Drug Class Review
on Cyclo-oxygenase (COX)-2 Inhibitors
and
Non-steroidal Anti-inflammatory Drugs (NSAIDs)

UPDATED FINAL REPORT #1
September 2003

Mark Helfand, MD, MPH
Oregon Evidence-based Practice Center
Oregon Health & Science University
TABLE OF CONTENTS

Introduction 4

Methods 7
Literature Search 7
Study Selection 8
Data Abstraction 8
Validity Assessment 8

Results 9
Key Question 1. In head-to-head comparisons, are there differences in efficacy or safety between different COX-2 inhibitors? 9

Key Question 2. Are there differences in efficacy between coxibs, COX-2 selective NSAIDs, nonselective NSAIDs, or other drugs? 12

Key Question 3. Are there clinically important differences in safety or adverse effects between coxibs, COX-2 selective NSAIDs, nonselective NSAIDs, and the combination of a nonselective NSAID plus antiulcer medication? 14

Key Question 4. Are there subgroups of patients based on demographics, other medications, or co-morbidities for which one medication is more effective or associated with fewer adverse effects? 22

Summary 22

References 26

Tables
Table 1. One year risk of GI bleeding due to NSAID
Table 2. Rofecoxib vs. celecoxib in chronic osteoarthritis of the knee
Table 3. Trials of celecoxib vs. nonselective NSAIDs
Table 4. Trials of celecoxib versus NSAIDs
Table 5. Trials of rofecoxib versus NSAIDs
Table 6. Trials of meloxicam versus NSAIDs
Table 7. Comparison of the VIGOR and CLASS trials
Table 8. Re-analysis of the CLASS and VIGOR Trials
Table 9. Tolerability in randomized controlled trials
Table 10. Strength of evidence by key question
INTRODUCTION

Compared with placebo, non-steroidal anti-inflammatory drugs (commonly called NSAIDs) reduce pain significantly in patients with arthritis, low back pain, minor injuries, and soft tissue rheumatism. However, NSAIDs have important adverse effects, including gastrointestinal (GI) bleeding, peptic ulcer disease, hypertension, edema, and renal disease.

NSAIDs reduce pain and inflammation by blocking cyclo-oxygenases (COX), enzymes that are needed to produce prostaglandins. Most NSAIDs block two different cyclo-oxygenases, called COX-1 and COX-2. COX-2, found in joint and muscle, contributes to pain and inflammation.

NSAIDs cause bleeding because they also block the COX-1 enzyme, which protects the lining of the stomach from acid. In the US, complications from NSAIDs are estimated to cause about six deaths per 100,000 population, a higher death rate than that for cervical cancer or malignant melanoma. A risk analysis based on a retrospective case-control survey of emergency admissions for upper GI disease in two UK general hospitals provided useful estimates of the frequency of serious GI complications from NSAIDs. In people taking NSAIDs, the 1-year risk of serious GI bleeding ranges from 1 in 2,100 in adults under age 45 to 1 in 110 for adults over age 75, and the risk of death ranges from 1 in 12,353 to 1 in 647:

<table>
<thead>
<tr>
<th>Age range (years)</th>
<th>Chance of GI bleed due to NSAID</th>
<th>Chance of dying from GI bleed due to NSAID</th>
</tr>
</thead>
<tbody>
<tr>
<td>16-45</td>
<td>2100</td>
<td>12,353</td>
</tr>
<tr>
<td>45-64</td>
<td>646</td>
<td>3800</td>
</tr>
<tr>
<td>65-74</td>
<td>570</td>
<td>3353</td>
</tr>
<tr>
<td>&gt; 75</td>
<td>110</td>
<td>647</td>
</tr>
</tbody>
</table>

Data are from Blower, recalculated in Moore and in Bandolier

NSAIDs differ in their selectivity for COX-2—how much they affect COX-2 relative to COX-1. An NSAID that blocks COX-2 but not COX-1 might reduce pain and inflammation in joints but leave the stomach lining alone. Appendix A summarizes the NSAIDs and their selectivity based on assay studies (done in the laboratory instead of in living patients). The table gives an idea of how widely NSAIDs vary in their selectivity, but should be interpreted with caution. Different assay methods give different results, and no assay method can predict what will happen when the drug is given to patients. Clinical studies, rather than these assay studies, are the best way to determine whether patients actually benefit from using more selective NSAIDs.

Scope and Key Questions

The purpose of this review is to compare the benefits and harms of different NSAIDs, focusing in particular on the COX-2 inhibitors and COX-2 selective NSAIDs. To refine this subject, the OHSU Evidence-based Practice Center (EPC) developed preliminary key questions, populations, interventions, outcome measures, and eligibility criteria for discussion. They were reviewed in several public meetings attended by members of the Health Resources Commission NSAIDs Subcommittee, OHSU EPC.
Based on input from these individuals, the OHSU EPC refined the questions and scope of the review. The key questions were:

1. In head-to-head comparisons, are there differences in efficacy or safety between different COX-2 inhibitors?
2. Are there differences in efficacy between coxibs, COX-2 selective NSAIDs, and nonselective NSAIDs?
3. Are there clinically important differences in safety or adverse effects between coxibs, COX-2 selective NSAIDs, nonselective NSAIDs, and the combination of a nonselective NSAID plus antiulcer medication?
4. Are there subgroups of patients based on demographics, other medications, or co-morbidities for which one medication is more effective or associated with fewer adverse effects?

Several aspects of the key questions merit comment:

1. **Patients.** We focused on patients with chronic pain from osteoarthritis, rheumatoid arthritis, soft-tissue pain, or back pain. We included ankylosing spondylitis. COX-2 inhibitors are also used to treat dysmenorrhea and acute pain (e.g., dental or surgical pain), and to prevent the formation of colorectal polyps. We did not examine studies of the use of coxibs for these indications.

2. **Efficacy.** The main efficacy measures are pain, functional status, and discontinuations due to lack of efficacy. Measures vary among studies. Frequently used measures are:

   - **Visual analogue scale (VAS):** The patient indicates their level of pain, function, or other outcome by making a mark on a scale labeled with numbers (such as 0 to 100) or descriptions (such as “none” to “worst pain I’ve ever had”).
   - **Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC):** A 24-item questionnaire used to assess the functional status of patients with osteoarthritis of the knee and hip. A lower score indicates better function.
   - **Patient Global Assessment of Disease Status and Investigator Global Assessment of Disease Status:** The patient or investigator answers questions about the overall response to treatment, functional status, and pain response, using a VAS or Likert scale.
   - **American College of Rheumatology (ACR) criteria:** Measure disease activity and response to treatment. ACR 20, ACR 50, or ACR 70 reflect either an improvement to the 20%, 50%, or 70% level in the parameters outlined.

3. **Safety and adverse effects.** The following events were included in the review:

   a. **Serious GI events (GI bleeding, symptomatic ulcer disease, perforation of the GI tract, and death).**
   b. **Serious cardiovascular events (myocardial infarction, angina, stroke, transient ischemic attack, cardiovascular death, and related measures).**
   c. **Tolerability and adverse events.** We recorded discontinuation due to any adverse event, any serious adverse event, the overall rate of adverse events, the rate of GI adverse events, and the combined rate of adverse events related to renal and cardiovascular function, including increased creatinine, edema, hypertension, or congestive heart failure. We also
recorded the frequency of, and discontinuations due to, abnormal laboratory tests, primarily elevated transaminases (liver tests).

Several types of adverse events were excluded:

d. The main non-clinical, or intermediate, outcome measure for GI adverse effect is **endoscopic ulcer**. Ulcers in the stomach or small intestine can be seen in up to 40% of patients taking NSAIDs.\(^\text{10, 11}\) As many as 85% of these ulcers can only be found by endoscopy because they do not cause symptoms or bleeding. All three COX-2 inhibitors in the US market significantly reduce the incidence of these asymptomatic ulcers. Based on input from the subcommittee, we did not include endoscopic ulcer as an outcome measure, since our focus is on clinically significant adverse events.

e. **Case reports.**

i. **Aseptic meningitis.** Near the end of March 2002, the Food and Drug Administration announced that rofecoxib use had been linked to seven cases of aseptic meningitis reported to the FDA’s Spontaneous Reporting System. Authors from the FDA published an article reporting five of the cases.\(^\text{12}\)

ii. **New case reports.** The May, 2003 searches identified 38 case reports involving the following adverse events:

   a. Celecoxib: anaphylaxis,\(^\text{13}\) fatal\(^\text{14}\) and nonfatal allergic vasculitis,\(^\text{15, 16}\) interstitial nephritis with\(^\text{17}\) and without\(^\text{18}\) nephritic syndrome, cholestatic hepatitis,\(^\text{19}\) toxic epidermal necrolysis,\(^\text{20-25}\) erythema multiforme,\(^\text{24}\) migratory pulmonary infiltrates,\(^\text{24}\) acute pancreatitis,\(^\text{26}\) torsade de pointes,\(^\text{27}\) and renal papillary necrosis.\(^\text{28}\)

   b. Rofecoxib: wrinkled palms,\(^\text{29}\) acute pancreatitis,\(^\text{30}\) acute colitis,\(^\text{31-33}\) cholestatic hepatitis,\(^\text{34}\) fatal\(^\text{35}\) and nonfatal hyperkalemia,\(^\text{36}\) fatal pulmonary hemorrhage,\(^\text{37}\) erythema multiforme,\(^\text{38}\) acute interstitial nephritis,\(^\text{39}\) and gynecomastia.\(^\text{40}\)

4. **Drugs.** We sought evidence about currently available coxibs (celecoxib [SC 58635], rofecoxib [MK 0966], valdecoxib), COX-2 selective NSAIDs (meloxicam, nambumetone), and nonselective NSAIDs.

### FDA Information

Since April, 2002 there have been several changes to FDA package insert information on the Cox-2 inhibitors.

Celecoxib. The most recent FDA review was posted in August, 2002. It approved changes in the product label based on review of the 9-month results of the
CLASS study. The FDA’s conclusions are appended to this report (pages 25 to 27 of 20-99S009_Celebrex_printbl.pdf). Of note, in their analysis of the 9-month CLASS trial data, FDA found no evidence that use of celecoxib was associated with an increased risk of coronary heart disease.

The usual recommended dose of celecoxib for osteoarthritis is 100 mg twice a day. The usual recommended dose for rheumatoid arthritis is 200 mg twice a day.

Rofecoxib. The most recent FDA review was posted in July, 2002 and the most recent product label is dated April 16, 2002. The revision (which was available to the subcommittee last year) included detailed information on the risk of cardiovascular events observed in the VIGOR trial. The recommended starting dose of rofecoxib for osteoarthritis is 12.5 mg once a day. The maximum recommended dose for osteoarthritis is 25 mg once a day. The recommended dose for rheumatoid arthritis is 25 mg once a day.

Valdecoxib was approved Nov. 16, 2001. The FDA review of this drug was posted May 30, 2002. Indications are “for the relief of the signs and symptoms of osteoarthritis, adult rheumatoid arthritis and for treatment of primary dysmenorrhea.” A warning was added to the product label in Nov, 2002. It was prompted by reports of cases of serious anaphylactic reactions and serious dermatologic adverse events in postmarketing surveillance.

The usual recommended dose of Valdecoxib for osteoarthritis and rheumatoid arthritis is 10 mg once a day. The recommended dose is 10 milligrams once a day.

METHODS

Literature Search

To identify articles relevant to each key question, we searched the Cochrane Database of Systematic Reviews (2002, Issue 1), EMBASE (1980 to December 2001), MEDLINE (1996 to April Week 3 2002), and Premedline. The Cochrane search identified high-quality systematic reviews of nonselective NSAIDs and of prevention of NSAID-induced ulcers. In EMBASE, we combined the terms Cyclooxygenase 2 inhibitors (Adverse Drug Reaction, Pharmacoeconomics, Clinical Trial, Drug Toxicity) with the term controlled study. The EMBASE search returned 311 citations. We repeated the EMBASE search using the terms meloxicam, nabumetone, and salsalate instead of cyclooxygenase inhibitor. In MEDLINE we searched on the terms rofecoxib, celecoxib, valdecoxib, and meloxicam (any field), limited to human studies in adults and relevant study design terms (clinical trial or controlled clinical trial or meta analysis or multicenter study or randomized controlled trial). The MEDLINE search returned 129 citations, 30 of which were not identified in the EMBASE searches. We also searched PREMEDLINE periodically for new citations, using the terms valdecoxib, rofecoxib, celecoxib, meloxicam, nembutamone, and salsalate. Other sources of citations were EULAR 2001 abstracts (which are published in searchable form at http://www.eular.org/eular2001/AbstractsOnline.cfm), and the reference lists of review articles and meta-analyses from Bandolier and other sources. Subcommittee members were invited to provide additional citations. Pharmaceutical manufacturers were invited to submit dossiers, including citations, using a protocol issued by the State of Oregon.
In May, 2003, we searched MEDLINE and EMBASE for articles published in 2002 or 2003 that were not cited in our original paper. The search used the names of each Cox-2 inhibitor and meloxicam, limited to human studies. (We did not search for new articles about salsalate.)

