

Observed Data Cohort

A patient's data at a specific visit were included in this analysis if he or she satisfied the requirements for the ITT Cohort and the corresponding assessment days after the first dose of study medication fell in the following intervals: 14 ± 5 days for Week 2; 42 ± 7 days for Week 6; and 84 ± 9 days for Week 12.

The analyses were performed for Evaluable and Observed Data Cohorts at all scheduled visits and also at the 'Final Visit', which consisted of the last valid observation of the patient.

The Physician's and Patient's Global Assessments were classified based on changes as "improved" (a reduction of at least two grades from Baseline for grades 3-5 or a change in grade from 2 to 1), "no change," or "worsened" (an increase of at least two grades from Baseline for grades 1-3 or a change in grade from 4 to 5) and analyzed by the Cochran-Mantel-Haenszel (CMH) test stratified by center.

Mean change analyses, including the linear trend test for all Cx and placebo groups, and overall and pairwise comparisons for all five treatment groups were performed by using analysis of covariance (ANCOVA) with treatment and center as factors, and the corresponding Baseline score as covariate. Additionally, the Q-Ratio with 95% confidence intervals was calculated by taking the ratio of adjusted mean changes for each Cx treatment groups vs. the naproxen treatment group.

The results of the pairwise comparisons for the Cx 100 mg BID and 200 mg BID treatment groups vs. placebo were interpreted using Hochberg's step-up procedure. P-values of comparisons between Cx 100 mg BID and Cx 200 mg BID vs. placebo for the ITT Cohort were ordered from larger to smaller. The larger p-value was examined first, and, if $p < 0.05$, then it was declared that both doses were significantly different from placebo and no further examination was performed. If the larger p value was > 0.05 , the smaller p-value was checked. If the smaller p-value was ≤ 0.025 , then the corresponding dose was claimed to be significantly different from placebo. For other comparisons, an alpha level of 0.05 was used to summarize the results.

The above categorical and mean change analyses were performed on the ITT Cohort, the Evaluable Cohort and the Observed Data Cohort.

The categorical status of "improved," "no change," or "worsened" for the Global Assessments of Arthritic Condition was also calculated for each patient based on a one-grade change from Baseline. These analyses were performed for the ITT Cohort. In addition, for the ITT Cohort with LOCF approach, differential effects of gender, age and duration of disease were examined by ANCOVA models including factors as follows:

1. Age or gender or duration and center, treatment, and Baseline;
2. Age by gender or age by duration or gender by duration, lower order terms, and center, treatment, and Baseline;
3. Age by treatment, lower order terms, center, and Baseline;
4. Gender by treatment, lower order terms, center, and Baseline; and
5. Duration by treatment, lower order terms, center, and Baseline.

Mean change from Baseline for quality of life data observed at Week 2 and Week 12 or Early Termination was analyzed using ANCOVA with treatment and center as factors and corresponding Baseline score as covariate. This analysis was performed on the ITT Cohort only. For the mean change, a positive value represents an improvement and a negative value represents a worsening.

Efficacy Results for OA:

Reviewer's comment: The following comments of OA efficacy refer ONLY to the ITT LOCF analysis.

Primary endpoints:

Patient and Physician globals in both studies showed that Cx at all doses studied (i.e. 50 mg BID, 100 mg BID, 200 mg BID) was efficacious vs. placebo. For example, the Physician's Global Assessment of Arthritis Pain (see *Appendix Table A.7.1 and A.7.2*) and Patient's Global Assessment of Arthritis Pain (see *Appendix Table A.8.1 and A.8.2*) in protocol 054 shows improvements (categorical and mean analyses) over time in all treatment groups; it should be noted that improvements were based upon a two (2) categorical change in globals (see above). Improvements in the global scores seemed to be maintained during the 12 weeks of this trial. Celecoxib at all doses is better than placebo and comparable to naproxen but, overall, patients are still symptomatic. With these endpoints, there does not appear to be any additional benefit from the higher doses of Cx. Comparison by Q values also suggests there are no differences between doses of Cx and Naproxen in either the patient or physician global. Results are similar for protocol 020.

Patient's Assessment of Arthritis Pain also demonstrated in both studies that Cx at all doses studied (with the exception of 50 mg BID in protocol 020) was efficacious as compared to placebo. After flare occurred, baseline pain scores appeared comparable (see *Appendix Table/Figure A.9.1-.2 and Table/Figure A.10.1-.2*) across treatment groups as well as across studies; these pain scores improved ($p < 0.05$, except Cx 50 mg BID in protocol 020) over time with Cx; improvements appeared comparable to those seen with naproxen. There seemed to be no additional improvement at 12 weeks (in fact, there are some suggestions of waning of response at 12 weeks). Interestingly,

comparison by Q values also suggests there are no differences between all doses of Cx and naproxen. Utilizing these endpoints, there does not appear to be any additional benefit from the higher doses of Cx. Also of note, patients are still apparently symptomatic as judged by the week 12 pain scores.

As noted above, the WOMAC scores were added as primary outcomes in the course of the IND development. Both the WOMAC subscales (i.e. pain, stiffness, function) and the WOMAC composite showed Cx at all doses in both trials to be efficacious compared to placebo (see *Appendix A.11-A.18*). Effect sizes were comparable to naproxen. There were generally consistent differences between the lower (i.e. 50 mg BID) and higher (i.e. 100/200 mg BID) doses of Cx; but not between the higher doses. As noted with the other primary endpoints, patients improved but were still apparently symptomatic.

Secondary endpoints:

In protocols 054 and 020, the OSI index correlated well with the results of the primary endpoints at all doses of Cx (data not shown). This index again suggested that there was a dose response between 50 mg BID and the higher doses of Cx but nothing consistently different between the higher doses.

The SF 36 index did not generally reveal any significance at the lower dose of Cx used for short periods of time (i.e. 25 and 40 mg BID in studies 047 and 013, respectively), but did show significance for the physical functioning, role physical, bodily pain, vitality, social functioning, and mental health at Cx doses of ≥ 50 mg BID.

Withdrawal due to lack of arthritis efficacy (i.e. treatment failure) shows a similar trend in both studies, although the placebo rates differ (see *Appendix A.19*). Higher doses of Cx generally lead to fewer patients withdrawing from the study, this was most evident in protocol 054. Top doses of Cx had similar rates of withdrawal to that of naproxen.

Time to withdrawal due to lack of arthritis efficacy (see *Appendix A.20-A.21*) showed that all doses of Cx were significantly different than placebo and tended to be similar to naproxen, especially at the higher doses of Cx; these trends were more obvious in the hip study (054).

As can be seen, the reasons for study termination (in all groups) were primarily due to treatment failure in study 020 and 054, as was the case for all the placebo-controlled trials (see *Appendix Table A.22.1-2*). There was a decrease in these treatment failure rates, compared to placebo, in the Cx-treated patients which tended to plateau at the higher doses and was similar (sometimes better, sometimes worse) to the rates seen with naproxen. On the other hand, termination for an adverse event tended to increase with increasing doses of Cx. With one exception (Cx at 100 mg BID, study 020), Cx was comparable to (or better than) naproxen while tending to be worse than (or comparable

to) placebo in terms of adverse event rates. Not unexpectedly, termination due to treatment failure and for adverse events were lower, in these six-week (Study 060, 087) vs. the twelve-week studies (Study 020, 021, 054).

