. The adjudicated events, as opposed to the reported SAEs, were considered primary. Results of these analyses are described in the Cardiovascular Events Analysis [2.1.6].

5.7.3.7.1 Adverse Experiences

Survival analysis methods were used to analyze prespecified adverse experiences. For such adverse experiences, time-to-event was analyzed and cases, patient-years at risk, relative risk, CIs, and p-values were determined. p-Values and 95% CIs for relative risk ratios (rofecoxib versus naproxen group) were computed using the Cox proportional hazard model with treatment as the explanatory factor. Numbers, proportions, and 95% CIs on the difference in proportions [3.5] were provided for all other adverse experiences.

Prespecified adverse experiences included:

- Serious clinical adverse experiences (overall)
- Drug-related (possibly, probably, definitely) clinical adverse experiences (overall)
- Clinical adverse experiences leading to study discontinuation (overall)
- Discontinuations due to Digestive adverse experiences including abdominal pain
- Discontinuations due to edema-related adverse experiences
- Discontinuations due to hypertension-related adverse experiences

- Discontinuations due to renal-related adverse experiences (clinical and/or laboratory adverse experiences)
- Discontinuations due to hepatic-related adverse experiences (clinical and/or laboratory adverse experiences)
- CHF adverse experiences
- Serious laboratory adverse experiences (overall)
- Drug-related (possibly, probably, definitely) laboratory adverse experiences (overall)
- Laboratory adverse experiences leading to study discontinuation (overall)

5.8.2 Protocol Amendments

The original protocol was amended 4 times. All amendments were made prior to unblinding of the database, after frozen file, and before the first interim analysis. Major changes included in the amendments were as follows:

1. The primary hypothesis was changed from an assessment of "cumulative incidence" to "relative risk" to better conform with the planned statistical analyses.
2. Secondary hypotheses assessing the relative risk in the 2 treatment groups of: (a) confirmed and unconfirmed PUBs; and (b) confirmed complicated PUBs were added. After discussions with the Steering Committee and regulatory agencies, the importance of obtaining sufficient data on complicated endpoints was made clear. Therefore the secondary hypotheses were added and the end-of-study stopping rule was changed such that a minimum of 40 confirmed complicated endpoints was required in addition to a minimum of 120 PUBs.
3. Endpoint definitions were refined after receiving feedback from regulatory agencies. These revised endpoint definitions were used by the Case Review Committee to adjudicate all endpoints.
4. At the request of regulatory agencies, the modified HAQ was added as an exploratory efficacy measurement in the United States.

References

1. Wollheim FA, Selective Cox-2 inhibition in man-therapeutic breakthrough or cosmetic advance? Rheumatology, 2000; 39:935-38
2. Wolfe MM Lichtenstein DR and Singh G, Gastrointestinal toxicity of nonsteroidal anti-inflammatory drugs, NEJM. 1999; 340:1888-99