### Study Selection

We included randomized controlled trials of at least 4 weeks’ duration that compared a coxib or Cox-2 selective agent with an active control group. We excluded trials in healthy volunteers and those that had only a placebo control group.

In contrast with many other drug classes, clinical trials of COX-2 inhibitors are often designed to assess adverse events. We also identified several observational studies that assessed adverse event rates, but we excluded them as these offered no advantage over the randomized trials.

### Data Abstraction

One reviewer abstracted the following data from included trials: study design, setting, population characteristics (including sex, age, ethnicity, diagnosis), eligibility and exclusion criteria, interventions (dose and duration), comparisons, numbers screened, eligible, enrolled, and lost to followup, method of outcome ascertainment, and results for each outcome. We recorded intention-to-treat results if available.

### Validity Assessment

We assessed the internal validity (quality) of systematic reviews and randomized trials based on the predefined criteria listed in Appendix B, which were submitted to the 3 Health Resources Commission in December 2001. These criteria are based on those developed by the US Preventive Services Task Force and the National Health Service Centre for Reviews and Dissemination (UK).\(^{55,56}\) We rated the internal validity of each trial based on the methods used for randomization, allocation concealment, and blinding; the similarity of compared groups at baseline; maintenance of comparable groups; adequate reporting of dropouts, attrition, crossover, adherence, and contamination; loss to followup; and the use of intention-to-treat analysis. Trials that had a fatal flaw in one or more categories were rated poor quality; trials which met all criteria were rated good quality; the remainder were rated fair quality. As the “fair quality” category is broad, studies with this rating vary in their strengths and weaknesses: the results of some fair-quality studies are likely to be valid, while others are only probably valid. A “poor quality” trial is not valid—the results are at least as likely to reflect flaws in the study design as the true difference between the compared drugs. External validity of trials was assessed based on whether the publication adequately described the study population, how similar patients were to the target population in whom the intervention will be applied, and whether the treatment received by the control group was reasonably representative of standard practice. We also recorded the funding source and role of the funder.

Overall quality ratings for an individual study were based on ratings of the internal and external validity of the trial. A particular randomized trial might receive two
different ratings: one for efficacy and another for adverse events. The overall strength of evidence for a particular key question reflects the quality, consistency, and power of the set of studies relevant to the question.

RESULTS

Overview

The original searches and review of reference lists identified 377 publications, of which 66 were included in the review. Three reported trials directly comparing different COX-2 inhibitors; 10 concerned cardiac safety; five were systematic reviews; and the remainder were controlled trials or meta-analyses of controlled trials comparing rofecoxib, valdecoxib, celecoxib, or meloxicam to a nonselective NSAID. The searches identified 43 non-English articles, 22 of which concerned nimesulide.

The May, 2003 searches returned 361 citations that were not identified in the original searches. Of these, 55 were identified as possibly eligible for inclusion in the update. Of these, one appeared to be a publication of results from a trial directly comparing rofecoxib and celecoxib in patients with chronic arthritis (SUCCESS). One randomized trial of celecoxib versus diclofenac plus omeprazole reported clinical bleeding outcomes. No other new trials reported clinical bleeding outcomes, but a new report from an older trial of rofecoxib versus naproxen (VIGOR) was identified. Three trials compared valdecoxib to nonselective NSAIDs; none of these were designed to examine clinical bleeding outcomes as a primary endpoint. We also selected for further examination 8 review articles, 12 observational studies, and five publications concerning the cardiac effects of the Cox 2 inhibitors.

We excluded the following trials:
- 2 head-to-head trials involving pain from dental surgery that would not be included in an update.
- A head-to-head trial that found no difference between celecoxib, rofecoxib, and ibuprofen in efficacy for acute pain in the emergency room setting.
- A trial of valdecoxib versus naproxen for dysmenorrhea.
- A trial of the effects of rofecoxib versus celecoxib and naproxen on renal function in healthy elderly subjects.

In head-to-head comparisons, are there differences in efficacy or safety between different COX-2 inhibitors?

We found four published randomized, multicenter, fair-to-good quality trials that directly compared COX-2 inhibitors for osteoarthritis of the knee. Three of the studies, funded by the maker of celecoxib, found no difference in efficacy between rofecoxib 25mg and celecoxib 200mg, but found a higher rate of adverse effects with rofecoxib. The other study (VACT, for Vioxx Acetominophen Celecoxib Trial), conducted by the maker of rofecoxib, found that rofecoxib 25mg was more effective than celecoxib 200mg, with no differences in rates of adverse effects.
**Efficacy results.** Two of the trials appeared to enroll patients with similar demographics and baseline levels of pain (see table below).\textsuperscript{93-95} Both compared rofecoxib 25mg qd and celecoxib 200mg qd in patients with flare-ups of chronic osteoarthritis of the knee. Both were 6-week trials.

| Table 2. Comparison of rofecoxib and celecoxib in flare-ups of chronic osteoarthritis of the knee |
|-----------------|-------|-------|
| **Characteristic** | **McKenna** | **Geba** |
| Rofecoxib 25mg (n) | 59 | 95 |
| Celecoxib 200mg (n) | 60 | 97 |
| Aspirin 325 qd permitted | Yes | No |
| Mean age | 62 | 62.6 |
| Mean osteoarthritis duration | 10.5 years | 10 years |
| Percent white | 80% | 85% |
| Baseline pain on walking (score) | 72 | 72 |
| Discontinued trial by 6 wks: | | |
| Rofecoxib 25mg | 16% | 19% |
| Celecoxib 200mg | 22% | 17% |

Both studies were probably adequately randomized and blinded, and neither had statistically significant differences in baseline characteristics. However, in both studies there were some discrepancies. In McKenna, the proportion of patients with a past history of ulcers was higher for celecoxib (10% vs. 5%), and the proportion that had a past history of nonspecific GI symptoms was higher for rofecoxib (38% vs. 46%). The proportion of white patients was the same in the celecoxib and rofecoxib groups (84% vs. 85%), but was lower in the placebo group (73%). In Geba, the rofecoxib 25mg group had a higher proportion of women (72.6% vs. 64.9%) and a lower proportion of white subjects (82.1% vs. 87.6%) than the celecoxib 200mg group. The main article did not report the baseline WOMAC and global assessment scores of patients in the different treatment groups; a response to a letter to the editor states that the baseline WOMAC scores were similar.

Despite these similarities, the efficacy results of the studies differed. In McKenna, there were no differences between rofecoxib and celecoxib in pain relief or mobility (walking) as measured by the VAS or WOMAC scores. In Geba, however, for each component of the WOMAC score, rofecoxib 25mg had larger mean reductions than celecoxib 200mg. The differences were statistically significant for rest pain and night pain, but not for walking pain or morning stiffness. After 6 weeks, the proportion of patients reporting a good or excellent response was 60% in the rofecoxib 25mg group and 46% in the celecoxib group (p<0.05).

Geba and his colleagues noted that, regarding the WOMAC scores, "There is no current consensus on the magnitude of effects that is clinically important." A 1992 consensus conference found that a difference of 15 to 20 points on a VAS for pain and global disease activity was "clinically significant," but this has never been validated in clinical studies.\textsuperscript{96} A more recent analysis of data from randomized trials estimated that the minimal perceptible improvement for each WOMAC scale was 11 mm.\textsuperscript{51} In the Geba trial, WOMAC scores differed by 8 points or less between celecoxib 200mg and rofecoxib 25mg. 

**Adverse events and safety.** The differences between Geba and McKenna in adverse effects were striking. In McKenna, more rofecoxib than celecoxib patients experienced
at least one GI adverse effect (34% vs. 11%, p=0.004), most of which were mild and involved diarrhea, dyspepsia, or abdominal pain. In Geba, there were no differences in rates of adverse effects.

It is difficult to explain the discrepancy between the two trials in adverse GI event rates. One possible explanation is that McKenna's study may have included a broader spectrum of patients with respect to previous GI disease. In McKenna, patients who had previously had ulcer disease or other GI disorders were permitted to enter the study, and almost 50% had at least one GI disorder before the study began. It is not clear how many patients in Geba's study had a history of GI disorders. Another potentially important difference is that McKenna's study permitted aspirin 325mg qd and occasional use of acetaminophen, which had to be discontinued before assessments of clinical response. Geba's patients were permitted to use acetaminophen throughout the trial but were not permitted to use any aspirin.

There were no differences in the percentage of patients who discontinued therapy by 6 weeks in either study. In Geba, the proportion of patients who discontinued for lack of efficacy was 8.4% in the rofecoxib 25mg group and 9.2% in the celecoxib 200mg group (not significant).

The other two head-to-head trials (SUCCESS VI and SUCCESS VII) focused on adverse events in patients 65 or older who had osteoarthritis and well-controlled hypertension. Both of these 6-week trials compared rofecoxib 25mg qd to celecoxib 200mg qd.

In the first trial, nearly twice as many rofecoxib-treated patients (n=399) as celecoxib-treated patients (n=412) experienced edema (9.5% vs. 4.9%, p = 0.014). Systolic blood pressure increased significantly in 17% of rofecoxib- compared with 11% of celecoxib-treated patients (p = 0.032) at any study time point. Diastolic blood pressure increased in 2.3% of rofecoxib-compared with 1.5% of celecoxib-treated patients (p = 0.44). At week 6, the change from baseline in mean systolic blood pressure was +2.6 mmHg for rofecoxib compared with -0.5 mmHg for celecoxib (p = 0.007).

In this trial, there was an important baseline difference in the proportion of patients who took an ACE inhibitor for hypertension (40% for celecoxib-treated patients vs. 29% for rofecoxib-treated patients, p=0.002). Although not statistically significant, at baseline fewer celecoxib-treated patients had angina (16.3% vs. 19.8%) or or a history of myocardial infarction (8% vs. 9.3%). These differences cast doubt on the quality of the trial: unbiased randomization is unlikely to have resulted in a baseline difference this large.

In the second trial, the primary endpoints were 1) systolic blood pressure >140 mm Hg or elevation of 20 or more mm Hg from baseline 2) edema and 3) change from baseline in mean blood pressure. Aspirin use was permitted. The mean age of subjects was 73 years. At baseline, fewer celecoxib-treated patients had coronary disease (15.5% vs. 12.8%), but the proportions of patients taking different anti-hypertensive medications were similar. The baseline differences do not suggest that there was a major flaw in randomization.

A similar proportion of patients taking rofecoxib (53/543, 9.8%) and celecoxib (51/549, 9.3%) withdrew from the study because of adverse events, treatment failure, or other reasons. By 6 weeks, 14.9% of the rofecoxib patients and 6.9% of the celecoxib patients had systolic blood pressure elevations (p<0.0001). In a subgroup analysis, rofecoxib was more likely than celecoxib to cause systolic blood pressure elevation for
patients taking an ACE inhibitor or a beta blocker, with or without a diuretic.

Two head-to-head trials have been reported in abstract form but have not been published fully and could not be included in this review. One,\textsuperscript{97} funded by the maker of celecoxib, found that rofecoxib, but not celecoxib or naproxen, induced a significant increase in 24-hour ambulatory systolic blood pressure in diabetics taking ACE inhibitors. The other (VACT II), funded by the maker of rofecoxib, found no difference in clinically important hypertension adverse events for rofecoxib and celecoxib.\textsuperscript{98}

**Are there differences in efficacy between coxibs, COX-2 selective NSAIDs, nonselective NSAIDs, or other drugs?**

**Celecoxib vs. nonselective NSAIDS.** Nearly all of the available efficacy data comparing celecoxib to nonselective NSAIDs in osteoarthritis and rheumatoid arthritis come from a series of trials sponsored, designed, and analyzed by the maker of celecoxib. These trials were large (536 to 1,214 subjects) and were mostly short-term. As shown in the table, not all of these trials have been published:

<table>
<thead>
<tr>
<th>NSAID</th>
<th>Number of trials</th>
<th>Number published fully in peer-reviewed literature</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>naproxen</td>
<td>6</td>
<td>3</td>
<td>6-12 weeks</td>
</tr>
<tr>
<td>diclofenac</td>
<td>3</td>
<td>2</td>
<td>6-24 weeks</td>
</tr>
<tr>
<td>ibuprofen</td>
<td>1</td>
<td>0</td>
<td>12 weeks</td>
</tr>
</tbody>
</table>

The lack of complete reporting raises concern about publication bias. For example, of the two trials that compared celecoxib to naproxen in patients with osteoarthritis, the one that found celecoxib and naproxen to be equally effective has been published.\textsuperscript{99, 100} The unpublished trial, which included 1191 subjects, found that naproxen was superior to celecoxib at 2 and 12 weeks.\textsuperscript{42} Data from the unpublished studies have been reported, but not in detail, in several meta-analyses sponsored by the manufacturer of celecoxib.\textsuperscript{42-47} After reviewing data from all the trials, the FDA found no difference in efficacy between celecoxib and nonselective NSAIDs.\textsuperscript{101} These data are not completely available for critical appraisal, however.