Other OA Studies

As noted in table 4 above, protocol 021 was also a 12 week trial in OA of the knee. The study design, treatments, patient demographics and number of patients treated, as well as primary and secondary endpoints were similar (if not identical) to both protocols 020 and 054. The results of the primary and secondary endpoints show no significant differences from those seen in the other 12 week trials in OA as noted above. The 4-week (047) OA trial showed Cx 25 mg BID to be ineffective while the 100 mg BID dose of Cx showed significance only for the globals and not the VAS pain scale or WOMAC. However, in this study, a dose of 400 mg BID showed efficacy in all these four primary endpoints.

Single-dose (QD) vs. multiple-dose (BID) OA trials:

There were two randomized, double-blind, placebo-controlled, parallel group, multicenter six week trials that addressed the issue of alternate dosing schedules for Cx (see table 4). Both of these studies (i.e. protocol 060 and 087) involved patients with OA of the ("Index joint") knee that was in "flare" (see definition above for 12 week studies) and they both employed the same doses of Cx, either 200 mg QD (evening, with placebo in the morning) or 100 mg BID (morning and evening). The schedule of observations and procedures in these two trials (for an example in study 060, see *Appendix, Table A.23*) differed primarily in that the SF-36 and samples for PK analysis were not collected in study 087. Inclusion and exclusion criteria for patients to participate in these studies were similar to the 12 week studies discussed above.

Between protocol 060 and 087, a total of 1399 patients were enrolled and received at least one dose of study drug as follows:

•placebo	474 patients
•SC-58635 100 mg BID	472 patients
•SC-58635 200 mg QD	453 patients

Primary measures of arthritis efficacy were Patient's Global Assessment of Arthritic Condition, Patient's Assessment of Pain-Visual Analog Scale (VAS), and Physician's Global Assessment of Arthritic Condition. **Secondary measures of arthritis efficacy** were Functional Capacity Classification, WOMAC Index, Incidence of Withdrawal Due

to Lack of Arthritis Efficacy, Time to Withdrawal Due to Lack of Arthritis Efficacy, and Osteoarthritis Severity Index.

The patient demographics were comparable to those of the 12 week studies being elderly, white females; there were no obvious imbalances between the treatment groups in these 6 week studies.

Primary endpoints:

Both the patient and physician's globals showed similar trends and effect sizes for patients treated with Cx to those seen at comparable times, and at comparable doses, in the 12 week OA studies (for example in protocol 087, see *Appendix Table A.24.1-2 and Table A.25.1-2*); the placebo responses appeared generally more robust in the 6 week studies. Both dosing regimens of Cx were significantly different than placebo while there did not appear to be any difference between the two doses of Cx (either by Q-ratios or p values).

The patient's assessment of arthritis pain also revealed the patients to have been comparable to the patients in the 12-week studies, both in terms of their baseline pain, and their response to treatment (see *Appendix Table A.26 and A.27*). Once again, both dosing regimens of Cx were significantly different than placebo and there did not appear to be any difference between the two doses of Cx.

Secondary endpoints:

The WOMAC index (composite plus subscales) was evaluated in both of these 6-week trials. As can be seen with the WOMAC pain index (*Appendix Table A.28 - A.29*), the baseline characteristics of both dosing regimens of Cx appeared similar to that of placebo as well as to the treatment groups in the 12-week studies. Similarly, Cx was significantly different than placebo and did not appear to differ between the two dosing schemes for Cx. Similar results were noted for the WOMAC function, stiffness, and composite scales.

The results of the OSI index were also comparable between these 6-week trials and the 12-week studies. It is not possible to comment on the SF-36 since this was not obtained in protocol 087.

The time to withdrawal due to lack of arthritis efficacy in these 6-week trials was, not unexpectedly, quite different than the results (for all treatment groups) obtained in the 12-week trials in that, overall, not as many patients withdrew in these shorter studies (see *Appendix Table A.30*). However, significantly fewer patients withdrew in the Cx groups compared to placebo and the two dosing schedules of Cx do not appear different in this regard. The results of both the 6-week trials are similar. The time to withdrawal due to lack of arthritis efficacy (see *Appendix Table A.31*) show these same trends.

Non-Flare vs. Flare Studies:

There were three studies (062, 071, and 042) that allowed patients with OA or RA to enter the trials without the requirement for "flares" as noted in the other OA studies above. One of these trials was conducted outside the U.S. (protocol 042) and will not be discussed here. These studies were intended to evaluate several endpoints as noted in the brief review of study 071 below.

Study 071:

This randomized, double-blind, parallel group, multicenter, 12-week study was designed primarily to compare the cumulative incidence of gastroduodenal ulcers associated with celecoxib 200 mg BID with that of diclofenac 75 mg BID and ibuprofen 800 mg TID in patients with OA or RA. The efficacy and overall safety of Cx compared to diclofenac and ibuprofen were also assessed in this trial.

Patients were eligible to participate in the study if they had a documented clinical diagnosis of OA or RA (not necessarily in flare) with a Functional Capacity Classification of I-III and required chronic NSAID treatment. At the time of study enrollment, patients underwent an endoscopy to ensure they did not have an esophageal, gastric, pyloric channel, or duodenal ulcer.

The efficacy endpoints for OA in this study were Patient's Global Assessment of Arthritic Condition and Physician's Global Assessment of Arthritic Condition. Arthritis Assessments were performed at Baseline and at the Weeks 4, 8, and 12 (or Early Termination) follow-up visits. UGI safety was assessed by serial endoscopy and biopsy and overall safety was assessed by comparison of physical examinations, clinical laboratory tests, and incidence of adverse events between treatment groups.

Table 7 below summarizes the numbers and types of patients studied in protocols 062 and 071.

Table 7. Types/Numbers of Patients Studied in Non-Flare Trials

Study	Diagnosis	Number of Patients Receiving:				Total
		Cx 200 mg BID	Naproxen 500 mg BID	Diclofenac 75 mg BID	Ibuprofen 800 mg BID	
062	OA	194	195	-	-	389
	RA	76	72	-	-	148
071	OA	271	-	285	254	810
	RA	94	-	102	91	287
Total		635	267	387	345	1634

Although patients had OA or RA, it is unclear exactly how this diagnosis was made in these protocols. As indicated by the Inclusion Criteria on the admission CRF, eligibility into these two studies was based upon a clinical diagnosis of OA or RA of at least three

months duration. The disease also had to be of sufficient severity to warrant the patient require chronic NSAID therapy. The determination of OA was made by the investigator; patients with OA of any joint (i.e. ankle, elbow, shoulder, knee, hip) were eligible. Radiologic evidence, other ACR criteria, or other methods of diagnosis were not specifically required by the study protocol.

While it is not possible to draw any accurate comparisons to the placebo-controlled OA and RA (see below) studies, it is of interest that these patients had about a one category difference (at baseline) from the patients studied in the flared OA designs (i.e. baseline of about 2.8 vs. 3.8). Similarly, the treatment responses (or effect size) based upon the patient and physician globals in these patients tended to be about half of those noted in the flared OA studies (i.e. 0.5 vs. 1.0). Also of note, there was a tendency in both trials for more patients to drop from the Cx treatment group vs. the comparator NSAIDs.

