Early response to COX-2 inhibitors as a predictor of overall response in osteoarthritis: pooled results from two identical trials comparing etoricoxib, celecoxib and placebo

Clifton O. Bingham III1, Steven S. Smugar2, Hongwei Wang2 and Andrew M. Tershakovec2

Objective. We evaluated whether early response to NSAIDs predicted later response, and when this was established.

Methods. We evaluated pooled data from two identical 26-week, double-blind, randomized trials comparing once-daily etoricoxib 30 mg (n = 475), celecoxib 200 mg (n = 488) and placebo (n = 244) in patients with knee or hip OA. The present analysis was limited to the 12-week placebo-controlled period. Patient-level OMERACT-OARSI response was determined at 2, 4, 8 and 12 weeks. The proportion of patients who maintained response status between these times was determined from binomial distribution using the exact method.

Results. After 12 weeks of treatment, there were significantly more responders in the etoricoxib (59.8%) and celecoxib (57%) groups compared with placebo (34%; P < 0.001 for etoricoxib or celecoxib vs placebo). About 77.2% of the patients receiving etoricoxib, 75.4% celecoxib and 58% placebo (P = 0.001 vs etoricoxib; P = 0.003 vs celecoxib) who were responders at 2 weeks were also responders at 12 weeks. When comparing response agreement (responder or non-responder) at 2 weeks and 12 weeks, 74.3% of the patients receiving etoricoxib, 73.2% celecoxib and 71.3% placebo had the same response status (κ-coefficient 0.459, 0.449 and 0.357, respectively). There were small incremental increases in agreement between Weeks 4 and 8 and 12 weeks. Logistic regression showed that agreement was not affected by index joint (P = 0.965).

Conclusions. The overwhelming majority of the patients who responded to treatment by 2 weeks remained responders at 12 weeks, with response status largely established within 2 weeks of treatment initiation. Early identification of NSAID response or non-response may allow clinicians to better and more rapidly adjust symptomatic OA management.

Key words: Celecoxib, Etoricoxib, NSAID, OMERACT-OARSI, Osteoarthritis, Responder criteria, Response.

Introduction

NSAIDs are the mainstays of symptomatic treatment of OA [1], as well as other diseases requiring chronic pain treatment [2, 3]. Although it is reported that some rheumatic diseases respond better to particular NSAIDs (e.g. indomethacin in gout) [4], multiple studies have revealed few clinically important differences in efficacy [5–8], despite great variety in structure and pharmacokinetics among different NSAIDs [9]. However, while overall efficacy appears to be similar when evaluated at the population or group level, individual patient responses are highly variable and may not represent group mean effects [10–12], making individual responses difficult to predict. This variability may be at least partially explained by extent of disease and also by a given NSAID’s physiochemical properties, such as enantiomeric state, differential effect on cyclo-oxygenase isoenzymes, kinetics of distribution within SF and inhibition of other mediators of pain and inflammation not directly related to cyclo-oxygenase or prostaglandin activity [4, 9, 12].

Due to the variability of individual responses, patients may require several medication changes in order to achieve a satisfactory clinical response [4, 9, 12]. As such, guidelines have suggested a trial period of at least 4 weeks when initiating NSAID therapy before switching therapies [13, 14].

Several OA studies have evaluated a variety of baseline clinical variables, ranging from demographics to pain to inflammatory markers and their relation to response. Unfortunately, none of these studies have demonstrated consistent associations [15–18].

Another method less frequently explored is to determine whether an early response (e.g. within 1 or 2 weeks) predicts a later response. While many studies have examined early response and late response, these have almost uniformly looked at population-level data. To the best of our knowledge, only one study has evaluated early and late response with NSAIDs at the patient level in OA. Battisti et al. [19] found that 74% of the patients with good or excellent patient global response at Week 1 maintained good or excellent global response at Week 6 [19]. This study examined only a single indicator of response at the group level. We are unaware of any study performing a similar analysis at the individual patient level using the composite Osteoarthritis Research Society International and the OMERACT (OMERACT-OARSI) responder criteria [20] to define response, or studies that evaluated response beyond 6 weeks.