In addition to the five published trials in the series, there were two others. The CLASS trial had a total of 7,968 patients randomized to celecoxib, ibuprofen, or diclofenac.\textsuperscript{102} CLASS focused on adverse effects rather than efficacy. A higher proportion of NSAID patients withdrew for lack of efficacy (14.8\% vs. 12.6\%, p=0.005), but no other efficacy results were reported. In the other trial, there were no statistically significant differences between celecoxib and ketoprofen in pain control or function in 246 patients with ankylosing spondylitis.\textsuperscript{103}

The seven published trials are summarized in Table 4.\textsuperscript{99, 100, 104-108} In one of the studies (Emery), randomization did not appear to result in equivalent groups; this study was rated fair-to-poor while the others were rated good-quality. The published trials provide good overall evidence that celecoxib is equivalent to nonselective NSAIDs in efficacy for osteoarthritis and rheumatoid arthritis.

One meta-analysis of trials of celecoxib versus NSAIDs focused on efficacy in
elderly patients.\textsuperscript{45} Celecoxib 200mg and 400mg and naproxen 1000mg were similar in

efficacy.

SUCCESS-1, a randomized trial of 13,274 patients with osteoarthritis, compared
celecoxib to diclofenac and naproxen. The trial is not yet published, but the authors
reported in an abstract that there were no differences in efficacy.\textsuperscript{109-111}

\textbf{Rofecoxib vs. nonselective NSAIDs.} We were unable to determine whether
all manufacturer-sponsored trials of rofecoxib versus nonselective NSAIDs have been
published. The results of nine published trials are summarized in Table 5, where they are
sorted by length of followup.\textsuperscript{10, 11, 112-117} All but one of the trials included osteoarthritis patients,
and all but one\textsuperscript{117} were supported by the manufacturer of rofecoxib. The published trials provide
good overall evidence that rofecoxib is equivalent to nonselective NSAIDs in efficacy for OA. In
addition, one large, good-quality trial indicates that rofecoxib is equivalent to nonselective
NSAIDs in efficacy for rheumatoid arthritis.

\textbf{Valdecoxib vs. nonselective NSAIDs.} In clinical trials submitted to the FDA,
valdecoxib was as effective as ibuprofen (800mg 3 times/day), diclofenac (75mg twice
daily), and naproxen (500mg twice daily) in treating osteoarthritis symptoms. One
published trial found no difference in efficacy between valdecoxib and naproxen.\textsuperscript{118} At the time
of this update, 2 more of the valdecoxib efficacy trials have been published.\textsuperscript{60, 61} In these trials,
various doses of valdecoxib were as efficacious as naproxen 500 mg twice a day.

\textbf{Coxib vs. selective NSAID.} One multicenter trial compared the efficacy of a
Cox-2 inhibitor (rofecoxib 12.5mg qd or 25mg qd) to nabumetone, a Cox-2 selective
NSAID\textsuperscript{119} in patients age 80 or older with knee or hip arthritis. This 6-week, fair-quality
randomized trial was interesting because many patients had other medical conditions, such as
hypertension and heart disease. A total of 341 patients were enrolled: 118 received rofecoxib
12.5mg and 115 received nabumetone 1500mg. The number of subjects who received rofecoxib
25mg was too small (n=56) to estimate the rate of adverse effects meaningfully. About 14% of
patients receiving active drug did not complete the study; the rates were similar for rofecoxib (both doses) and nabumetone.
There were no differences between the two doses of rofecoxib and nabumetone in
symptom relief as judged by the VAS for pain, global assessment of disease function, and
WOMAC scores for pain, physical function, stiffness, joint tenderness, or acetaminophen
use.

\textbf{Selective NSAID vs. nonselective NSAID.} In double-blinded trials of
meloxicam 7.5mg or 15mg versus nonselective NSAIDs (Table 6) there were generally
no differences in efficacy.\textsuperscript{120-127} In two of the trials, however, patients taking
nonselective NSAIDs were significantly less likely to withdraw due to lack of efficacy
than patients taking meloxicam.\textsuperscript{122, 127}

\textbf{Nonselective NSAID vs. nonselective NSAID.} Several recent good-quality
systematic reviews by the Cochrane Collaboration found no clear differences among
nonselective NSAIDs in efficacy for treating knee,\textsuperscript{128} back,\textsuperscript{2} or hip pain.\textsuperscript{129} These reviews did not
include the COX-2 inhibitors.
Are there clinically important differences in safety or adverse effects between coxibs, COX-2-selective NSAIDs, nonselective NSAIDs, and the combination of a nonselective NSAID plus antiulcer medication?

**Significant GI events (GI bleeding, hospitalization for GI bleeding, symptomatic ulcer disease, perforation of the GI tract, and death).**

**Coxibs.** Three trials were sufficiently large to evaluate complications of peptic ulcer disease as a primary endpoint. The CLASS trial evaluated celecoxib versus ibuprofen and diclofenac; the VIGOR trial evaluated rofecoxib versus naproxen. The third trial compared celecoxib to the combination of diclofenac plus omeprazole in patients who presented with a bleeding ulcer. There are no trials of valdecoxib in this category.

**VIGOR (Vioxx Gastrointestinal Outcomes Research) Trial.** VIGOR, a randomized, double-blind trial, compared twice the highest recommended dose of rofecoxib to naproxen 500 mg twice a day in 8,076 patients with rheumatoid arthritis. VIGOR found a statistically significant reduction in complicated upper GI events (see Table 7). The number-needed-to-treat (NNT) to prevent a complicated, confirmed upper gastrointestinal event was 191.

VIGOR met all but one of the criteria for a good-quality study. The one weakness was the number of subjects who had incomplete followup. VIGOR was designed to be a 13-month study, but half of the patients were followed up for 9 months or less, and only about 1,000 patients (13%) were followed up longer than 10 months. By 13 months, about 29% of the subjects had discontinued the study drugs. Similar proportions discontinued naproxen or rofecoxib because of an adverse event (naproxen—16.1%, rofecoxib—16.4%).

In 2003, the VIGOR investigators published a post hoc analysis of lower GI events, defined as bleeding with a 2 g/dL drop in hemoglobin or hospitalization, or hospitalization for perforation, ulceration, diverticulitis, or obstruction. There were 11 events in the rofecoxib group and 24 events in the naproxen group (p<0.001, NNT=309.)

**CLASS (Celecoxib Long-term Arthritis Safety Study).** CLASS combined two randomized trials: one compared celecoxib 400 mg twice a day with ibuprofen 800 mg three times a day, and the other compared celecoxib 400 mg twice a day with diclofenac 75 mg twice a day. The main publication of the study in JAMA reported only 6 months of data and combined the ibuprofen and diclofenac results. Subsequently, additional details of the study have been made public on the FDA web site. The findings reported on the FDA site suggest that, as reported in the JAMA publication, the results of CLASS are incomplete and, in part, misleading.

There were 3,987 subjects in the celecoxib group and 3,981 subjects in the NSAID groups. For the primary endpoint of CLASS, confirmed serious ulcer complications, by 6 months there were more events in the celecoxib group than in the NSAID groups, but the difference was not statistically significant. Moreover, by 12 months, according to FDA documents, there was no longer a trend favoring celecoxib (see Figure 4, Shieman review).

Twenty per cent of the patients in the CLASS trial took aspirin in addition to their study drug. When patients taking aspirin were excluded from the analysis, there were fewer confirmed serious ulcer complications in the celecoxib group than in the
ibuprofen group (p=0.03). However, celecoxib was less likely to cause serious ulcer complications than ibuprofen, but was equivalent to diclofenac.\textsuperscript{130} In summary, celecoxib was unable to demonstrate a statistically significant advantage over either diclofenac or ibuprofen for the primary endpoint. Celecoxib was superior to ibuprofen, but not diclofenac, in the subgroup of subjects not taking aspirin.

The incidence of serious ulcer complications in CLASS was much higher than it had been in previous trials of celecoxib. A meta-analysis examined the endpoint of "UGI ulcer complications" in 14 RCTs of celecoxib versus placebo or nonselective NSAIDs (usually naproxen).\textsuperscript{44} The trials ranged in duration from 2 to 24 weeks, with most lasting 6 or 12 weeks. The strength of this meta-analysis was that the endpoint was similar to those used in the VIGOR and CLASS trials. The endpoint consisted of upper GI bleeding with endoscopic findings of an ulcer or large erosion, perforation, or gastric outlet obstruction. Potential ulcer complications were adjudicated by a Safety Committee in a blinded manner. These endpoints were ascertained through a monitoring program which appears to have been superimposed on all of the trials; it is not clear how assiduously investigators complied with this program. As mentioned above, not all of these trials have been published, and their quality was not assessed as part of the meta-analysis.

In the 14 trials, there were 2/6,376 UGI ulcer complications in the celecoxib group (3 per 10,000) and 9/2,768 in the NSAIDs group (33 per 10,000) and none in the placebo group (0/1,864). This corresponded to annual rates of 2 per 1,000 per year for celecoxib and about 17 per 1,000 per year for NSAIDs (p=0.002). In the CLASS trial, the annualized rate of serious ulcer complications was 7.6 per 1,000 per year for celecoxib and 14.5 per 1,000 per year for the two NSAIDs combined.

There are several possible reasons why rofecoxib, but not celecoxib, was found to reduce ulcer complications significantly. Patient populations and features of the study designs differed (Table 7). VIGOR included patients with rheumatoid arthritis over 50 years, while CLASS had a broader age range of patients with either osteoarthritis or rheumatoid arthritis. VIGOR prohibited the use of aspirin while CLASS did not. The rate of ulcers in the patients taking a control drug was almost 3 times as high in VIGOR as in CLASS, although rates of ulcer complications were similar.

In the third major trial, conducted in Hong Kong, patients who presented with a bleeding ulcer were randomized to the combination of celecoxib 200 mg twice a day plus placebo once daily (n=144) or to the combination of extended-release diclofenac 75 mg twice a day plus omeprazole 20 mg once daily (n=143). The primary endpoint was recurrent ulcer bleeding within six months.

By six months, there were 16 cases of recurrent bleeding: 7 (4.9%) in the celecoxib group and 9 (6.3%) in the diclofenac+omeprazole group (the difference, -1.5%, CI –6.8% to 3.8%, was not significant).\textsuperscript{58} There were also no differences between the groups in other adverse events, including GI symptoms and renovascular complications. Overall 13.3% of celecoxib patients and 11.9% of diclofenac+omeprazole patients had discontinued medications because of adverse events or lack of efficacy (not significant.) It is not possible to determine whether (or by how much) celecoxib reduced the incidence of recurrent bleeding compared to what would be expected from using diclofenac alone. The high rates of recurrent bleeding in both the celecoxib-treated patients and in the diclofenac+omeprazole group—over 10 times as high as the rate in the CLASS trial—suggest that NSAIDs and coxibs should be used with caution, if at all, in patients who have a recent history of a bleeding ulcer.
**COX-2 selective NSAIDs.** Evidence that meloxicam and nabumetone prevent ulcer complications is weaker than that for coxibs. Meta-analyses have been performed of published and unpublished trials both of these agents. For nabumetone, a fair-quality meta-analysis included six nonendoscopic studies (five published and one abstract), the largest of which had 3,315 nabumetone patients and 1,096 NSAID patients. The studies had 3 to 6 months of followup. The main endpoint used in this meta-analysis was "PUB", meaning perforation, symptomatic ulcer, or bleeding. Because of the inclusion of symptomatic ulcer, this is a broader endpoint than that used in the VIGOR and CLASS trials or in the meta-analysis of celecoxib trials. The methods to ascertain the endpoint (that is, how well and consistently investigators identified complications) is unknown. There was one PUB event among 4,098 patients taking nabumetone versus 17 events among 1,874 nonselective NSAID patients; this was highly statistically significant. The rates per 1,000 patients per year were about 2 versus 6. (For comparison, in a similar meta-analysis of rofecoxib studies, the rates of PUBs per 1,000 patients per year were 13 for rofecoxib and 26 for NSAIDs.) There was also a significant reduction in treatment-related hospitalizations in the nabumetone group (6.4 per 1,000 patients per year vs. 20.3 per 1,000 patients per year).