Conclusions from the OA trials:

Efficacy in the treatment of the signs and symptoms of OA has been demonstrated in NDA 20-998 by adequate and well-controlled trials. Endpoints in these trials that helped to establish (but not limited to these) efficacy include the patient and physician's globals (Appendix Table A.7-A.8), assessment of pain-categorical/VAS (Appendix Table A.9-A.10), WOMAC scores-subcales and composite (Appendix Table A.11-A.18) along with the OSI Index (Appendix Table A.19) and time to withdrawal due to lack of arthritis effect (Appendix Table A.20-A.21).

There was generally a dose-response between Cx 50 mg BID and higher doses, but not between Cx 200 mg BID and lower doses. Patients were generally still symptomatic. Treatment responses generally appeared durable for 12 weeks but some endpoints in certain trials (i.e. primary endpoints-Patient Assessment of Arthritis Pain) suggest a waning of this response over time. Celecoxib, given as a single evening dose of 200 mg is equally as efficacious as 100 mg BID by the same endpoints noted above. It is difficult to draw any conclusions from the non-flare trials. Celecoxib, especially at the higher doses, is comparable in efficacy to naproxen.

Therefore, following conclusions regarding Cx and treatment of the signs and symptoms of OA are drawn from the information (ITT/LOCF) presented to this point in the randomized clinical trials:

- Cx from 100 mg BID to 200 mg BID is consistently efficacious vs. placebo
- Cx 50 mg BID is not consistently efficacious vs. placebo
- Cx (200 mg BID) is not consistently more efficacious vs. Cx (100 mg BID)
- Cx (100-200 mg BID) has efficacy comparable to Naproxen 500 mg BID
- Cx (100 mg BID) is as efficacious as Cx (200 mg QD)

Rheumatoid Arthritis Efficacy Trials:

Seven studies (see table 1: "Studies Included in NDA 20-998") were conducted in patients with RA; two designated "pivotal" and five "supportive" (including the long-term safety study 024). These trials are summarized in table 8.

Table 8. Summary Characteristics of Rheumatoid Arthritis Trials:

Placebo- and Active-Controlled Pivotal Studies			
Protocol No. Report No. Short Title	No. of Investigators Country(ies) Start Date	Study Design (Duration of Treatment)	Treatment Regimen(s)
P: N49-96-02-022 R: N49-98-06-022 Celecoxib Comparative Efficacy and UGI Safety vs Naproxen in RA	81 Investigators U.S. and Canada 6 Sep 1996	Randomized, Double-Blind, Placebo-Controlled, Active Controlled, Multicenter, Parallel (12 Weeks)	Celecoxib 100 mg BID, 200 mg BID, or 400 mg BID or Naproxen 500 mg BID or Placebo
P: N49-96-02-023 R: N49-98-06-023 Comparative Efficacy and Safety vs Naproxen in RA	77 Investigators U.S. and Canada 7 Aug 1996	Randomized, Double-Blind, Placebo-Controlled, Active Controlled, Multicenter, Parallel (12 Weeks)	Celecoxib 100 mg BID, 200 mg BID, or 400 mg BID or Naproxen 500 mg BID or Placebo
Placebo-Controlled Supportive Study			
Protocol No. Report No. Short Title	No. of Investigators Country(ies) Start Date	Study Design (Duration of Treatment)	Treatment Regimen(s)
P: N49-96-02-012 R: N49-97-06-012 Pilot Efficacy in RA	29 Investigators United States 1 Feb 1996	Randomized, Double-Blind, Placebo-Controlled, Multicenter, Parallel (4 Weeks)	Celecoxib 40 mg BID, 200 mg BID or 400 mg BID or Placebo
Active-Controlled Supportive Studies			
Protocol No. Report No. Short Title	No. of Investigators Country(ies) Start Date	Study Design (Duration of Treatment)	Treatment Regimen(s)
P: I49-96-02-041 R: I49-98-06-041 Ex-U.S. Efficacy/GI Safety vs Diclofenac in RA	132 Investigators 21 countries in Australia, Europe and South Africa 28 Nov 1996	Randomized, Double-Blind, Active Controlled, Multicenter, Parallel (24 Weeks)	Celecoxib 200 mg BID or Diclofenac SR 75 mg BID
P: N49-97-02-062 R: N49-98-06-062 Comparative Incidence of UGI Ulcers: Celecoxib vs Naproxen in Patients with OA and RA	75 Investigators United States 13 May 1997	Randomized, Double-Blind, Active Control, Multicenter, Parallel (12 Weeks)	Celecoxib 200 mg BID or Naproxen 500 mg BID
P: N49-97-02-071 R: N49-98-06-071 Comparative Incidence of UGI Ulcers: Celecoxib vs Diclofenac and Ibuprofen in Patients with OA and RA	121 Investigators United States 21 Jul 1997	Randomized, Double-Blind, Active Control, Multicenter, Parallel (12 Weeks)	Celecoxib 200mg BID or Diclofenac 75 mg BID or Ibuprofen 800 mg TID

Uncontrolled Supportive Study

Protocol No. Report No. Short Title	No. of Investigators Country(ies) Start Date	Study Design (Duration of Treatment)	Treatment Regimen(s)
P: N49-96-02-024 R: pending N49-98-08-024 (Interim Data Listings) Long-term Safety in OA and RA	278 Investigators U.S. and Canada 17 Jun 1996	Open Label, Multicenter (1-2 Years)	Celecoxib 200-400mg BID (for RA)

Reviewer's comment: The only protocols of adequate duration and characteristics for review are 022 and 023; this review will focus primarily on the efficacy and dose-response characteristics of Cx from these 12-week trials. Protocols 062 and 071 have been discussed in the OA efficacy section as well as in the GI differentiation section. Similarly, the endoscopic portion of protocol 022 will not be discussed here since this will be covered in detail in the GI safety section. The open-label experience, protocol 024, is discussed further on in this review.

Study characteristics:

As noted in table 8 above, studies 022 and 023 were twelve-week, double-blind, placebo-controlled, multicenter, parallel group comparisons of Cx versus placebo and naproxen in patients with RA. Table 9 below summarizes the experience with RA in this NDA:

Table 9. Number of Patients with RA Studied in All Protocols (excludes open-label)

Study	Treatment								Total
	Placebo	Cx (mg, BID)				Naproxen (mg, BID)	Diclofenac (mg, BID)	Ibuprofen (mg, TID)	
		40	100	200	400	500	75	800	
012	85	81	-	82	82	-	-	-	330
022	231	-	240	235	218	225	-	-	1149
023	221	-	228	219	217	218	-	-	1108
041	-	-	-	326	-	329	-	-	655
062 ¹	-	-	-	76 (270)	-	72 (267)	-	-	148 (537)
071 ¹	-	-	-	94 (366)	-	-	102 (387)	91 (346)	287 (1089)
Total	537	81	468	1032	517	844	102	91	3672

1. Numbers in () = total number patients with RA studied in these protocols (i.e. remainder had OA)

As can be seen in table 9, between protocols 022 and 023, a total of 2252 patients with RA were enrolled and received at least one dose of study medication as follows:

• placebo	452
• Cx 100 mg BID	468
• Cx 200 mg BID	454
• Cx 400 mg BID	435
• Naproxen 500 mg BID	443

These studies (i.e. 022 and 023) were both double-blind, placebo-controlled, multicenter, parallel group comparisons of Cx versus placebo and naproxen in patients with RA. They consisted of 12 weeks of treatment with visits occurring at Pretreatment/Screening, Baseline, and at Weeks 2, 6, and 12 following the first dose of study drug (see *Appendix Table A.32* for details of Protocol 022 as an example of the schedule of observations and procedures). The studies differed primarily in that protocol 022 included an assessment of the UGI safety of Cx with endoscopies performed at Baseline and Week 12 (or Early Termination) with testing done for *Helicobacter pylori* (*H. pylori*) at Baseline and the Week 12 (or Early Termination) Visit. In protocol 023, blood samples were taken (approximately 40 patients/treatment group) at selected sites between day 7 and 28 after the first dose for determination of Cx plasma levels.

