Author's response to reviews

Title: An economic model of long-term use of celecoxib in patients with osteoarthritis

Authors:

Michael Loyd (mloydassoc@shaw.ca)
Dale A Rublee (dale.rublee@pfizer.com)
Philip Jacobs (pjacobs@ihe.ca)

Version: 2 Date: 11 May 2007

Author's response to reviews: see over
Reviewer report and authors’ response:
An economic model of long-term use of celecoxib in patients with OA
Loyd, Rublee, Jacobs

Comment 1. “This paper presents an economic model of long-term celecoxib versus non-selective NSAID use. The main issue here is that the findings of this analysis are dramatically different from another well-known cost-effectiveness study of coxibs by Speigel, et al. In that study, an ICER of over $240,000 per QALY was reported versus $31,000 reported in this paper. The reasons for these large differences need to be very transparent to the reader.”

Reply. Professor Brookhart is correct that results in our paper differ sharply from Spiegel et al, and we agree completely that it is essential for readers to be clear about model differences. We would draw attention to the text, beginning on page 18, referring to the “celecoxib-rofecoxib hybrid”.

A sentence has been added on that page about what factors lay behind the differences; otherwise the text seems to be straightforward. Another new sentence emphasizes that our key ulcer probabilities and relative risks (RRs) are based on the SUCCESS trial of celecoxib rather than Spiegel’s use of those for a coxib hybrid with predominant weightings from rofecoxib trials and the CLASS trial, which used supratherapeutic doses of celecoxib. The text now also points out that the main difference is that the Spiegel model’s base case ICER results from his theory that annual new patient incidence rates of PUBs would decrease 35% per year over the entire 21 years of the model, an inference which the authors based on a comparison between the incidence rates of nsNSAID POBs in the first 6 months and the remainder of the CLASS trial. With this assumed “decay function,” the cumulative incidence of PUBs in the Spiegel model was 2.6% for nsNSAID patients after the first year and totaled only 7.2% over a 21-year time span; whereas our base model’s corresponding 21-year rate was 34.6%, a difference of almost fivefold. Sentences on the differences in the 2 studies have been added on page 19 in yellow.

We note for this reply that when Spiegel employs a sensitivity test for “high-risk” patients which incorporates 21-year PUB cumulative incidence rates of 19% for nsNSAIDs and 4.9% for his coxib hybrid (a difference of 14.1% in cumulative incidence in the 2 arms), the resulting ICER is $55,803. This demonstrates that the main differences between Spiegel’s model and our base case model (15.6% difference in cumulative incidence in the 2 arms) stem from differences in the PUB probabilities.

Spiegel’s results are dominated by the “theory,” as the authors termed it in their Appendix, that the annual new patient incidence rates of PUBs would decrease 35% per year each year over the entire 21 years of the model. The theory is based on an extrapolation of controversial, special case, and questionable data from CLASS. We do not find any basis in the literature for assuming such a progressive long-term decay function, even if one subscribes to the principle of depletion of the susceptibles, which is controversial in itself.

May 11, 2007
Other differences are that Spiegel et al. did not take into account increased UGI risks after PUB events or the price reductions from the patent expirations on celecoxib and omeprazole, while ours does. Supporting references are included in the paper.

Comment 2. “The model is not presented in such a way so that it would be reproducible; and it is therefore hard to evaluate its plausibility. For example, it is unclear how the timing of events is tracked. A Markov model seems like it would have been appropriate choice for this analysis, but it is not clear if that was used.”

Reply. We have expanded the wording on page 6 of the revised manuscript to provide more information about the methodology of the decision model. The approach is suitable for the available clinical probabilities and their limitations, and the fact that base-model results are most sensitive to differences in cumulative incidence rates of UGI adverse events.

Comment 3. “Projecting results from a 12-week trial to 21 years of follow-up is speculative. It is reasonable to expect that there is "depletion of susceptibles" phenomenon with NSAIDs, i.e., there are some patients who are vulnerable to developing GI complications on NSAIDs and that may happen quickly. So, that the risk difference between nsNSAIDs and coxib might be large initially but then converge over time. This should be explored in a sensitivity analysis.”

Reply. We realize that short-term clinical trials do not provide adequate information on long-term outcomes in chronic diseases, so we made use of longer-term studies to assess the behavior of NSAID risks over time. First we reviewed results of all longer NSAID clinical trials ranging from 1 to 3 years and found that the overall evidence strongly supports a constancy of PUB risks over these periods (page 19 of manuscript). Laine drew the same conclusion based on earlier information. The impact of the aging of the study population on UGI risk is inconsequential over such 1-3 year observational periods.