The meloxicam (7.5mg or 15mg) meta-analysis included seven double-blind trials, one single-blind trial, and two unblinded trials. Most of the patients were followed for only 4 weeks. The main endpoint was "PUBs". The meta-analysis did not provide event rates, but reported that the rate of PUBs was reduced in the meloxicam patients (Odds ratio 0.52, 95% CI 0.28-0.96).

Because of the lack of detail about the quality and results of the included studies, and lack of a more stringent endpoint than PUBs, this meta-analysis provides insufficient evidence to make any judgment about the safety of meloxicam. While the nabumetone meta-analysis is somewhat better, the results are unlikely to apply to actual practice: as for the celecoxib meta-analysis, the rates of events with nabumetone are probably underestimated, and the similarity of the subjects in the efficacy trials to a broader group of NSAID users was not addressed.

**Misoprostol, PPIs, and H2 blockers.** One good-to-fair-quality double-blind, randomized trial (MUCOSA) of 8,843 patients found that misoprostol 800µg/day prevented symptomatic ulcers and ulcer complications among patients taking nonselective NSAIDs. As in the VIGOR trial, patients were older (> age 50) and had rheumatoid arthritis. Both the misoprostol and placebo groups took NSAIDs. Misoprostol reduced the frequency of gastric outlet obstruction and perforation but not the rate of serious bleeding. The NNT to prevent one serious ulcer complication was 263. Misoprostol was associated with high rate GI adverse effects, such as diarrhea, nausea, dyspepsia, and flatulence. These side effects were not correlated with serious GI events but led to a significantly higher rate of discontinuation of the drug than NSAID plus placebo (42%, vs. 36% in the placebo group).

Proton pump inhibitors (PPIs) and H2-receptor antagonists have also been used to prevent GI side effects in patients taking NSAIDs. A Cochrane review summarized four placebo-controlled trials of PPIs and seven placebo-controlled trials of H2-receptor antagonists. These trials provide strong evidence that PPIs and double-dose H2-receptor antagonists reduce the risk of endoscopic gastric and duodenal ulcers in patients taking NSAIDs, but the total number of subjects was too low to examine whether symptomatic ulcers or ulcer complications are reduced. Unlike misoprostol, PPIs and double-dose H2-receptor blockers also
reduced GI symptoms.

No head-to-head comparisons of high-dose H2-receptor blockers vs. PPIs have been done. Since the Cochrane review was last updated, a trial comparing lansoprazole 15 or 30mg/day and misoprostol 800µg/day in patients who had a history of NSAID-induced ulcer has been published. The outcomes were endoscopic ulcer and withdrawal due to adverse events. Withdrawals were much higher for misoprostol, due primarily to diarrhea. On an intention-to-treat basis, misoprostol and lansoprazole were equally efficacious in preventing recurrent endoscopic ulcers. This was because the higher rate of withdrawals was balanced by a lower rate of endoscopic ulcers among those who continued to take misoprostol. The incidence of GI symptoms and the amount of antacid use was significantly higher in the misoprostol group.

**Cardiac risk.** The main publication of the VIGOR trial reported that “the incidence of myocardial infarction was lower among patients in the naproxen group than among those in the rofecoxib group (0.1 percent vs. 0.4 percent; relative risk, 0.2; 95 percent confidence interval, 0.1 to 0.7); the overall mortality rate and the rate of death from cardiovascular causes were similar in the two groups.” This corresponds to 1 additional heart attack for every 333 patients treated with rofecoxib instead of with naproxen. Blinded re-review of the VIGOR trial data classified 45/4047 (1 in every 90) rofecoxib patients and 20/4029 (1 in 201) naproxen patients as having serious thrombotic events (heart attack, stroke, unstable angina, transient ischemic attack, resuscitated cardiac arrest, and sudden death). This corresponds to one additional serious thrombotic event for every 162 patients taking rofecoxib. For patients who had an indication for aspirin prophylaxis, rofecoxib patients were 4.89 (95% CI, 1.41-16.88) times as likely to have a cardiovascular event than those who took naproxen. For the other, lower-risk subjects in VIGOR, rofecoxib patients were 1.89 (1.03 - 3.45) times as likely as naproxen patients to have a serious cardiovascular event.

An independent review article re-examined the findings of VIGOR, CLASS, and two unpublished trials of rofecoxib that reported cardiovascular outcomes. The latter three trials all permitted patients to use low-dose aspirin, and none of them found an increase in cardiovascular events. Both rofecoxib patients (three trials) and celecoxib patients (CLASS) had annual rates of heart attacks higher than the average for the patients in several placebo-controlled studies of aspirin prophylaxis.

In October 2001, an article published in *Circulation* reported a pooled analysis from 23 rofecoxib Phase IIb through V trials conducted by Merck. The investigators examined results by patient group (rheumatoid arthritis, osteoarthritis, or Alzheimer’s disease) and by control group (placebo, naproxen, or non-naproxen NSAID). The risk of cardiovascular events was 1.69 times higher for rofecoxib than for naproxen, but was not elevated in trials comparing rofecoxib versus placebo or other NSAIDs. The authors hypothesized that rofecoxib might have been an “innocent bystander” in the VIGOR trial; that is, rather than rofecoxib increasing the rate of cardiovascular events, naproxen might have reduced it. They also criticized the Mukherjee article, correctly in our view, for comparing trials of patients under treatment for rheumatoid arthritis to an entirely different group of patients in studies of aspirin prophylaxis.

The problem with the Konstam analysis is that the non-naproxen studies and naproxen studies are not directly comparable. VIGOR, the only COX-2 trial to demonstrate a significant reduction in serious GI events, used a supratherapeutic dosage of rofecoxib (50mg), prohibited aspirin, and followed patients for 9 months. All but one of the non-naproxen-controlled studies were shorter than 6 weeks in duration or used lower doses of rofecoxib. The one exception is a combination of data from eight phase IIb/III trials in osteoarthritis patients (see below). The data
presented in the meta-analysis are also inadequate to judge the quality of the included studies and how concomitant aspirin use might have affected rates of cardiovascular events.

A subsequent meta-analysis provided a more detailed analysis of the subjects in the eight phase IIb/III trials mentioned in the Konstam analysis. The total number of subjects in the eight trials is given as 5,435, versus 5,505 in the Konstam analysis; the reason for the discrepancy is unclear and the second analysis does not provide a detailed accounting of the excluded subjects. The mean duration of treatment was 3½ months.

The conclusion of the analysis—that there were no significant differences between rofecoxib and placebo or non-naproxen NSAIDs—is valid for this set of studies, but does not address the more general question of whether rofecoxib is safe at the dosage proven to reduce serious GI events. The analysis combined data from all rofecoxib doses (12.5, 25, and 50mg/day); only 545 of the patients received the 50mg/day dose. The issue of dosage is important because only the 50mg dose has been shown to prevent GI adverse events. It is possible that lower doses do not increase cardiovascular events compared with non-naproxen NSAIDs, but the benefit of lower, conventional doses are uncertain.

The original publication of the CLASS trial found that celecoxib had no effect on the rate of cardiovascular events compared with diclofenac and ibuprofen. The CLASS data on thrombotic events were recently analyzed in more detail. There were no differences in the rates of any significant cardiovascular event for the overall sample and for the subgroup who did not use aspirin. In their analysis, White and colleagues pooled the data from the ibuprofen and diclofenac groups. In fact, more detail is needed about the duration of the CLASS trial to judge the validity of these results. While the original publication reported results at 6 months, a concurrent publication by the Director of Clinical Research at Pfizer stated that "CLASS…consisted of two trials lasting 1 year." The recent analysis by White seems to contradict this statement, saying: "After all patients had the opportunity to participate for at least 6 months, a monthly blinded review of ulcer complications by the CLASS oversight committees indicated a marked decrease and then cessation in the accrual rate. After two successful reviews with no further complications noted, the oversight committee voted to terminate the study." This may mean that the study was terminated after all patients had the opportunity to be followed for 8 months, which may not be long enough to assess cardiovascular risk. At 8 months in the VIGOR trial there was no significant difference between rofecoxib and naproxen in the cumulative incidence of events. From 8 to 12 months, the incidence of events in the rofecoxib group rose sharply (Figure 1 of Mukherjee), while that of naproxen did not. Based on this pattern in VIGOR, if celecoxib were associated with an increased risk of cardiovascular events, it would probably not be seen until 10 or 12 months of followup. In the VIGOR trial, 2,140 subjects, about one-fourth of the original sample, were available for 10 months of followup, and 1,045 were available for 12 months. Comparable data are not available for CLASS, so it is difficult to know whether the analysis had sufficient power to detect a difference.

The authors of the study argue that their results show that celecoxib is safer than rofecoxib. To support this point, they note that the annualized rate of all cardiovascular thromboembolic events in the naproxen group in the VIGOR trial and the non-aspirin users in the CLASS trial were similar. However, this comparison of rates across the VIGOR and CLASS studies is imprecise. After 8 months in the VIGOR trial, about 0.4% of naproxen patients had experienced an event; after 8 months in CLASS, about 0.8% of non-aspirin users had. It is not clear whether or not this is a clinically or statistically significant difference.
New clinical information about the cardiac effects of the coxibs falls into 2 categories: new analyses of the CLASS and VIGOR data, and new data from observational studies.

**New analyses of the CLASS and VIGOR data.** One meta-analysis from Canadian investigators used FDA materials to analyze the rates of serious adverse events, defined as death, hospitalization, “any life-threatening event, or event leading to severe disability.” This measure combines the rates of serious upper GI complications (in which coxibs are expected to have an advantage over NSAIDs) with other serious adverse events. The numbers of all serious adverse events shown in the following table were drawn directly from FDA materials, pages 7 and 57.

In the Canadian re-analysis, shown in the table, the rates are calculated using the number of patients as the denominator. These simple rates are compared with the number of serious upper GI events, which constitute only about 10% of all serious adverse events (the 2 columns to the right in the table). Using the data from the table, the number-needed-to-harm 1 person was 82 for celecoxib vs. diclofenac, 129 for celecoxib vs. ibuprofen, 100 for celecoxib vs. diclofenac and ibuprofen, and 65 for rofecoxib vs. naproxen. The Canadian authors pooled the results for celecoxib and rofecoxib, assigning more weight to VIGOR, which had a longer duration than CLASS. In the pooled analysis, the number needed to harm was 78 and was statistically significant.

<table>
<thead>
<tr>
<th>Trial</th>
<th><strong>ALL SERIOUS ADVERSE EVENTS</strong></th>
<th><strong>SERIOUS UPPER GI EVENTS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td>Control</td>
</tr>
<tr>
<td>CLASS</td>
<td>270/3987 (6.8%)</td>
<td>230/3981 (5.8%)</td>
</tr>
<tr>
<td>VIGOR</td>
<td>378/4047 (9.3%)*</td>
<td>315/4029 (7.8%)</td>
</tr>
</tbody>
</table>

*statistically significant vs. control group.

For the VIGOR trial, the FDA calculated rates of serious adverse events in exactly the same manner as the Canadian investigators. The FDA noted that the rates of each serious adverse event (except upper GI adverse events) were higher for rofecoxib than for naproxen. For the CLASS trials, the FDA used normalized patient-years as the denominator instead of a simple proportion to calculate rates of serious adverse events. (This approach was used because the 2 trials that make up CLASS had different durations.) In the FDA analysis, the rates of all serious adverse events combined were 11.6 per 100 patient-years for celecoxib; 10.3 per 100 patient-years for diclofenac, and 10.6 per 100 patient-years for ibuprofen. The FDA interpreted these differences as being insignificant.

In summary, the FDA data clearly show that these 2 coxibs, in doses higher than those used in practice, do not reduce the overall rate of serious adverse events, and may have increased them. It should also be noted that not all serious adverse events are equal in importance to patients and physicians. A reduction in the rate of one kind of adverse event might be considered more important than an increase in another one.