Patients with diagnosed RA in a flare state were enrolled and randomized to receive Cx 100 mg BID, Cx 200 mg BID, or Cx 400 mg BID, naproxen 500 mg BID, or placebo.

To qualify for study participation, candidates must have:

1. Been of legal age of consent or older;
2. For women of childbearing potential, confirmed use of adequate contraception since last menses and confirmed continued use of adequate contraception during the study, were not lactating, and had a negative serum pregnancy test within 7 days prior to the Baseline Arthritis Assessments
3. Been diagnosed as having adult-onset RA of at least three month's duration as defined by the 1987 American College of Rheumatology (ACR) classification criteria
4. Had a Functional Capacity Classification of I-III at the Baseline Visit
5. Been stable on NSAID therapy and had a Functional Capacity Classification that had not changed for at least one month immediately preceding the NSAID washout period
6. Had RA in a flare state within two to seven days after discontinuing NSAID therapy (within four to seven days for patients who received either oxaprozin, piroxicam, or both)
7. Provided written informed consent before undergoing any study procedures

Candidates were not eligible for admission if they met any one of the following:

1. Had been diagnosed with any other inflammatory arthritis
2. Had been diagnosed with a secondary, non-inflammatory type of arthritis (e.g., osteoarthritis or fibromyalgia) that, in the Investigator's opinion, was symptomatic enough to interfere with the evaluation of the effect of Cx on the patient's primary diagnosis of RA
3. Had begun taking any of the following medications or had changed the dosing regimen of any of these medications within 12 weeks before receiving the first dose of study medication:
 - a) Gold salts (including oral gold)
 - b) Sulfasalazine (doses of up to 3 g/day were allowed)
 - c) Azathioprine
 - d) Antimalarials
 - e) Penicillamine;
4. Had begun taking or had changed the dosing regimen of methotrexate within the eight weeks preceding the first dose of study medication. The methotrexate dose was not to exceed 20 mg/week
5. Had begun taking oral corticosteroids or had changed the dose regimen of oral corticosteroids within four weeks before receiving the first dose of study medication (doses of up to 10 mg prednisone or equivalent/day were allowed), or the patient had received intramuscular, intra-articular, or soft-tissue injections of corticosteroids within four weeks before receiving the first dose of study medication
6. Had received any antineoplastic (other than methotrexate \leq 20 mg/week or azathioprine as therapy for RA) during the eight weeks preceding the first dose of study medication
7. Had taken any NSAID (including aspirin) within two days before the Baseline Arthritis Assessments or any analgesic within 24 hours before the Baseline Arthritis Assessments. (Patients taking \leq 325 mg aspirin per day for non-arthritic reasons for at least 30 days before the first dose of study medication were allowed to continue their aspirin regimen for the duration of the study. Patients must have discontinued oxaprozin or piroxicam at least four days before the Baseline Arthritis Assessments.)
8. Had an active malignancy of any type or history of malignancy. (Patients who had a history of basal cell carcinoma that had been treated were eligible. Patients with a history of other malignancies that had been surgically removed and who had no evidence of recurrence for at least five years before study enrollment were also eligible.)
9. Had been diagnosed with or had received treatment for esophageal, gastric, pyloric channel, or duodenal ulceration within 30 days before receiving the first dose of study medication
10. Had active GI disease (e.g., inflammatory bowel disease) or has an esophageal, gastric, pyloric channel or duodenal ulcer (an ulcer was defined as any break in the mucosa at least 3 mm in diameter with unequivocal depth) or more than ten erosions in the stomach or more than ten erosions in the duodenum on the Baseline UGI endoscopy
11. Had a history of any gastric or duodenal surgery other than simple oversew;
12. Had chronic/acute renal or hepatic disorder or a significant coagulation defect
13. Had abnormal screening laboratory test values within seven days before the Baseline Arthritis Assessments that were >1.5 x upper limit of normal (ULN) for either AST (SGOT) or ALT (SGPT) or any other laboratory abnormality considered by the Investigator to be clinically significant

14. Had a known hypersensitivity to COX-2 inhibitors, sulfonamides, or NSAIDs
15. Had received any investigational medication within 30 days before the first dose of study medication or was scheduled to receive an investigational drug, other than study medications described in the protocol, during the course of the study
16. Had previously been admitted to this study.

Reviewer's comment: *It should be noted that the exclusion criteria for study 023 (022 included endoscopy) did not include items 10 and 11 noted above.*

All study patients had to demonstrate an arthritis flare within two to seven days after discontinuing their NSAID or analgesic. Patients receiving oxaprozin or piroxicam must have discontinued these NSAIDs at least four days before the Baseline Arthritis Assessments. An RA flare was demonstrated if the Physician's Global Assessment of Arthritic Condition and the Patient's Global Assessment of Arthritic Condition were "fair," "poor," or "very poor" at the Baseline Visit AND if a comparison of the Screening Arthritis Assessments and the Baseline Arthritis Assessments met criteria 1 and 2 described below plus either criterion 3 or 4:

1. MUST have had a minimum of six tender joints at Baseline AND an increase of at least two tender or painful joints (or 20% increase in the number of tender/painful joints, whichever was greater) at the Baseline as compared to the Screening Visit
2. MUST have had a minimum of three swollen joints at Baseline AND an increase of at least two swollen joints (or 20% increase in the number of swollen joints, whichever was greater) at the Baseline as compared to the Screening Visit
3. A minimum of 45 minutes of morning stiffness at Baseline AND an increase in the duration of morning stiffness of at least 15 minutes as compared to the Screening Visit
4. Patient's Assessment of Pain-Visual Analog Scale measurement of at least 40 mm (on a visual analog scale) at Baseline AND an increase of 10 mm (or 20% increase, whichever was greater) at the Baseline as compared to the Screening Visit

At each follow-up visit, patients were asked the following question: "Since your last visit, have you experienced any symptoms that are not associated with your arthritis?" Any symptom was recorded on the Adverse Signs and Symptoms CRF. Patients who withdrew before the end of the study had all final assessments performed at the time of withdrawal (Early Termination Visit).

Demographics:

There did not appear to be any remarkable differences in baseline characteristics between treatment groups in the 12-week RA trials (see *Appendix Table A.33*). These patients tended to be white females, in their 50's, with a disease duration of approximately 10 years. About 40% of patients used corticosteroids, 65% other disease-modifying anti-rheumatic drugs (DMARDs), and about 50% used methotrexate (MTX). Approximately 75% of patients in all groups were, therefore, concurrently using corticosteroids or MTX

or other DMARDs (data not shown). Patients and their physicians rated the baseline global assessments of arthritis condition as fair. The number of tender and swollen joints were comparable across treatment groups in both studies. Of note, the mean average of the weights of all the patients with RA (i.e. 77-78 kg, data not shown) was substantially different than those patients noted earlier with OA. The demographic characteristics, arthritis history and co-therapy for each individual study were consistent with these pooled results.