We also reviewed the evidence from the few observational studies—mostly published in the late 1980s and early 1990s—that found higher POB risks in the initial 1 to 3 months of treatment and thereby occasioned the theory of depletion of susceptibles. The evidence is relatively weak, there were other explanations for the observed behavior of risks, and none of the studies supported Spiegel’s theory of a long-term progressive decay function. Additionally, the results of other observational studies including newer studies with observation periods ranging up to 15 year are inconsistent with depletion of susceptibles.

---

1 Some of the other explanations offered for the higher POB event rates in a few observational studies include design issues, differences in compliance rates over time, and recall bias.
We adopted a conservative assumption in the base model of constant peptic ulcer and POB risks over the analytic horizon. In sensitivity analysis, we adopted an alternate assumption that incorporates evidence that PUB risks increase with age.

The notion of depletion of the susceptibles is usually based on the premise that patients first exposed to NSAIDs have higher POB rates than previous users. Hence, if all patients in a clinical trial were lifetime first users, the implication is that POB event rates per period of time would fall after the most susceptible early first users are weeded out. Hence, if one extrapolated event rates derived from the entire follow-up period of such a clinical trial, the resulting cumulative incidence rates would be overestimated.

However, most subjects in coxib clinical trials have had prior exposure to NSAIDs, since their average ages are about 60 years with approximate 8-year average durations of arthritis. Indeed, in the CLASS trial (for which this information is available) over 80% of patients were currently on NSAID therapy at the baseline. Others would have had a history of NSAID use even if they were not on NSAIDs at baseline. Hence, even if depletion of the susceptibles is a valid concept, it is unlikely to bias probabilities derived from the coxib clinical trials. As stated above, regardless of underlying reasons, the data from longer trials are most consistent with constant risks.

The data supporting the theory of depletion of the susceptibles applies to POBs. If we reduce the celecoxib and nsNSAID probabilities from SUCCESS by 30%, the ICER would be $40,841/QALY. If we extend the reduction to peptic ulcers as well, the ICER would decline to $45,955/QALY. This sensitivity analysis was added to the revised manuscript on page 15.

Comment 4. “Are QALYs discounted? This is not mentioned in methods anywhere. Given that costs were discounted at 3% per year, if QALYs aren’t discounted, results are going to be biased in favor of coxibs.”

Reply. Yes, QALYs were discounted. Changes in Tables 1 & 2 have been made to underscore this fact.

Comment 5. The sensitivity analysis results need to be presented rather than just mentioned.

Reply. We are unclear on the purport of this comment. The main sensitivity test results were presented either in the text or in Table 5 or both. We have changed some of the headings in this table to be clearer. The comment may also relate to our decision not to present individual results for our alternate post event treatment strategies (page 15). We provided the range of results (Table 5 and page 16) instead of providing individual results for each alternative because some of the underlying clinical probabilities (especially the PPI RR) were not robust. The main conclusion of these sensitivity analyses was that the ICERs of all alternates were well below $60,000/QALY. An extra sentence has now been added on page 15.

May 11, 2007
Comment 6. “0.13 is a big utility decrement for ulcer/dyspepsia relative to the other numbers in the literature. This is not varied and the results may be quite sensitive to this number. It is labeled as the disutility for moderate to severe dyspepsia, but in the paper it came from, it is actually the value for severe dyspepsia. Mild to moderate are in the 0.07 to 0.09 range. Other numbers in the literature are generally smaller. For example, 0.02 in Sullivan Med Care paper, 0.07 in Beaver Dam study, 0.06 in Gerson Am J Gastroenterol”

Reply. In the original manuscript, we used a utility decrement of 0.13 for dyspepsia and peptic ulcers, and varied it (±20%) from 0.104 to 0.156 as mentioned at the bottom of Table 2. (We have now made our sensitivity testing of this parameter clearer in the revised Table 2 by showing the ranges). Our cohort of OA patients are older patients aged 60 to 81 years. Groeneveld did estimate the median QALY decrement to be 0.13 for severe dyspepsia in patients of all ages (his Table 3). However, we inferred that 0.13 would be an appropriate estimate for moderate to severe dyspepsia in our older patient population based on Groeneveld’s estimates for elderly patients in his Table 4.

Modeling for dyspepsia is complex because of various definitional issues (eg, includes or excludes predominant UGI symptoms of heartburn/gastroesophageal reflux disease or nausea or discomfort or bloating) and criteria issues relating to frequency, duration, and severity.(13),(14) Hence, the incidence rates of “dyspepsia” in journal articles range from 10% to 40% among all patients, not just NSAID users, whose rates are higher.(15),(13),(14),(16) One clinical trial found dyspepsia-like adverse event cumulative incidence rates of 35.5% to 43.3% in NSAID users.(17) And the utility decrement for someone with a mild 1-hour episode of heartburn over a health state measurement period for utility determination would be much different from that for someone with a continuous severe abdominal pain over this health state measurement period for utility determination.