**New data from observational studies.** Two retrospective cohort studies focused on the cardiovascular complications of coxibs. In both studies, rofecoxib 25 mg or less, naproxen, and celecoxib were not associated with an increased rate of cardiac events. In one of the studies, rofecoxib 50 mg, the dose studied in the VIGOR trial, was associated with an increased risk of cardiac events.
One of these examined a database from a Medicaid population in Tennessee. In an earlier publication, the investigators had examined the effects of different NSAIDs on cardiac events in the era before coxibs became available. In the more recent article, the authors added rofecoxib and celecoxib use to their analysis. The investigators compared rates of cardiovascular events within one year of starting one of several NSAIDs or one of the 2 coxibs, versus not taking any NSAID or coxib, as identified by pharmacy benefit claims. The unit of analysis was the patient-year: over the course of the study, a particular patient could be in the same group several times, or in several different groups. For example, if a patient had the following history:

Year 1 No NSAID or coxib
Year 2 No NSAID or coxib
Year 3 Took rofecoxib
Year 4 No NSAID or coxib
Year 5 Took celecoxib

she could be in the control group for year 2, the rofecoxib group for year 3, the control group again for year 4, and the celecoxib group for year 5. The investigators attempted to adjust for baseline cardiovascular risks using a patient’s inpatient diagnostic codes and medication history.

The subjects mean age was 61 years. Approximately 42% of rofecoxib and celecoxib users had a history of treatment for major cardiovascular diseases, as estimated by the claims information. The main findings of the study was an increased adjusted 1-year risk of a cardiovascular event (death or myocardial infarction) in patients taking more than 25 mg of rofecoxib daily. Rates for patients taking ibuprofen, naproxen, rofecoxib 12.5 mg or 25 mg daily, or celecoxib were not different from the rate in nonusers of NSAIDs or coxibs.

The other study, from Ontario, Canada, used similar methods. It found no difference in 1-year rates of cardiovascular events between nonusers of NSAIDs and patients taking various NSAIDs, rofecoxib, or celecoxib. It did not specifically examine patients taking doses of rofecoxib greater than 25 mg daily.

The strength of these studies is that they reflect how coxibs are actually used in practice. The studies have important weaknesses. Perhaps the most serious weakness of these two observational studies is that they did not report serious GI events. Doing so would have provided a sense of the balance of benefits and harms of using the coxibs in these populations.

Inadequate data were available to control for confounders in a convincing manner. Neither study could control for (or estimate the degree of use of) over-the-counter aspirin and ibuprofen. The lack of information about aspirin use makes it even more important to report the frequency of GI adverse events, and the methods used to estimate baseline cardiovascular risk were crude.

The clinical information available from claims was insufficient and, possible, unreliable. For example, in the earlier of the two Tennessee reports, 22% of NSAID users and 22% of the nonuser controls were described as having “serious cardiovascular disease in the past year” at baseline. In the 2nd report, 43% of coxib users, 38% of NSAID users, and 34% of nonusers were described as having “major cardiovascular disease” in the past year by similar criteria. In the first report, 67% of NSAID users took a cardiovascular drug in the previous year; in the 2nd report, 82% did. In the 2nd report,
although only 3 to 5% of the subjects had rheumatoid arthritis, 30% of the patients taking coxibs, 24% of the patients taking naproxen or ibuprofen, and 16% of the controls had been prescribed oral corticosteroids within the past year. These examples underscore that, when a study relies on claims data, it may not be able to provide a meaningful, consistent sense of the type, severity, and frequency of cardiovascular morbidity and comorbidity in the compared groups.

**Tolerability.** Table 9 summarizes tolerability information from published randomized trials. There was no difference between celecoxib versus nonselective NSAIDs or rofecoxib versus nonselective NSAIDs in withdrawals due to adverse events. In general total adverse events were similar for coxibs and nonselective NSAIDs, but celecoxib was better than ibuprofen 800mg tid or diclofenac 75mg bid in the largest celecoxib study.\(^{102}\) GI adverse events, primarily abdominal pain, diarrhea, and nausea, were consistently lower for coxibs than for nonselective NSAIDs, even in studies which permitted the use of H2 blockers in the NSAID group. In most studies these adverse events are described as mild or moderate because they did not result in discontinuation of the drug. The magnitude of difference is probably one GI adverse event for every 20 patients treated. Coxibs may be associated with a lower incidence of anemia than nonselective NSAIDs, but because only a few studies report this outcome publication bias cannot be ruled out.

The effects of celecoxib on renal function were reviewed in a meta-analysis of manufacturer’s data; the overall incidence was similar to that of nonselective NSAIDs.\(^{47}\) In the CLASS trial, there was one fewer episode of edema, hypertension, or increased creatinine for every 62 patients treated with celecoxib instead of ibuprofen 800mg tid or diclofenac 75 bid. Data on renal events of rofecoxib are less well-reported, there were no differences in the studies that reported these events, but many studies did not report them. As discussed earlier, one head-to-head trial found higher rates of these complications with rofecoxib than with celecoxib.\(^{92}\)

There is some evidence that meloxicam (7.5mg or 15mg) is better tolerated than nonselective NSAIDs. The meta-analysis of meloxicam studies mentioned earlier found lower rates of any gastrointestinal event, dyspepsia, and withdrawals due to GI events compared with NSAIDs, but as mentioned before it included some inadequately blinded studies; only blinded studies are reliable for assessing withdrawals and attributing the cause of adverse events. In the nabumetone meta-analysis, the incidence of GI adverse events was significantly different (25.3% vs. 28.2%, p=.007), corresponding to about 1 fewer event for every 34 patients treated with nabumetone.

**Are there subgroups of patients based on demographics, other medications, or co-morbidities for which one medication is more effective or associated with fewer adverse effects?**

An original data meta-analysis of three celecoxib trials found that, in the elderly, celecoxib 200mg/day or 400mg/day and naproxen 1,000mg/day were similar in WOMAC scores and SF-36 scores.\(^{45}\) For the SF-36, there were no statistically significant differences: naproxen scored better than celecoxib 200mg on four of 10 components of the SF-36, while celecoxib 200mg scored better on six, including general health. Celecoxib 200mg was significantly better than placebo on nine of the 10 components, while naproxen was significantly better than placebo on seven. The study confirmed that, as discussed above, the overall incidence of GI adverse events was lower
with celecoxib; the difference was about one event in 20 patients for celecoxib 200mg and one in 10 for celecoxib 400mg.

In most of the published trials, a majority of subjects were women. We did not find any publications focusing on the differential efficacy or safety of coxibs in African-Americans, Hispanics, or other ethnic minorities.

**SUMMARY**

The table below summarizes the strength of evidence and results for each key question. Publication bias is an issue for all of these questions, because we do not know the complete results of unpublished trials submitted to the FDA or trials that have been done but not published or submitted to the FDA.
<table>
<thead>
<tr>
<th>Key Question</th>
<th>Level of Evidence</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. In head-to-head comparisons, are there differences in efficacy or safety between different COX-2 inhibitors?</td>
<td>Overall grade: fair. Grade for valdecoxib: poor. There are no head-to-head comparisons involving valdecoxib.</td>
<td>The trials had conflicting results on efficacy and on GI adverse effects. Two trials found that in patients 65 years or older, rates of edema and hypertension were higher for rofecoxib than for celecoxib (Number-needed-to-harm (NNH) was 21 for edema and 26 for hypertension).</td>
</tr>
<tr>
<td>2. Are there differences in efficacy between coxibs, COX-2 selective NSAIDs, and nonselective NSAIDs?</td>
<td><strong>a. celecoxib or rofecoxib vs. nonselective NSAIDs</strong></td>
<td>No difference.</td>
</tr>
<tr>
<td></td>
<td>Good. Consistent evidence from many published trials.</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>b. valdecoxib</strong></td>
<td>No difference.</td>
</tr>
<tr>
<td></td>
<td>Fair. (1 published and several unpublished trials)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>c. coxib vs. selective NSAID</strong></td>
<td>No difference.</td>
</tr>
<tr>
<td></td>
<td>Fair. (1 published and several unpublished trials)</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>d. selective NSAID vs. nonselective NSAID</strong></td>
<td>No difference.</td>
</tr>
<tr>
<td></td>
<td>Good. Consistent evidence from many published trials</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>e. nonselective NSAID vs. nonselective NSAID</strong></td>
<td>No difference.</td>
</tr>
<tr>
<td></td>
<td>Good. Consistent evidence from trials included in several good-quality systematic reviews</td>
<td></td>
</tr>
<tr>
<td>3. Are there clinically important differences in safety or adverse effects between coxibs, COX-2 selective NSAIDs, nonselective NSAIDs, and the combination of a nonselective NSAID plus antiulcer medication?</td>
<td>Good for rofecoxib (1 fair quality RCT)</td>
<td>Rofecoxib significantly reduced symptomatic ulcers and serious ulcer complications in patients with RA. NNT was 62 to prevent one symptomatic ulcer and 191 to prevent one serious complication.</td>
</tr>
<tr>
<td>a. Significant GI events (GI bleeding, hospitalization for GI bleeding, symptomatic ulcer disease, perforation of the GI tract, and death).</td>
<td>Fair for celecoxib (subgroup analysis in an RCT; fair quality meta analysis)</td>
<td>Celecoxib did not significantly reduce ulcer complications in a large, fair-to-poor quality trial, but did reduce them in the subgroup of patients who did not take aspirin.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>b. Cardiac risk.</td>
<td>Insufficient evidence for valdecoxib.</td>
<td>Valdecoxib has not been shown to reduce ulcer complications.</td>
</tr>
<tr>
<td>c. Tolerability: discontinuations due to adverse effects</td>
<td>Fair-poor for nabumetone.</td>
<td>Nabumetone reduced PUBs and hospitalizations in a fair-quality meta-analysis, but the results are unlikely to apply to practice.</td>
</tr>
<tr>
<td>c.2. Tolerability: total adverse effects</td>
<td>Insufficient for meloxicam.</td>
<td>Meloxicam reduced PUBs poor quality meta-analysis; the results are unlikely to apply to practice.</td>
</tr>
<tr>
<td></td>
<td>Good for misoprostol (1 good-quality RCT)</td>
<td>Misoprostol prevented serious ulcer complications (perforations and gastric outlet obstruction, but not bleeding); NNT 263.</td>
</tr>
<tr>
<td></td>
<td>Good for omeprazole vs. celecoxib. No data for high-dose H2 blockers.</td>
<td>The combination of a PPI plus diclofenac was equivalent to celecoxib. High-dose H2 blockers were associated with fewer endoscopic ulcers, but there are no data regarding serious ulcer complications.</td>
</tr>
<tr>
<td>b. Cardiac risk.</td>
<td>Fair-quality for rofecoxib.</td>
<td>Overall a relationship is not proven.</td>
</tr>
<tr>
<td>c. Tolerability: discontinuations due to adverse effects</td>
<td>Fair-to-poor-quality for celecoxib. No data for valdecoxib, meloxicam and nabumetone.</td>
<td>Rofecoxib: An unplanned endpoint in a good-quality trial found an increased risk of serious thromboembolic events, primarily myocardial infarction, in patients using rofecoxib 50 mg daily. Observational cohort studies of short-term use found no relationship for doses of 25 mg daily or lower.</td>
</tr>
<tr>
<td>c.2. Tolerability: total adverse effects</td>
<td>1. Good. (consistent results from randomized trials)</td>
<td>Celecoxib: A meta-analyses and a re-analysis of a good-quality trial found no risk, but these analyses were flawed.</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td>In most trials total adverse events were not lower with coxibs. Total serious adverse events were as high or higher for coxibs than for NSAIDs.</td>
</tr>
</tbody>
</table>
### c.3. Tolerability: GI adverse effects and GI adverse events leading to discontinuation

<table>
<thead>
<tr>
<th>Tolerability</th>
<th>Celecoxib and rofecoxib, certain measures</th>
<th>Rofecoxib reduced the frequency of withdrawals for GI adverse events in one good-quality trial (for overall GI adverse events, data are insufficient).</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fair-to-good. (celecoxib and rofecoxib, certain measures)</td>
<td>2. Insufficient. (valdecoxib)</td>
<td>3. Fair-to-poor. (meloxicam, nabumetone)</td>
</tr>
</tbody>
</table>

### c.4. Elevated creatinine or blood pressure, edema, CHF, and abnormal LFTs (coxibs vs. nonselective NSAIDs).

| Elevated creatinine or blood pressure, edema, CHF, and abnormal LFTs (coxibs vs. nonselective NSAIDs). | Fair |