Primary/Secondary Endpoints:

The primary measures of arthritis efficacy were:

- **ACR-20 Responder Index;**
- **Patient's Global Assessment of Arthritic Condition;**
- **Number of Tender/Painful Joints;**
- **Number of Swollen Joints;**
- **Physician's Global Assessment of Arthritic Condition.**

ACR-20 responder index:

In order to examine the overall effect of the study drug on the patient's condition, a categorical analysis was performed on all patients who met the ACR-20 criteria as improved compared to Baseline. A patient was classified as "improved" if (compared to baseline) the patient experienced:

- A. $\geq 20\%$ improvement in
 - tender/painful joint count (TJC)
 - swollen joint count (SJC) 20%
- AND
- B. $\geq 20\%$ improvement in at least three of the following five assessments
 - Physician's Global Assessment of Arthritic Condition
 - Patient's Global Assessment of Arthritic Condition
 - Patient's Assessment of Pain-VAS
 - CRP (as example of acute phase reactant)
 - HAQ Functional Disability Index.

The Patient's and Physician's Global Assessments of Arthritic Condition were made independently. Patient's were asked to answer the question, "Considering all the ways your arthritis affect you, how are you doing today?" Patients rated, and physician's graded, according to the 5-point categorical scale below:

- | | |
|---------------------|--|
| 1. Very good | Asymptomatic and no limitation of normal activities |
| 2. Good | Mild symptoms and no limitation of normal activities |
| 3. Fair | Moderate symptoms and limitation of some normal activities |
| 4. Poor | Severe symptoms and inability to carry out most normal activities |
| 5. Very poor | Very severe symptoms that are intolerable; inability to carry out all normal activities |

To determine the Number of Tender/Painful Joints, sixty-eight joints (right and left) were examined for joint tenderness/pain. The joints were as follows:

- Temporomandibular
- Sternoclavicular
- Acromioclavicular
- Shoulder
- Elbow
- Wrist (radiocarpal, carpal, and carpometacarpal considered as one unit)
- Metacarpophalangeals (MCP I, II, III, IV, V)
- Thumb interphalangeal (IP)
- Proximal interphalangeals (PIP II, III, IV, V)
- Distal Interphalangeals (DIP II, III, IV, V)
- Knee
- Hip
- Ankle
- Tarsus (includes subtalar, transverse tarsal, and tarsometatarsal as one unit)
- Metatarsophalangeals (MTP I, II, III, IV, V)
- Great Toe interphalangeal (IP)
- Proximal and distal interphalangeals combined (PIP II, III, IV, V)

In response to pressure or motion, each joint was graded as painful or tender using the scale shown below:

- | | |
|----------|--|
| 0 | No response (not tender) |
| 1 | Positive response to questioning (tender) |
| 2 | Spontaneous response elicited (tender and winced) |
| 3 | Withdrawal by patient on examination (tender, winced, and withdrew) |

To determine the Number of Swollen Joints, sixty-six joints were also graded for swelling using the same joints as those listed above (for joint pain/tenderness) except that the hip joints were not assessed. The joint swelling scale was graded using the scale below:

- | | |
|----------|---|
| 0 | None |
| 1 | Detectable synovial thickening without loss of bony contours |
| 2 | Loss of distinctiveness of bony contours |
| 3 | Bulging synovial proliferation with cystic characteristics |

Secondary Measures of Efficacy were:

- Patient's Assessment of Pain-Visual Analog Scale
0 mm = no pain, 100 mm = very severe pain
- Tender/Painful Joints Score
- Swollen Joints Score
- SF 36 (eight domains, see OA section)
- Duration of Morning Stiffness
average duration for the previous three days
- HAQ Functional Disability Index (eight areas of daily living, graded on scale from 0 = without any difficulty to 3 = unable to do)
- CRP
- Incidence of Withdrawal Due to Lack of Arthritis Efficacy
- Time to Withdrawal Due to Lack of Arthritis Efficacy
- ACR-50 Responder Index

Patient Populations Analyzed/Statistics:

The ITT Cohort included all patients with RA who were randomized to treatment and who had taken at least one dose of study medication. The Last Observation Carried Forward (LOCF) approach was used for either missing data or data that was obtained on days that fell outside the observation window (i.e. >19 days for Week 2, >49 days for Week 6, and >93 days for Week 12). The LOCF approach was employed in the ITT analyses only.

Evaluable Cohort

A patient was considered evaluable for analysis of arthritis assessments for Week 2, Week 6, Week 12 or Early Termination if, in addition to satisfying the requirements for the ITT Cohort, he or she:

1. Was diagnosed by ACR criteria as having adult onset RA
2. Had a Functional Capacity Classification of I-III at the Baseline Visit
3. Had RA in a flare state at the Baseline Visit
4. Did not have any other inflammatory arthritis or any secondary, noninflammatory-type arthritis that, in the Investigator's opinion, would interfere with the evaluation of Cx
5. Did not receive IM, IA, or soft-tissue injections of corticosteroids or begin or change dose regimen of oral corticosteroids within four weeks before the first dose of study medication
6. Did not begin or change dose regimen of the following within 12 weeks before the first dose of study medication: gold salts, sulfasalazine, azathioprine, antimalarials, or penicillamine;
7. Did not begin or change the dose regimen of methotrexate within eight weeks before the first dose of study drug;
8. Did not take any antineoplastic, other than methotrexate (≤ 20 mg/week) or azathioprine as therapy for RA within eight weeks before the first dose of study medication
9. Did not take any NSAID or analgesic within 24 hours before the Baseline Arthritis Assessments

10. Underwent the Baseline Arthritis Assessments within seven days before the first dose of study drug;
11. Did not take any of the following proscribed medications during the course of the study:
 - any antineoplastic (other than methotrexate \leq 20 mg/week or azathioprine as treatment for RA)
 - any NSAID (other than aspirin \leq 325 mg/day)
 - any injectable corticosteroid
 - any analgesic (other than acetaminophen up to 2 g/day for nonarthritic reasons)
12. Did not change dose regimen or initiate treatment with the following during the study: corticosteroids, gold salts, penicillamine, methotrexate, antimalarials, azathioprine, or sulfasalazine;
13. Was compliant with study medication as described below:
 - for the Week 2 Visit, the patient took at least 70% of the doses prescribed from Day 1 through the Week 2 Visit
 - for the Week 6 Visit, the patient took at least 70% of the doses prescribed from the Week 2 Visit through the Week 6 Visit AND at least 50% of the doses prescribed from Day 1 through the Week 2 Visit
 - for the Week 12 Visit, the patient took at least 70% of the doses prescribed from the Week 6 Visit through the Week 12 Visit AND at least 50% of the doses prescribed from the Week 2 Visit through the Week 6 Visit AND at least 50% of the doses prescribed from Day 1 through the Week 2 Visit
14. Underwent the Arthritis Assessments for each visit under consideration according to the following schedule:
 - a. 14 ± 5 days after the first dose of study medication for the Week 2 Visit
 - b. 42 ± 7 days after the first dose of study medication for the Week 6 Visit
 - c. 84 ± 9 days after the first dose of study medication for the Week 12 Visit
 - d. ≤ 2 days after the last dose of study medication for the Final Visit
15. Had complete primary efficacy data available for each visit under consideration.