Our analysis is conservative in its use of a restrictive definition of dyspepsia and relatively low cumulative incidences rates of 7.8% and 12.0% over the lifetime horizon. Our use of an “annual” utility decrement of –0.13 is consistent with this approach of excluding the mild and short duration episodes. Further, our use of this disutility for dyspepsia and peptic ulcers is generally conservative relative to those employed in the coxib cost-utility evaluation literature.(1),(18),(19),(20),(21)

We corroborated this peptic ulcer disutility by comparing differences in average Health Utility Index values for older OA patients with and without ulcers in a large Canadian database.

---

2 Similar arguments apply to peptic ulcers because most such ulcers are asymptomatic or insufficiently symptomatic for the patient to seek medical advice. Clinical studies of peptic ulcers may be based on endoscopic screening, which would identify both symptomatic and asymptomatic peptic ulcers, or on “symptomatic” ulcers identified by endoscopy after meeting clinical criteria for a medically-necessary endoscopy. Utility decrements from asymptomatic or largely asymptomatic peptic ulcers would be minimal.

May 11, 2007
In response to Professor Brookhart’s concern, we have reduced the lower end of the range of the annual utility decrements for peptic ulcers and dyspepsia to –0.06 in Table 2 of the manuscript to deal with any remaining uncertainties. Additionally, for this reply, we tested the base model with the assumption that both values were at –0.06 and generated an ICER of $44,685/QALY.

Comment 7. “Finally, all of the relevant treatment options are not represented here. The ICER of celecoxib benefits from patent expiration in 2013. ….It would also be of interest to know the ICER of continual treatment with a PPI+naproxen/H2-blocker.”

Reply. We believe that the 2 initial treatment strategies in our model are appropriate and frequently-used strategies in populations of average-risk patients aged 60 years.

We agree with Professor Brookhart that evaluations of other initial treatment options involving PPIs would be a worthwhile study, too, especially in higher-risk patients. However, the data on the relative risk reductions in UGI adverse events with PPIs are very weak. We believe that such an evaluation warrants a separate study given the resultant complexity of workarounds and extensive sensitivity testing required to deal with the inadequacies in the data on other options, and the associated uncertainty.

We believe that the more general issue of whether celecoxib is cost effective relative to nsNSAIDs in populations of OA patients with average UGI risks needed to be the focus of our manuscript. Many evaluations of celecoxib versus nsNSAIDs in high-risk patients show that celecoxib is more cost effective, whereas its cost-effectiveness versus nsNSAIDs in average-risk patients is more controversial.

Comment 8. That begs the question of whether treating with nsNSAIDs until 2013 and then switching to celecoxib might be cost-effective.

First, we note that the ICER in the base model would have been $57,363 per QALY in the absence of patent reduction (See page 14) and $38,594 per QALY in the model with age-related UGI event risks (not reported in manuscript).

Additionally, his comment on patent expiration reflects on issues of lifetime treatment horizons, like that used in our model. We chose the lifetime horizon because our evaluation applies to the use of NSAIDs for OA, which is a progressive and degenerative disease whose symptoms often worsen through time. Our approach is consistent with cost-effectiveness analysis guidelines suggesting that the analytic horizon should reflect the duration of the condition being treated and should capture all relevant information (eg, patent expiration, cumulative effects, and differences in mortality rates). Our

---

3 We use PPI strategies as subsequent treatment options after UGI adverse events, which increase patient risks of recurrent events. The potential impacts of weakness in PPI RR data are minimal in our use of PPI strategies only in patients who have experienced a UGI adverse event.
sensitivity tests of analytic horizon duration allow readers to make inferences about the cost-effectiveness of shorter-term use.

We believe that our approach is consistent with the principles of cost-effectiveness from a societal perspective and its implied implications for resource allocation. Further, the base-case ICER satisfies our criteria for cost-effectiveness, even if we assume no patent-related price reduction in celecoxib.

Reference List


(19) Yun HR, Bae SC. Cost-effectiveness analysis of NSAIDs, NSAIDs with concomitant therapy to prevent gastrointestinal toxicity, and COX-2 specific inhibitors in the treatment of rheumatoid arthritis. Rheumatol Int 2003 Sep 12.


(22) Moore A, Derry S. Gastrointestinal harm from nsaid: an operations research approach [abstract]. Digestive Disease Week 2006 May.