### 4. Are there differences in efficacy or safety of COX-2 inhibitors in different demographic groups (age, sex, race)?

<table>
<thead>
<tr>
<th>Demographic Groups (age, sex, race)</th>
<th>Good (age, sex).</th>
<th>Poor (race).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most studies included a majority of women. The data that coxibs are safe and efficacious in different racial groups have been presented to the FDA, but no differences have been described in the peer-reviewed literature.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial data comparing celecoxib to nonselective NSAIDs are inconsistent. Data comparing rofecoxib and valdecoxib to NSAIDs are insufficient. (see Key Question 1 for coxib vs. coxib comparisons.)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


52. Watson DJ, Harper SE, Zhao PL, et al. Gastrointestinal tolerability of the selective cyclooxygenase-2 (COX-2) inhibitor rofecoxib compared with nonselective COX-1 and


Table 4. Trials of celecoxib versus NSAIDs

<table>
<thead>
<tr>
<th>Trial</th>
<th>Subjects</th>
<th>Celecoxib doses (mg)</th>
<th>NSAIDs (mg)</th>
<th>Number of subjects*</th>
<th>Duration (weeks)</th>
<th>Aspirin permitted?</th>
<th>Efficacy measures</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dougados</td>
<td>Ankylosing spondylitis with flare</td>
<td>100 bid</td>
<td>ketoprofen 100 bid</td>
<td>170</td>
<td>6</td>
<td>?</td>
<td>PGA, Sleep, BASFI</td>
<td>Trend favoring celecoxib</td>
</tr>
<tr>
<td>McKenna</td>
<td>OA of the knee with flare</td>
<td>100 bid</td>
<td>diclofenac 50 tid</td>
<td>400</td>
<td>6</td>
<td>Yes</td>
<td>Index joint pain, WOMAC</td>
<td>No difference</td>
</tr>
<tr>
<td>Goldstein</td>
<td>OA and RA with no ulcer on EGD; many had a history of GI disease</td>
<td>200 bid</td>
<td>naproxen 500 bid</td>
<td>537</td>
<td>12</td>
<td>Yes</td>
<td>PGA, WOMAC, withdrawals</td>
<td>No difference</td>
</tr>
<tr>
<td>Bensen/Zhao</td>
<td>OA of the knee with flare</td>
<td>50, 100, or 200 bid</td>
<td>naproxen 500 bid</td>
<td>1004</td>
<td>12</td>
<td>Yes</td>
<td>PGA, WOMAC, withdrawals</td>
<td>No difference</td>
</tr>
<tr>
<td>Simon/Zhao</td>
<td>RA with flare and no ulcer on EGD</td>
<td>100, 200, or 400 bid</td>
<td>naproxen 500 bid</td>
<td>918</td>
<td>12</td>
<td>Yes</td>
<td>PGA, pain, duration of morning stiffness</td>
<td>No difference</td>
</tr>
<tr>
<td>Emery</td>
<td>RA</td>
<td>200 bid</td>
<td>diclofenac 75 bid</td>
<td>655</td>
<td>24</td>
<td>No</td>
<td>PGA, ACR-20</td>
<td>No difference except significantly more improvement in morning stiffness in the NSAID group</td>
</tr>
<tr>
<td>Silverstein</td>
<td>OA and RA</td>
<td>400 bid</td>
<td>ibuprofen 800 tid or diclofenac 75 bid</td>
<td>7968</td>
<td>24</td>
<td>Yes</td>
<td>no efficacy measures reported except withdrawal</td>
<td>Not reported</td>
</tr>
</tbody>
</table>

*Excludes subjects randomized to placebo; PGA - patient global assessment; BASFI - Bath Ankylosing Spondylitis Functional Index; WOMAC - Western Ontario and McMaster Universities Osteoarthritis Index; OA - osteoarthritis; RA - rheumatoid arthritis; EGD - esophagastroduodenoscopy; GI - gastrointestinal; ACR-20 - American College of Rheumatology criteria
Table 5. Trials of rofecoxib versus NSAIDs

<table>
<thead>
<tr>
<th>Study</th>
<th>Subjects</th>
<th>Rofecoxib doses (mg)</th>
<th>NSAIDs (mg)</th>
<th>Number of subjects*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chrubasik</td>
<td>Low back pain</td>
<td>12.5</td>
<td>Assalix 1 qid †</td>
<td>228</td>
</tr>
<tr>
<td>Acevado</td>
<td>OA, negative FOBT</td>
<td>12.5</td>
<td>diclofenac 50/misoprostol 200 mcg bid</td>
<td>483</td>
</tr>
<tr>
<td>Saag</td>
<td>OA of knee or hip and flare (for NSAID users) or acetaminophen user. Excluded aspirin 81mg users.</td>
<td>12.5, 25</td>
<td>ibuprofen 800 tid</td>
<td>667</td>
</tr>
<tr>
<td>Day</td>
<td>OA of knee or hip and flare (for NSAID users) or acetaminophen user.</td>
<td>12.5, 25</td>
<td>ibuprofen 800 tid</td>
<td>735</td>
</tr>
<tr>
<td>Hawkey</td>
<td>OA with no ulcer on EGD</td>
<td>25, 50</td>
<td>ibuprofen 800 tid</td>
<td>581</td>
</tr>
<tr>
<td>Laine (044)</td>
<td>OA with no ulcer or esophagitis on EGD</td>
<td>25, 50</td>
<td>ibuprofen 800 tid</td>
<td>565</td>
</tr>
<tr>
<td>Bombadier (VIGOR)</td>
<td>RA, negative FOBT</td>
<td>50</td>
<td>naproxen 500 bid</td>
<td>8076</td>
</tr>
<tr>
<td>Cannon (035)</td>
<td>OA of knee or hip and flare (for NSAID users) or acetaminophen user.</td>
<td>12.5, 25</td>
<td>diclofenac 50 tid</td>
<td>784</td>
</tr>
<tr>
<td>Saag</td>
<td>OA of knee or hip and flare (for NSAID users) or acetaminophen user. Excluded aspirin 81mg users.</td>
<td>12.5, 25</td>
<td>diclofenac 50 tid</td>
<td>693</td>
</tr>
</tbody>
</table>

* Excludes subjects randomized to placebo; ** If underlined, for lack of efficacy; otherwise, for all reasons; †Willow bark extract containing 15% salicin, total dose 240mg of salicin a day; OA - osteoarthritis; FOBT - fecal occult blood test; PGA - patient global assessment; WOMAC - Western Ontario and McMaster Universities Osteoarthritis Index; EGD - esophagogastroduodenoscopy; RA - rheumatoid arthritis
<table>
<thead>
<tr>
<th>Study</th>
<th>Duration (weeks)</th>
<th>Aspirin permitted?</th>
<th>Efficacy measures</th>
<th>Rofecoxib Withdrawals** (%)</th>
<th>NSAIDs Withdrawals** (%)</th>
<th>Other outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chrubasik</td>
<td>4</td>
<td>Yes</td>
<td>Pain</td>
<td>21%</td>
<td>18.0%</td>
<td>No difference</td>
</tr>
<tr>
<td>Acevado</td>
<td>6</td>
<td>No</td>
<td>PGA</td>
<td>7%</td>
<td>10.80%</td>
<td>No difference</td>
</tr>
<tr>
<td>Saag</td>
<td>6</td>
<td>No</td>
<td>WOMAC, PGA, pain while walking</td>
<td>7.8% (12.5 mg) 4.0% (25 mg)</td>
<td>8.6%</td>
<td>No difference</td>
</tr>
<tr>
<td>Day</td>
<td>6</td>
<td>No</td>
<td>WOMAC, PGA, pain while walking</td>
<td>3.5% (12.5 mg) 2.8% (25 mg)</td>
<td>3%</td>
<td>No difference in 3 primary endpoints, but trend favored rofecoxib 25 mg for 2 of the 3.</td>
</tr>
<tr>
<td>Hawkey</td>
<td>24</td>
<td>No</td>
<td>PGA</td>
<td>3% (12.5 mg) 1.6% (25 mg)</td>
<td>5%</td>
<td>No difference</td>
</tr>
<tr>
<td>Laine (044)</td>
<td>24</td>
<td>No</td>
<td>PGA</td>
<td>3% (12.5 mg) 2.1% (25 mg)</td>
<td>4.9%</td>
<td>No difference</td>
</tr>
<tr>
<td>Bombadier</td>
<td>52</td>
<td>No</td>
<td>PGA</td>
<td>6.30%</td>
<td>6.50%</td>
<td>No difference</td>
</tr>
<tr>
<td>Cannon (035)</td>
<td>52</td>
<td>No</td>
<td>WOMAC, PGA, pain while walking</td>
<td>13.9% (12.5 mg) 21.8% (25 mg)</td>
<td>16%</td>
<td>Trend favoring diclofenac for 2 of 3 primary measures</td>
</tr>
<tr>
<td>Saag</td>
<td>52</td>
<td>No</td>
<td>WOMAC, PGA, pain while walking</td>
<td>12.1% (12.5 mg) 11.2% (25 mg)</td>
<td>7.0%</td>
<td>No difference</td>
</tr>
</tbody>
</table>

* Excludes subjects randomized to placebo; ** If underlined, for lack of efficacy; otherwise, for all reasons; †Willow bark extract containing 15% salicin; total dose 240 mg of salicin a day; OA - osteoarthritis FOBT - fecal occult blood test; PGA - patient global assessment; WOMAC - Western Ontario and McMaster Universities Osteoarthritis Index; EGD - esophagogastroduodenoscopy; RA - rheumatoid arthritis
<table>
<thead>
<tr>
<th>Trial</th>
<th>Subjects</th>
<th>Meloxicam doses (mg)</th>
<th>NSAIDs (mg)</th>
<th>Number of subjects*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valat</td>
<td>OA lumbar spine</td>
<td>7.5</td>
<td>diclofenac 100</td>
<td>229</td>
</tr>
<tr>
<td>Linden§</td>
<td>OA hip</td>
<td>15</td>
<td>piroxicam 20</td>
<td></td>
</tr>
<tr>
<td>Hawkey (Melissa)</td>
<td>OA hip, knee, hand, or spine</td>
<td>7.5</td>
<td>diclofenac 100</td>
<td>9323</td>
</tr>
<tr>
<td>Dequeker (Select) †</td>
<td>OA hip, knee, hand, or spine</td>
<td>7.5</td>
<td>piroxicam 20</td>
<td>8656</td>
</tr>
<tr>
<td>Goei The</td>
<td>OA knee</td>
<td>7.5</td>
<td>diclofenac 100</td>
<td>258</td>
</tr>
<tr>
<td>Hosie 1996</td>
<td>OA hip or knee</td>
<td>7.5</td>
<td>diclofenac 100</td>
<td>336</td>
</tr>
<tr>
<td>Hosie 1997§</td>
<td>OA hip or knee</td>
<td>15</td>
<td>piroxicam 20</td>
<td>455</td>
</tr>
<tr>
<td>Wojtulweski</td>
<td>RA</td>
<td>7.5</td>
<td>naproxen 750</td>
<td>379</td>
</tr>
</tbody>
</table>

* Excludes subjects randomized to placebo; Bold type - statistically significant; §review incomplete at time of draft; †design identical to Hawkey et al; OA - osteoarthritis; PGA - patient global assessment; RA - rheumatoid arthritis
<table>
<thead>
<tr>
<th>Trial</th>
<th>Duration (weeks)</th>
<th>Aspirin permitted?</th>
<th>Efficacy measures</th>
<th>Withdrawals**</th>
<th>Other outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valat</td>
<td>2</td>
<td>unclear</td>
<td>pain on motion</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Linden§</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawkey (Melissa)</td>
<td>4</td>
<td>unclear</td>
<td>pain, PGA, withdrawals</td>
<td>1.7%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Dequeker (Select) †</td>
<td>4</td>
<td>unclear</td>
<td>pain, PGA, withdrawals</td>
<td>1.7%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Goei The, 1997</td>
<td>6</td>
<td>yes</td>
<td>pain during active movement, PGA, acetominophen use.</td>
<td>3.9%</td>
<td>2.3%</td>
</tr>
<tr>
<td>Hosie 1996</td>
<td>24</td>
<td>unclear</td>
<td>pain, quality of life</td>
<td>4%</td>
<td>4%</td>
</tr>
<tr>
<td>Hosie 1997§</td>
<td></td>
<td></td>
<td>overall pain, pain on movement, joint stiffness, global efficacy and quality of life</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wojtulweski</td>
<td>24</td>
<td>no</td>
<td>PGA, several others</td>
<td>23.6%</td>
<td>14.4%</td>
</tr>
</tbody>
</table>