We previously reported the data pooled from two identical studies comparing etoricoxib 30 mg, celecoxib 200 mg and placebo in knee or hip OA [21] showing that the proportion of OMERACT-OARSI responders was significantly greater in the etoricoxib (66.2%) and celecoxib (63.5%) groups compared with the placebo group (43%; P < 0.001 for active drug vs placebo) at a pre-specified primary endpoint of 12 weeks [22]. The purpose of the current analysis was to identify the proportion of responders, defined by OMERACT-OARSI criteria, after 2, 4 and 8 weeks of treatment who were also responders after 12 weeks and to examine whether there were differences between active treatments and placebo, using patient-level data pooled from these two studies [21].

Patients and methods

Study design and patients

Details from the original studies have been reported [21, 22]. Briefly, the studies enrolled patients of ≥ 40 years of age with a diagnosis of OA of the hip or knee >6 months (ARA functional class I, II or III), who required prescription-strength NSAIDs on a regular basis with a history of therapeutic benefit. Clinical efficacy and safety data were collected at 2, 4, 8, 12, 16...
and 26 weeks. The three co-primary efficacy endpoints were the changes from baseline to 12 weeks in WOMAC Index Pain subscale, WOMAC Physical Function subscale and patient global assessment of disease status (PGADS). The protocols for the original studies were approved by the institutional review board or ethics review board for each site, and all patients provided written informed consent prior to participating in the studies, consistent with the principles of the Declaration of Helsinki.

**Statistical analysis**

The present analysis was limited to data from the 12-week placebo-controlled part of the studies before placebo patients were switched to active treatment. Because of the identical and concurrent nature of the two separate trials, data from the two trials were pooled to provide a more robust dataset from which to perform these post hoc analyses. In addition, because of the generally similar baseline characteristics and efficacy results from the two studies [21], and because the number of responders in the two studies were similar to each other as well as to the pooled population [21, 22], we felt that the pooled patient groups were the representative of the individual treatment groups and would not likely be subject to a confounding cohort effect. As there was no significant interaction between index study joint and co-primary efficacy outcomes as reported in the main study (P-value range 0.765–0.870) [21], and as only ~20% of the patients identified hip as the index OA joint [21], we also pooled patients together irrespective of primary joint studied (i.e. hip or knee). However, to confirm whether index joint was a factor for response agreement, we performed a logistic regression with covariates of treatment and index joint for remaining in same response category across Weeks 2, 4, 8 and 12.

The present analysis used a modified intent-to-treat approach; all patients with a baseline efficacy measurement and at least one post-baseline observation were included. Only observed data were included to define OMERACT-OARSI response status; missing data were not carried forward or imputed. OMERACT-OARSI response was determined for each patient at each timepoint (Weeks 2, 4, 8 or 12), using the corresponding changes from baseline in WOMAC Pain and Physical Function subscales, and PGADS, all measured on a 100-mm visual analogue scale (VAS). OMERACT-OARSI criteria were defined as: (i) improvement in pain or physical function ≥50% with an absolute change ≥20 mm; or (ii) improvement of ≥20% with an absolute change ≥10 mm in at least two of the following: pain, physical function and patient global assessment, as measured on a 100-mm VAS [20].

We evaluated the proportion of patients who remained in the same responder status (responder or non-responder) from the early timepoints of Weeks 2, 4 or 8 to the later timepoint of Week 12. To evaluate the durability of response, we also evaluated the proportion of patients who remained as consistent responders or non-responders at all timepoints (i.e. at 2, 4, 8 and 12 weeks). Finally, to measure overall agreement between early and later timepoints, we assessed the combined consistency at Weeks 2, 4, 8 and 12 of response irrespective of response status.