Evaluability determinations were made prior to unblinding the data and no subsequent revisions were made.

Observed Data Cohort

A patient's data at a specific visit was included in this analyses if he or she satisfied the requirements for the ITT Cohort and the corresponding assessment days after the first dose of study medication fell in the following intervals: 14 days \pm 5 days for Week 2; 42 ± 7 days for Week 6; and 84 ± 9 days for Week 12.

Statistical analyses were performed for the Evaluable and Observed Data Cohorts at all scheduled visits and at the Final Visit which consisted of the last valid observation of the patient.

Mean change analyses, including the linear trend test for all Cx and placebo groups and overall and pairwise comparisons for all five treatment groups, were performed on all primary measures of efficacy with the exception of the ACR-20 responder index, using an analysis of covariance (ANCOVA) with treatment and center

as factors, and the corresponding Baseline value as a covariate. Additionally, the Q-Ratio with 95% confidence intervals was calculated by taking the ratio of adjusted mean changes for each Cx treatment group versus the naproxen treatment group.

The results of the pairwise comparisons for the Cx 200 mg BID and 400 mg BID treatment groups versus placebo for the ITT Cohort were interpreted using Hochberg's step-up procedure.

For Assessment of Joint Tenderness/Pain and the Assessment of Joint Swelling, a joint was classified as "improved" if a reduction in grade to 0 or a change from 3 to 1 was observed. A joint was classified as "worsened" if an increase in grade from 0, a change in grade from 1 to 3, or a change in grade from 2 to 3 was observed. The median number of "improved" joints was compared between treatment groups using ANCOVA with the Baseline number of joints that had a score greater than zero as the covariate and center and treatment as factors. The number of "worsened" joints was similarly analyzed. In addition, the patient's overall status was considered as "improved" if the difference between the number of improved and the number of worsened joints was greater than or equal to 50% of the number of Baseline joints that had a score greater than zero. A patient was classified as "worsened" if the difference between the number of worsened and the number of improved joints was greater than or equal to 50% of the number of Baseline joints that had a score greater than zero. Patient's overall status was analyzed by the CMH test stratified by center.

For Physician's and Patient's Global Assessments of Arthritic Condition a patient was classified as "improved" if a reduction of at least two grades from Baseline for grades 3 to 5 or a change in grade 2 to 1 was observed. A patient was classified as "worsened" if an increase of at least two grades from Baseline for grades 1 to 3 or a change in grade 4 to 5 was observed. The changes were analyzed by the CMH Test stratified by center. The linear trend test (naproxen group excluded) and pairwise comparisons were performed based on the above CMH tests.

Efficacy Results for RA:

Primary endpoints

Patient and Physician globals

With reference to the ITT analyses, patient and physician globals in both studies (022, 023) showed the baseline characteristics of the patients in all the treatment groups were comparable within, and between trials. Celecoxib, at all doses studied (i.e. 100 mg BID, 200 mg BID, 400 mg BID) was consistently efficacious compared with placebo. Naproxen had the same results with the one exception it did not show significance at week 12 in the categorical analysis of trial 022. For example, the Physician's Global Assessment of Arthritis Condition (see *Appendix Table A.34.1-2*) and the Patient's Global Assessment of Arthritis Condition (see *Appendix Table A.35.1-2*) for study 023 (results are similar for 022), shows improvements (categorical and mean change analyses) versus placebo over time in all Cx treatment groups. Although there may be some suggestions of waning over time, improvements in the Cx-treated global scores seemed to be maintained during the 12 weeks of this trial. There does appear to be a difference between 100 mg BID and the higher doses of Cx, but not a consistent dose-response relationship for higher doses of Cx. In certain situations, such as the Physician Globals for protocol 022, higher doses of Cx also appear to be more efficacious than Naproxen; the Q-ratio analysis suggests the same. However, these same trends regarding comparison to naproxen are not evident in the protocol 023.

With reference to Cx, there were **NO** statistically significant differences compared to placebo (categorical or mean change analyses) at any time point (except the 2 week assessments in both trials and categorical analysis for Cx 200 mg BID in study 022) in either the **evaluable or observed cohorts** in either trial (022, 023) in the Patient or Physician's global assessments (data not shown).

With reference to naproxen (and considering only protocol 023) and the evaluable or observed cohorts, there were statistically significant differences compared to placebo at all time points (categorical and mean change analysis), with the exception of the Physician global (mean analysis) at week 12 and the Patient Global (categorical) at week 12. In study 022, on the other hand, only the 2 week time points revealed any significant difference compared to placebo for both the physician and patient globals.

Tender/Painful Joint Counts

Considering the ITT analyses, the tender/painful joint counts (TJC) were comparable at baseline (though high, mean of approximately 29 joints) between groups within each study as well as between the two protocols. The placebo response in protocol 022 (see *Appendix, Table A.36.1-2*) was more robust than that seen in study 023 (data not shown). This may account for the fact that naproxen did not show significance vs. placebo at week

12 (mean or categorical) but it did in trial 023. However, all doses of Cx were significantly different (categorical and mean change analyses) than placebo at all times (i.e. weeks 2, 6, 12) in both trials. There were no consistent dose-response trends between the various doses of Cx but the responses appeared durable. Overall, Cx appears comparable to naproxen; the Q-ratio analysis suggests the same.

With reference to Cx and the evaluable and observed cohorts there were NO statistically significant differences compared to placebo (categorical or mean change analyses) at any time point (except the 2 week assessments in both trials and a single mean analysis for Cx 100 and 400 mg BID in study 023 at week 6; evaluable and observed, respectively) in either trial (022, 023) (data not shown). Naproxen also showed significance at all two week time points and at 6 weeks (both mean analyses-evaluable and observed).

Swollen Joint Counts

Looking at the ITT analyses, the swollen joint counts (SJC) were comparable at baseline (again high, mean of approximately 21 joints) between groups within each study as well as between the two protocols. Once more, the placebo response was a little more robust for trial 022 (data not shown). Similar to the TJC, Cx was significantly different (categorical or mean change analyses) from placebo at all times points and at all doses in both trials with the notable exception (categorical analysis) of the 100 and 400 mg BID doses in trial 023 (see *Appendix, Table A.37.1-2*). No obviously consistent dose-response trends were evident between the three doses of Cx, but the responses noted appeared durable throughout the trials. Again, Cx appears comparable to Naproxen; the Q-ratio analysis suggests the same.

With reference to Cx and the evaluable and observed cohorts, there were NO statistically significant differences (categorical or mean change analyses) compared to placebo at any time point (except for a various doses at the 2 week assessments in both trials) in either trial 022 or 023 (data not shown). The same can be said regarding naproxen.

ACR-20 and ACR-50 Responder Index

Based on the ACR-20 Responder Index (ITT cohort), there was a statistically significant difference in the percentage of patients classified as responders in all doses of Cx compared to placebo at all time points in both protocols. The one exception to this statement was that significance was not achieved with CX at 100 mg BID in protocol 023 (see *Appendix Table A.38.1-3*). Once again, there did appear to be a difference in response between 100 mg BID and the higher doses, but not between the higher doses. This distinction between 100 mg BID of Cx and the higher doses is more evident in the ACR-50 (ITT cohort) Responder Index (see *Appendix Table A.39.1-2*).