* Excludes subjects randomized to placebo; ** If underlined, for lack of efficacy; otherwise, for all reasons; Bold type - statistically significant; § Review incomplete at time of draft; † Design identical to Hawkey et al; OA - osteoarthritis; PGA - patient global assessment; RA - rheumatoid arthritis
Table 7. Comparison of the VIGOR and CLASS trials

a) Evidence table

<table>
<thead>
<tr>
<th>Trial</th>
<th>Sites</th>
<th>Patients</th>
<th>Aspirin use</th>
<th>Definition of significant GI events</th>
<th>Number screened/enrolled</th>
<th>Number analyzed</th>
<th>Withdraw for lack of efficacy (Coxib group / NSAID groups)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIGOR (rofecoxib 50mg qd)</td>
<td>301 centers, 22 countries</td>
<td>RA, over 50</td>
<td>Not allowed</td>
<td>Perforation, obstruction, upper GI bleeding, or symptomatic ulcer</td>
<td>9539/8076</td>
<td>8076</td>
<td>6.3% / 6.5%</td>
</tr>
<tr>
<td>CLASS (celecoxib 400mg bid)</td>
<td>386 centers, US and Canada</td>
<td>RA or OA, 18 or older</td>
<td>20%</td>
<td>Perforation, obstruction, upper GI bleeding</td>
<td>9764/8059</td>
<td>7968</td>
<td>12.6% / 14.8%</td>
</tr>
</tbody>
</table>

* In VIGOR, there was no difference

b) Comparison of outcomes

<table>
<thead>
<tr>
<th></th>
<th>VIGOR NSAID group**</th>
<th>CLASS NSAID group†</th>
<th>VIGOR NNT</th>
<th>CLASS NNT †</th>
</tr>
</thead>
<tbody>
<tr>
<td>ulcers</td>
<td>0.030</td>
<td>0.011</td>
<td>62</td>
<td>265</td>
</tr>
<tr>
<td>perforation</td>
<td>0.001</td>
<td>0.000</td>
<td>no effect</td>
<td>no effect</td>
</tr>
<tr>
<td>obstruction</td>
<td>0.000</td>
<td>0.000</td>
<td>no effect</td>
<td>no effect</td>
</tr>
<tr>
<td>bleeding from an ulcer</td>
<td>0.008</td>
<td>0.008</td>
<td>268</td>
<td>199</td>
</tr>
<tr>
<td>Complicated confirmed UGI events</td>
<td>0.009</td>
<td>0.008</td>
<td>191</td>
<td>199</td>
</tr>
</tbody>
</table>

**average 9 months of followup; † adjusted to replicate 9 months of followup; RA - rheumatoid arthritis; GI - gastrointestinal; OA - osteoarthritis; NNT - number needed to treat; UGI - upper gastrointestinal
### Table 9. Tolerability in randomized controlled trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Focus</th>
<th>Subjects</th>
<th>Coxib dose</th>
<th>NSAIDs (mg)</th>
<th>Number of subjects*</th>
<th>Duration (weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Celecoxib</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dougados</td>
<td>efficacy</td>
<td>Ankylosing spondylitis with flare</td>
<td>100 bid</td>
<td>ketoprofen 100 bid</td>
<td>170</td>
<td>6</td>
</tr>
<tr>
<td>McKenna</td>
<td>efficacy and tolerability</td>
<td>OA of the knee with flare</td>
<td>100 bid</td>
<td>diclofenac 50 tid</td>
<td>400</td>
<td>6</td>
</tr>
<tr>
<td>Bensen/Zhao</td>
<td>efficacy</td>
<td>OA of the knee with flare</td>
<td>200 bid</td>
<td>naproxen 500 bid</td>
<td>1004</td>
<td>12</td>
</tr>
<tr>
<td>Goldstein</td>
<td>endoscopic ulcers</td>
<td>OA and RA with no ulcer on EGD</td>
<td>200 bid</td>
<td>naproxen 500 bid</td>
<td>537</td>
<td>12</td>
</tr>
<tr>
<td>Simon/Zhao</td>
<td>efficacy and endoscopic ulcers</td>
<td>RA with flare and no ulcer on EGD</td>
<td>100, 200,</td>
<td>naproxen 500 bid</td>
<td>918</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>or 400 bid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emery</td>
<td>endoscopic ulcers</td>
<td>RA</td>
<td>200 bid</td>
<td>diclofenac 75 bid</td>
<td>655</td>
<td>24</td>
</tr>
<tr>
<td>Silverstein (CLASS)</td>
<td>serious GI events</td>
<td>OA and RA</td>
<td>400 bid</td>
<td>ibuprofen 800 tid or diclofenac 75 bid</td>
<td>7968</td>
<td>24</td>
</tr>
<tr>
<td><strong>Rofecoxib</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acevado</td>
<td>adverse events</td>
<td>OA, negative FOBT</td>
<td>12.5 mg</td>
<td>diclofenac 50 mg/misoprostol 200 mcg bid</td>
<td>483</td>
<td>6</td>
</tr>
<tr>
<td>Saag</td>
<td>efficacy and tolerability</td>
<td>OA of knee or hip with flare (for NSAID users) or acetaminophen user. Excluded aspirin 81mg users.</td>
<td>25 mg</td>
<td>ibuprofen 800 tid</td>
<td>667</td>
<td>6</td>
</tr>
</tbody>
</table>

*Excludes subjects randomized to placebo; **inadequately reported; § Reported GI adverse events leading to discontinuation, but did not report total GI adverse events; †statistically significant; GI - gastrointestinal; HTN - hypertension; CHF - congestive heart failure; NR - not reported; OA - osteoarthritis; EGD - esophagastroduodenoscopy; RA - rheumatoid arthritis; FOBT-fecal occult blood test; LFT - liver function test
Table 9. Tolerability in randomized controlled trials (continued)

<table>
<thead>
<tr>
<th>Trial</th>
<th>Withdrawals due to adverse events</th>
<th>Total adverse events</th>
<th>GI adverse events</th>
<th>Elevated creatinine, HTN, CHF, or edema</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coxib</td>
<td>NSAID</td>
<td>coxib</td>
<td>NSAID</td>
<td>coxib</td>
</tr>
<tr>
<td>Celecoxib</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dougados</td>
<td>6.3%</td>
<td>1.1%</td>
<td>68.0%</td>
<td>60.0%</td>
<td>32.2%</td>
</tr>
<tr>
<td>McKenna</td>
<td>7.0%</td>
<td>11.0%</td>
<td>50.0%</td>
<td>54.0%</td>
<td>18.0%</td>
</tr>
<tr>
<td>Bensen/Zhao</td>
<td>10.0%</td>
<td>8.0%</td>
<td>65.0%</td>
<td>63.0%</td>
<td>24.0%</td>
</tr>
<tr>
<td>Goldstein</td>
<td>7.0%</td>
<td>9.0%</td>
<td>70.0%</td>
<td>70.0%</td>
<td>34.0%</td>
</tr>
<tr>
<td>Simon/Zhao</td>
<td>5.5%</td>
<td>5.3%</td>
<td>62%-68%</td>
<td>65.0%</td>
<td>26.0%</td>
</tr>
<tr>
<td>Emery</td>
<td>nr</td>
<td>nr</td>
<td>68.0%</td>
<td>73.0%</td>
<td>36.0%</td>
</tr>
<tr>
<td>Silverstein (CLASS)</td>
<td>18.4%</td>
<td>20.6%</td>
<td>48.5%</td>
<td>56.8%</td>
<td>31.4%†</td>
</tr>
<tr>
<td>Rofecoxib</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acevado</td>
<td>4.1%</td>
<td>9.1%</td>
<td>52.9%†</td>
<td>73.0%</td>
<td>28.9%</td>
</tr>
<tr>
<td>Saag</td>
<td>no difference</td>
<td></td>
<td>nr</td>
<td>nr</td>
<td>3.5%</td>
</tr>
</tbody>
</table>

*Excludes subjects randomized to placebo; **inadequately reported; § Reported GI adverse events leading to discontinuation, but did not report total GI adverse events; †statistically significant; GI - gastrointestinal; HTN - hypertension; CHF - congestive heart failure; NR - not reported; OA - osteoarthritis; EGD - esophagogastroduodenoscopy; RA - rheumatoid arthritis; FOBT fecal occult blood test; LFT - liver function test
### Table 9. Tolerability in randomized controlled trials (continued)

<table>
<thead>
<tr>
<th>Trial</th>
<th>Focus</th>
<th>Subjects</th>
<th>Coxib Dose (mg)</th>
<th>NSAIDs (mg)</th>
<th>Number of subjects*</th>
<th>Duration (weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>efficacy and tolerability</td>
<td>OA of knee or hip with flare (for NSAID users) or acetaminophen user</td>
<td>25</td>
<td>ibuprofen 800 tid</td>
<td>735</td>
<td>6</td>
</tr>
<tr>
<td>Truitt</td>
<td>efficacy and tolerability</td>
<td>OA knee or hip with flare, &gt;80 years old</td>
<td>25</td>
<td>nabumetone 1500</td>
<td>250</td>
<td>6</td>
</tr>
<tr>
<td>Hawkey</td>
<td>endoscopic ulcers</td>
<td>OA with no ulcer on EGD</td>
<td>25</td>
<td>ibuprofen 800 tid</td>
<td>581</td>
<td>18</td>
</tr>
<tr>
<td>Laine (044)</td>
<td>endoscopic ulcers</td>
<td>OA with no ulcer or esophagitis on EGD</td>
<td>25, 50</td>
<td>ibuprofen 800 tid</td>
<td>565</td>
<td>24</td>
</tr>
<tr>
<td>Bombadier</td>
<td>serious GI events</td>
<td>RA, negative FOBT</td>
<td>50</td>
<td>naprofen 500 bid</td>
<td>8076</td>
<td>52</td>
</tr>
<tr>
<td>Cannon (035)</td>
<td>efficacy</td>
<td>OA of knee or hip with flare (for NSAID users) or acetaminophen user</td>
<td>25</td>
<td>diclofenac 50 tid</td>
<td>784</td>
<td>52</td>
</tr>
<tr>
<td>Saag</td>
<td>efficacy and tolerability</td>
<td>OA of knee or hip with flare (for NSAID users) or acetaminophen user. Excluded aspirin 81 mg users.</td>
<td>25</td>
<td>diclofenac 50 tid</td>
<td>693</td>
<td>52</td>
</tr>
<tr>
<td>Makarowski</td>
<td>efficacy and tolerability</td>
<td>OA of the hip</td>
<td>5, 10</td>
<td>naprofen 500 bid</td>
<td>349</td>
<td>12</td>
</tr>
</tbody>
</table>

*Excludes subjects randomized to placebo; **inadequately reported; § Reported GI adverse events leading to discontinuation, but did not report total GI adverse events; †statistically significant; GI - gastrointestinal; HTN - hypertension; CHF - congestive heart failure; NR - not reported; OA - osteoarthritis; EGD - esophagogastroduodenoscopy; RA - rheumatoid arthritis; FOBT fecal occult blood test; LFT - liver function test
Table 9. Tolerability in randomized controlled trials (continued)

<table>
<thead>
<tr>
<th>Trial</th>
<th>Withdrawals due to adverse events</th>
<th>Total adverse events</th>
<th>GI adverse events</th>
<th>Elevated creatinine, HTN, CHF, or edema</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coxib</td>
<td>NSAID</td>
<td>coxib</td>
<td>NSAID</td>
<td>coxib</td>
</tr>
<tr>
<td>Day</td>
<td>3.7%</td>
<td>8.4%</td>
<td>53.3%</td>
<td>51.8%</td>
<td>higher for NSAID (numbers not given)</td>
</tr>
<tr>
<td>Truitt</td>
<td>8.9%</td>
<td>7.0%</td>
<td>nr</td>
<td>nr</td>
<td>nr</td>
</tr>
<tr>
<td>Hawkey</td>
<td>5.6%</td>
<td>9.8%</td>
<td>80.1%</td>
<td>80.0%</td>
<td>no difference</td>
</tr>
<tr>
<td>Laine (044)</td>
<td>10.3%</td>
<td>14.0%</td>
<td>78.3%</td>
<td>74.7%</td>
<td>nr</td>
</tr>
<tr>
<td>Bombadier (VIGOR)</td>
<td>16.4%</td>
<td>16.1%</td>
<td>3.5%§†</td>
<td>4.9%</td>
<td>1.2%</td>
</tr>
<tr>
<td>Cannon (035)</td>
<td>12.5%</td>
<td>15.3%</td>
<td>84.0%</td>
<td>86.2%</td>
<td>no difference (numbers not given)</td>
</tr>
<tr>
<td>Saag</td>
<td>significantly higher for NSAID (numbers not given)</td>
<td>nr</td>
<td>nr</td>
<td>5.2%</td>
<td>8.3%</td>
</tr>
<tr>
<td>Valdecoxib</td>
<td>Makarowski</td>
<td>9%</td>
<td>12.70%</td>
<td>53%</td>
<td>60.20%</td>
</tr>
</tbody>
</table>

*Excludes subjects randomized to placebo; **inadequately reported; § Reported GI adverse events leading to discontinuation, but did not report total GI adverse events; †statistically significant; GI - gastrointestinal; HTN - hypertension; CHF - congestive heart failure; NR - not reported; OA - osteoarthritis; EGD - esophagogastroduodenoscopy; RA - rheumatoid arthritis; FOBT fecal occult blood test; LFT - liver function test.
Appendix A.

Table 1. Cyclooxygenase Selectivity of 25 Nonsteroidal Anti-Inflammatory Drugs*

<table>
<thead>
<tr>
<th>Nonsteroidal Anti-Inflammatory Drug</th>
<th>Ratio†</th>
<th>Reference‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flurbiprofen</td>
<td>10.27</td>
<td>4, 20, 21</td>
</tr>
<tr>
<td>Ketoprofen</td>
<td>8.16</td>
<td>4</td>
</tr>
<tr>
<td>Fenoprofen</td>
<td>5.14</td>
<td>4</td>
</tr>
<tr>
<td>Tolmetin</td>
<td>3.93</td>
<td>4, 20</td>
</tr>
<tr>
<td>Aspirin</td>
<td>3.12</td>
<td>4</td>
</tr>
<tr>
<td>Oxaprocin</td>
<td>2.52</td>
<td>4</td>
</tr>
<tr>
<td>Naproxen</td>
<td>1.79</td>
<td>4, 20, 21</td>
</tr>
<tr>
<td>Indomethacin</td>
<td>1.78</td>
<td>4, 20, 21</td>
</tr>
<tr>
<td>Ibuprofen</td>
<td>1.69</td>
<td>4</td>
</tr>
<tr>
<td>Kеторолак</td>
<td>1.64</td>
<td>4, 20, 21</td>
</tr>
<tr>
<td>Proxicam</td>
<td>0.79</td>
<td>4, 20, 21</td>
</tr>
<tr>
<td>Nabumetone, 6-MNA</td>
<td>0.64</td>
<td>4, 21</td>
</tr>
<tr>
<td>Etodolac</td>
<td>0.11</td>
<td>4, 25</td>
</tr>
<tr>
<td>Celecoxib</td>
<td>0.11</td>
<td>Cryer B, Feldman M. Unpublished data</td>
</tr>
<tr>
<td>Meloxicam</td>
<td>0.09</td>
<td>21</td>
</tr>
<tr>
<td>Mefenamic acid</td>
<td>0.08</td>
<td>4</td>
</tr>
<tr>
<td>Flurbinitroxybutylester</td>
<td>0.08</td>
<td>22</td>
</tr>
<tr>
<td>NS-398</td>
<td>0.07</td>
<td>4, 20</td>
</tr>
<tr>
<td>Diclofenac</td>
<td>0.05</td>
<td>4, 20, 21</td>
</tr>
<tr>
<td>DuP-697</td>
<td>0.05</td>
<td>20</td>
</tr>
<tr>
<td>Rofecoxib</td>
<td>0.05</td>
<td>19</td>
</tr>
<tr>
<td>Nimesulide</td>
<td>0.04</td>
<td>4</td>
</tr>
<tr>
<td>Flusulide (CGP 28238)</td>
<td>0.02</td>
<td>20</td>
</tr>
<tr>
<td>SC-58125</td>
<td>0.007</td>
<td>20, 21</td>
</tr>
<tr>
<td>L-745,337</td>
<td>0.004</td>
<td>23</td>
</tr>
</tbody>
</table>

* Modified from reference 4.
† Expressed as the ratio of the 50% inhibitory concentration of cyclooxygenase-2 to the 50% inhibitory concentration of cyclooxygenase-1 in whole blood. Nonsteroidal anti-inflammatory drugs with a ratio <1 indicate selectivity for cyclooxygenase-2.
‡ If more than one study determined a ratio, the median ratio was used.

From:

Appendix B: Methods for Drug Class Reviews for Oregon Health Plan Practitioner-Managed Prescription Drug Plan

Oregon Evidence-based Practice Center

December 14, 2001
Updated February 4, 2003

Overview

The purpose of this document is to outline the methods used by the Oregon Evidence-based Practice Center (EPC), based at Oregon Health & Science University, in developing drug class reviews for the Oregon Health Plan Practitioner-Managed Prescription Drug Plan.

The methods outlined in this document ensure that the products created in this process are methodologically sound, scientifically defensible, reproducible, and well-documented. This document has been adapted from the Procedure Manual developed by the Methods Work Group of the United States Preventive Services Task Force (version 1.9, September 2001), with additional material from the NHS Centre for Reviews and Dissemination (CRD) report on Undertaking Systematic Reviews of Research on Effectiveness: CRD’s Guidance for Carrying Out or Commissioning Reviews (2nd edition, 2001) and “The Database of Abstracts of Reviews of Effects (DARE)” in Effectiveness Matters, vol. 6, issue 2, December 2002, published by the CRD. To ensure scientific rigor and relevance of the work, the Oregon EPC develops key questions and criteria for admissible evidence, and uses these to create a literature search strategy that best captures the appropriate evidence. To consider papers identified by the searches, the teams use the criteria for admissible evidence (explicit inclusion and exclusion criteria) to select papers that provide information to help answer the key questions. They abstract key data from these selected papers. The teams use established criteria to assess the internal validity of the evidence in each paper, as well as the total internal validity, external validity, and coherence of the evidence for each key question.

Key Questions and Inclusion/Exclusion Criteria

Key questions are essential in focusing the literature review on a manageable and clinically relevant topic. All key questions are reviewed and approved by the topic team in the process of assessing and refining the topic before the detailed literature review. The EPC teams work with the subcommittee members of the Oregon Health Resources Commission assigned to a particular drug class to finalize the key questions for that drug class.

We clearly document the criteria by which the team chooses to admit evidence on a given key question. Such criteria might include, for example, study design (e.g., randomized...
controlled trials, cohort studies), setting, sample size, population studied, language(s) of publication, and year(s) of publication.

No generic criteria for admissible evidence have been established. Rather, the criteria are determined on a topic-by-topic and key question-by-key question basis, depending on the questions involved and the amount and quality of evidence available. All inclusion/exclusion criteria are reviewed and approved by the entire topic team.

Databases to Be Searched and Documenting Search Terms

At a minimum, all topics include a review of the English-language literature in MEDLINE and EMBASE bibliographic databases and the Cochrane Controlled Trials Register. Other databases (e.g., nursing or psychology databases) are searched as deemed necessary by the topic team. Evidence reviews document the databases used.

Search terms used for each key question, along with the yield associated with each term, are documented in a table or set of tables; these appear in the final evidence review.

Database of Abstracts

The EPC, for each review, establishes a database of all abstracts (i.e., both those included and those eventually excluded from the final set of full-text articles reviewed). Information captured in the database includes the key question(s) associated with each included abstract and reason for exclusion if the abstract does not meet inclusion criteria.

Abstraction Forms

Although the EPC has no standard or generic abstraction form, the following broad categories are always abstracted from included articles: study design, study participant description, quality information, and outcomes. Each team uses these (and, if indicated, other) general categories to develop an abstraction form specific to the topic at hand.

Double Abstraction of Included Articles

The EPC teams abstract only those articles that, after review of the entire article, meet criteria for both quality and focus on the key question at hand. Key articles are always read and checked by more than one team member. All reviewers are trained in the topic, the analytic framework and key questions, and the use of the abstraction instrument. Initial reliability checks are done for quality control.
Quality Criteria

Assessment of Internal Validity
To assess the internal validity of individual studies, the EPC adopted criteria for assessing the internal validity of individual studies from the US Preventive Services Task Force and the NHS Centre for Reviews and Dissemination.

For Controlled Trials:

Assessment of Internal Validity

1. Was the assignment to the treatment groups really random?
   Adequate approaches to sequence generation:
   - Computer-generated random numbers
   - Random numbers tables
   Inferior approaches to sequence generation:
   - Use of alternation, case record numbers, birth dates or week days
   - Not reported

2. Was the treatment allocation concealed?
   Adequate approaches to concealment of randomization:
   - Centralized or pharmacy-controlled randomization
   - Serially-numbered identical containers
   - On-site computer based system with a randomization sequence that is not readable until allocation
   - Other approaches sequence to clinicians and patients
   Inferior approaches to concealment of randomization:
   - Use of alternation, case record numbers, birth dates or week days
   - Open random numbers lists
   - Serially numbered envelopes (even sealed opaque envelopes can be subject to manipulation)
   - Not reported

3. Were the groups similar at baseline in terms of prognostic factors?

4. Were the eligibility criteria specified?

5. Were outcome assessors blinded to the treatment allocation?

6. Was the care provider blinded?

7. Was the patient kept unaware of the treatment received?
8. Did the article include an intention-to-treat analysis, or provide the data needed to calculate it (i.e., number assigned to each group, number of subjects who finished in each group, and their results)?

9. Did the study maintain comparable groups?

10. Did the article report attrition, crossovers, adherence, and contamination?

11. Is there important differential loss to followup or overall high loss to followup? (give numbers in each group)

Assessment of External Validity (Generalizability)

1. How similar is the population to the population to whom the intervention would be applied?

2. How many patients were recruited?

3. What were the exclusion criteria for recruitment? (Give numbers excluded at each step)

4. What was the funding source and role of funder in the study?

5. Did the control group receive the standard of care?

6. What was the length of followup? (Give numbers at each stage of attrition.)

For Reports of Complications/Adverse Effects

Assessment of Internal Validity

1. Was the selection of patients for inclusion non-biased (Was any group of patients systematically excluded)?

2. Is there important differential loss to followup or overall high loss to followup? (Give numbers in each group.)

3. Were the events investigated specified and defined?

4. Was there a clear description of the techniques used to identify the events?

5. Was there non-biased and accurate ascertainment of events (independent ascertainer; validation of ascertainment technique)?
6. Were potential confounding variables and risk factors identified and examined using acceptable statistical techniques?

7. Did the duration of follow-up correlate to reasonable timing for investigated events? (Does it meet the stated threshold?)

**Assessment of External Validity**

1. Was the description of the population adequate?

2. How similar is the population to the population to whom the intervention would be applied?

3. How many patients were recruited?

4. What were the exclusion criteria for recruitment? (Give numbers excluded at each step)

5. What was the funding source and role of funder in the study?

**Economic Studies**

**Assessment of Internal Validity**

**Framing**

1. Was a well-defined question posed in answerable form?

2. Was a comprehensive description of the competing alternatives given?

3. Are the interventions and populations compared appropriate?

4. Is the study conducted from the societal perspective?

5. Is the time horizon clinically appropriate and relevant to the study question?

**Effects**

1. Are all important drivers of effectiveness included?

2. Are key harms included?

3. Is the best available evidence used to estimate effectiveness?

4. Are long-term outcomes used?

5. Do effect measures capture preferences or utilities?
Costs
1. Are costs and outcomes measured accurately?
2. Are costs and outcomes valued credibly?
3. Are costs and outcomes adjusted for differential timing?
4. Are all appropriate downstream medical costs included?
5. Are charges converted to costs appropriately?
6. Are the best available data used to estimate costs? (like first question)
7. Are all important and relevant costs and outcomes for each alternative identified?

Results
1. Are incremental cost-effectiveness ratios presented?
2. Are appropriate sensitivity analyses performed?
3. How far do study results include all issues of concern to users?

Assessment of External Validity
1. Are the results generalizable to the setting of interest in the review?

Systematic Reviews:
1. Is the systematic review recent and relevant?
2. Is the review comprehensive in considering sources and in searching databases to find all relevant research?
3. Are inclusion/exclusion criteria reported relating to the primary studies that address the review question? If so, are they explicit and relevant?
4. Are the primary studies summarized appropriately?
5. Is sufficient detail of the primary studies presented?
6. Is there standard appraisal of the primary studies?
7. Is the validity of primary studies adequately assessed?
8. Are there valid conclusions in the systematic review?