However, in the evaluable cohort of protocol 023 (see *Appendix Table A.38.3*), only the naproxen group showed significance in the ACR-20 index; results were the same in the observed cohort group of this trial (i.e. only naproxen showed significance at week 12). On the other hand, there were no significant differences from placebo in the ACR-20 index for any of the treatments (Cx or naproxen), at any time point, in these other cohorts in protocol 022 (data not shown).

Secondary endpoints

Reviewer's comment: The reader will notice that not all secondary endpoints will be discussed and that some of these endpoints are part of the ACR -20/50 primary endpoints. Only the ITT/LOCF results are noted.

Patient's Assessment of Arthritis Pain (VAS)

The baseline VAS scores were comparable between the groups in both study 023 and 022, as well as between the studies (baseline of approximately 66). In both studies, the analyses of mean changes revealed that there were statistically significant differences from placebo at all doses of Cx and at all time points. The same is true for naproxen (see *Appendix Table A.40* for example in protocol 023).

C-reactive protein (CRP)

The baseline CRPs showed differences which, in light of the variation in results, is difficult to interpret. There were no statistically significant differences from placebo for any of the doses of Cx at any time point, in either protocol. Naproxen did show significance at only one time point, week 12 in trial 023 (see *Appendix Table A.41*). Of note, as discussed below for protocol 012, Cx also did not seem to effect ESR or Serum Amyloid A levels.

HAQ Functional Disability Index

The baseline HAQ scores were comparable between groups in both studies as well as between the studies (mean around 1.4). There were consistent statistically significant differences for the 200 and 400 mg BID doses of Cx and naproxen as compared to placebo, but not for Cx at 100 mg BID compared to placebo, in both studies (see *Appendix Table A.42* for example in protocol 023). Q-ratio analysis suggests there is no difference between the higher doses of Cx and naproxen, but there are differences between Cx at 100 mg BID and naproxen.

SF 36 Health Survey

Mean change analyses (from baseline to week 12 or early termination) were performed for scores for the eight SF-36 Health Survey domains: Physical Functioning, Role-

Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, and Mental Health.

Most domains showed statistically significant improvement compared to placebo in both the Cx and naproxen doses. The most notable exceptions were the General Health and Role Emotional where protocol 023 did not show any significance for these domains with all doses of Cx whereas the results were exactly opposite in protocol 022 (i.e. all doses of Cx did show significance). In a few other domains in protocol 022, such as the Physical Functioning and Role Physical, there was a separation of the lower dose of Cx (i.e. 100 mg BID) and the higher doses with the latter showing significance (data not shown).

Incidence of Withdrawal Due to Lack of Arthritis Efficacy

The Incidence of Withdrawal Due to Lack of Arthritis Efficacy (treatment failure) for both protocols (see *Appendix Table A.43*) reveals withdrawal of a total of 774 patients (345 and 429 for study 022 and 023, respectively) regardless of treatment. As would be expected if there was a favorable treatment effect over placebo, there were more patients in the placebo groups who withdrew due to lack of arthritis efficacy (51%) compared to any of the Cx treatment groups (27-34%, see pooled results). The differences were in withdrawal rate for all doses of Cx were statistically significant ($p < 0.001$) compared to placebo as noted in both individual trials (data not shown). Although there were more patients in the Cx 100 mg BID group (34%) compared to the Cx 200 mg BID (27%) and 400 mg BID (29%) groups who withdrew due to lack of arthritis efficacy, these differences were not statistically significantly different.

There were also fewer patients in the naproxen group who withdrew due to lack of arthritis efficacy (30%) than in the placebo group and this difference was again statistically significant ($p < 0.001$). However, there were no significant differences between patients taking any dose of Cx compared to naproxen as noted in the individual studies.

Time to Withdrawal Due to Lack of Arthritis Efficacy

The results of the analysis of the Time to Withdrawal Due to Lack of Arthritis Efficacy are presented as Kaplan-Meier estimates (see *Appendix Table A.44* for example from study 023). Again, as would be expected if treatment had an effect, in both studies placebo patients tended to withdraw earlier than patients in the Cx treatment groups and this difference in the time to withdrawal was statistically significant ($p < 0.001$); the same can be said for naproxen. While there was a statistically significant difference noted in study 023 between Cx 100 mg BID and 400 mg BID, this was not the case in study 022 ($p = 0.954$) and so there were no obvious differences between any of the Cx doses in either study. Patients in the naproxen group also tended to withdraw later than patients in the placebo. Differences between naproxen and Cx were inconsistent comparing to the lower doses of Cx (i.e. Cx 100 and 200 mg BID) and naproxen between studies; however, there were consistently no differences seen between Cx 400 mg BID and naproxen.

Other RA Studies:

Of the remaining studies submitted in support of the indication of RA (Table 8), only trial 012 will be described briefly here. The other trials (041, 062, 071) are intended primarily to address the GI safety issue (see UGI Safety Review) and/or have a mixed patient populations with entry criteria unsuited for adequate interpretation.

Study 012

Protocol 012 was a pilot, Phase II, double-blind, placebo-controlled, parallel-group study evaluated the safety and effectiveness of Cx in treating the signs and symptoms in patients with RA in a flare state. Three hundred thirty (330) patients received treatment for four weeks as follows: placebo, 85 patients; Cx 40 mg BID, 81 patients; Cx 200 mg BID, 82 patients; and Cx 400 mg BID, 82 patients. Arthritis assessments and safety evaluations were performed at Baseline and at Weeks 1, 2, and 4.

The measures of arthritis efficacy included: the Patient's and Physician's Global Assessment of Arthritic Condition, Patient Assessment of Arthritis Pain, Number of Tender/Painful Joints, Number of Swollen Joints, Incidence of Withdrawal due to lack of Arthritis Efficacy, Time to Withdrawal, and the ACR 20. With reference to these assessments at week 4 in the ITT population, Cx 40 mg BID was not different than placebo but Cx 200 and 400 mg BID were consistently statistically different than placebo and from Cx 400 mg BID. There were no consistent differences between Cx 200 and 400 mg BID. The ACR 20 response at week 4 was 31%, 51% and 52% for Cx 40, 200, and 400 mg BID, respectively (placebo = 29%). Also of interest, the ESR, CRP and serum amyloid A levels did not seem consistently effected by any of the doses of Cx at any time point.

Conclusions from the RA trials:

Efficacy in the treatment of the signs and symptoms of RA has been demonstrated in NDA 20-998 by adequate and well-controlled trials. Endpoints in these trials that have helped to establish (but not limited to these) efficacy include the patient and physician's globals (Appendix Table A.34-A.35), assessment of pain-categorical/VAS (Appendix Table A.40), Tender and Painful Joint Counts (Appendix Table A.36), Swollen Joint Counts (Appendix Table A.37), ACR 20 Responder Index (Appendix Table A.38) ACR 50 Responder Index (Appendix Table A.39) and HAQ Functional Index Disability Index (Appendix Table A.42). Of note, Cx (and naproxen), did not seem to effect the acute phase responses as witnessed by a lack of effect on C-reactive protein levels (Appendix Table A.41). Celecoxib, in study 012, also did not seem to effect ESR or Serum Amyloid A levels.

There was a general (though not universal) dose-response between Cx 100 mg BID and higher doses, but not between Cx 400 mg BID and lower doses. Treatment responses generally appeared durable for 12 weeks but some endpoints in certain trials (i.e. primary endpoints-Physician and Patient's Globals) suggest a waning of this response over time. Generally speaking, fewer patients withdrew for lack of treatment effect at the higher doses of Cx (Appendix Table A.43); all doses were superior to placebo in this regard. Placebo patients also tended to withdraw earlier from studies than Cx-treated patients (Appendix Table A.44).

Neither the "evaluable" nor the "observed" cohort analyses for Cx generally demonstrated a consistent statistically significant difference from placebo at weeks 6 and 12. This was also the overall pattern seen with naproxen, but results were inconsistent between trials. Celecoxib, especially at the higher doses, is comparable in efficacy to naproxen.

Therefore, the following conclusions regarding Cx and treatment of the signs and symptoms of RA are drawn from the information (ITT/LOCF) presented to this point in the randomized clinical trials:

- Cx from 100 mg BID to 400 mg BID is consistently efficacious vs. placebo
- Cx 200 and 400 mg BID is frequently more efficacious vs. Cx 100 mg BID
- Cx 200 mg BID and 400 mg BID generally have comparable efficacy
- Cx (100 mg-400 mg BID) has efficacy comparable to Naproxen 500 mg BID

Long-Term, Open-Label Experience in OA and RA:

Study N49-96-02-024 is an ongoing, long-term open-label safety study of patients who previously participated in one of following nine phase II or III double-blind controlled studies:

- N49-96-02-012 (RA)
- N49-96-02-013 (OA)
- N49-96-02-020 (OA)
- N49-96-02-021 (OA)
- N49-96-02-022 (RA)
- N49-96-02-023 (RA)
- N49-96-02-054 (OA)
- N49-97-02-062 (OA/RA)
- N49-97-02-071 (OA/RA)

All patients treated in the long-term, open-label study previously participated in one of the nine controlled studies listed above. A 14-day rule was used to determine direct transfer status as follows:

- If a patient received any celecoxib dose in the controlled study and transferred into the open label study within 14 days, the patient was considered a direct transfer patient and all previous study data were included in the long-term analysis (Day 1 of celecoxib is the first day of the double-blind study);
- If a celecoxib patient transferred after 14 days then Day 1 of celecoxib is the first day of the open label study
- Patients who received placebo or an active control agent in the double-blind study are evaluated as Day 1 of celecoxib in the open-label study regardless of the gap between studies.

This multicenter study is/was designed to determine the long-term (up to two years) safety, including an evaluation of the incidence of any clinically significant gastrointestinal events, of Cx administered to patients with osteoarthritis OA or RA. Efficacy assessments (see below) are also being performed. **The data cutoff date for the interim data listings included in this NDA is November 21, 1997.** The results of the completed trial are pending at this time; it is anticipated to be completed in 12/99.

For two-year patients, visits included the Baseline, at Weeks 2 and 6, and at Months 3, 6, 9, 12, 15, 18, 21, and 24. For patients enrolled for one year, the Month 12 visit is the final study visit. For both two-year and one-year patients, study visits are to include review of any treatment-emergent signs and symptoms experienced since the previous visit. Safety assessments are generally combined for OA and RA.

Measures of arthritis efficacy include:

- Patient's Assessment of Arthritis Pain on the Visual Analog Scale (VAS);
- Patient's Global Assessment of Arthritis;
- Physician's Global Assessment of Arthritis
- Functional Capacity Classification.

These assessments will be performed on all patients at every visit, with the exception of the Patient's Assessment of Arthritis Pain on the Visual Analogue Scale (VAS), with the exception of patients previously enrolled in N49-97-02-062 or N49-97-02-071. Patients will undergo a physical examination at the Baseline visit and every six months thereafter. Clinical laboratory tests will be performed at every study visit. Two-year patients will complete a quality of life assessment (SF-36 Health Survey) at Baseline and every six months thereafter, and a Health Resource Utilization Questionnaire at every visit except Baseline. One-year patients will not complete the SF-36 Health Survey or the Health Resource Utilization Questionnaire at any study visit.

A radiologic examination (i.e., hand and wrist x-rays for patients with RA and either the Index knee or the Index hip for patients with OA) will also be performed at Baseline and the Month 12, or Early Termination, visit for all patients, except those previously enrolled in N49-97-02-062 or N49-97-02-071.

As of the cutoff date, a total of 4499 patients had entered the long-term, open-label safety study. A total of 3256 patients were still active in the study at the cutoff date; the remaining 1243 had prematurely terminated from the study. The longest duration of treatment (patient 0150001) was 522 days.

Table 10 briefly summarizes the disposition of patients to the NDA cutoff (Nov. 21, 1997) for study 024:

Table 10. Disposition of Patients in Long-Term, Open-Label Protocol (024)

Category	Placebo	Cx (all doses)	NSAIDs	Total
Pts able to enroll	1270	4422	2073	7765
Pts enrolled (%)	860 (68)	2776 (63)	863 (42)	4499 (58)
Pts at 12 months	-	-	-	3256 (72)

Reviewer's comment: There is a discrepancy between the number of patients still active and those that have terminated between this text (i.e. 3256 and 1243, respectively) and tables cited below of 61 patients. In other words, the tables suggest there are 61 patients still receiving Cx that the text states have been terminated from study 024.

In study 024, the doses of Cx allowed have ranged from 100-200 mg BID for OA and 200-400 mg BID for RA. This range was allowed to control symptoms (increased) or for tolerability reasons (reduced). As can be seen (*Appendix, Table A.45*), approximately 70 % of patients with either OA or RA, increased their dose beyond what is felt to be the therapeutic dose during the randomized controlled studies presented in this NDA (i.e. 100 mg BID for OA, 200 mg BID for RA). Of those that did increase their dose, most moved to a dose twice as high (i.e. 200 mg BID for OA, 400 mg BID for RA). This appears to be a demonstration of the clinical observation of "dose creep".

Of the efficacy parameters assessed in protocol 024, the Patient's Global Assessment of Arthritic Condition for OA and RA are presented (see *Appendix, Figure A.1*); results are very similar for the Patient Assessment of Pain (VAS) and the Physician's Global Assessment for both the patients with OA and RA. Regarding figure 7 (OA) and figure 10 (RA) of Appendix figure A.1, it is noted by the sponsor:

"Although approximately 70% of OA patients did escalate the dose, there was no worsening of arthritis status compared to Baseline prior to dose escalation. In addition, following dose escalation, little additional improvement was noted in mean scores compared to patients who took celecoxib 200 mg BID without escalating their dose. This data lends further support to the conclusion that celecoxib 100 mg BID is an efficacious dose and an increase to 200 mg BID does not significantly enhance the efficacy in treating the signs and symptoms of OA."

"Although approximately 75% of RA patients did escalate their dose (to 300 or 400 mg BID), there was no evidence of worsening arthritis status compared to Baseline prior to dose escalation. In addition, following dose escalation, little additional improvement was noted in mean scores compared to patients who took celecoxib 200 mg BID without escalating their dose. This finding lends further support to the conclusion that celecoxib 100 mg BID and 200 mg are efficacious doses and 400 mg BID does not significantly enhance the efficacy in treating the signs and symptoms of RA"

Reviewer's comment: It could just as easily be argued that an escalation of the dose was required to maintain any long-term efficacy of Cx in OA and RA.