Of the 1207 total patients, ~20% had an undefined responder status due to missing data. As the drop-out rate in the placebo group (41%) was significantly higher than in the etoricoxib (15%) or celecoxib (19%) groups (P < 0.001), primarily due to lack of efficacy (29% placebo, 6% etoricoxib and 9% celecoxib; P < 0.001 for active treatments vs placebo), an undefined response was treated as a non-responder for the purpose of this analysis. While this more conservative approach overestimates the proportion of non-responders and underestimates the proportion of responders, it has no impact on the number of patients meeting the responder criteria. A separate analysis excluding patients with undefined response was also conducted.

Among patients who were responders or non-responders at early timepoints (Weeks 2, 4 or 8), the 95% CIs for the proportion of patients who remained in their early response category at Week 12 were reported from binomial distribution using the exact method. The agreement between the status at the early timepoint and at Week 12 was evaluated using κ-coefficient and reported with corresponding 95% CIs. The P-values comparing two proportions were reported using Fisher’s exact test. A P-value < 0.05 was considered statistically significant. Due to the exploratory nature of this post hoc analysis, no multiplicity adjustment was employed.

**Results**

**Patient accounting**

In the pooled study population, 1207 patients were randomly assigned to etoricoxib 30 mg (n = 475), celecoxib 200 mg (n = 488) or placebo (n = 244) once daily for 12 weeks during the placebo-controlled portion of the study. Overall, patient characteristics were consistent across treatment groups (Table 1). The proportion of OMERACT-OARSI responders was significantly greater in the etoricoxib (59.8%) and celecoxib (57%) groups compared with the placebo group (34%; P < 0.001 for active drug vs placebo) after 12 weeks of treatment when treating undefined status as non-responder.

**Early vs late response**

**Responders.** A large majority (~76%) of the patients receiving active treatment who were responders at Week 2 remained responders at Week 12, and significantly more patients receiving etoricoxib (P = 0.001) or celecoxib (P < 0.01) who were responders at Week 2 remained responders at Week 12 than patients receiving placebo (Fig. 1A). The proportion of patients in the active treatment groups at Week 8 remaining as responders at Week 12 was higher than the proportion at Weeks 2–12, although the absolute number of responders was fairly stable. In contrast, in the placebo group more patients became responders at later timepoints. Significantly more patients remained as consistent responders over time with etoricoxib (P < 0.01) or celecoxib (P < 0.05) than with placebo.

**TABLE 1. Summary of patient characteristics and select outcomes, pooled population**

<table>
<thead>
<tr>
<th></th>
<th>Etoricoxib 30 mg, n = 475</th>
<th>Celecoxib 200 mg, n = 488</th>
<th>Placebo n = 244</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>323 (68)</td>
<td>321 (65.6)</td>
<td>159 (65.2)</td>
</tr>
<tr>
<td>Age, mean±s.d., years</td>
<td>62 (9.9)</td>
<td>62.4 (9.4)</td>
<td>61.9 (9.2)</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>4 (0.8)</td>
<td>3 (0.6)</td>
<td>2 (0.8)</td>
</tr>
<tr>
<td>Black</td>
<td>33 (6.9)</td>
<td>40 (8.2)</td>
<td>22 (9)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>13 (2.7)</td>
<td>15 (3.1)</td>
<td>8 (3.3)</td>
</tr>
<tr>
<td>White</td>
<td>423 (89.1)</td>
<td>427 (87.5)</td>
<td>209 (85.7)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (0.4)</td>
<td>3 (0.6)</td>
<td>3 (1.2)</td>
</tr>
<tr>
<td><strong>Primary study joint, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knee</td>
<td>367 (77.5)</td>
<td>385 (78.9)</td>
<td>202 (82.8)</td>
</tr>
<tr>
<td>Hip</td>
<td>108 (22.7)</td>
<td>103 (21.1)</td>
<td>42 (17.2)</td>
</tr>
<tr>
<td><strong>Prior medicine use, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetaminophen</td>
<td>58 (12.2)</td>
<td>80 (16.4)</td>
<td>40 (16.4)</td>
</tr>
<tr>
<td>Nsaid/Coxib</td>
<td>417 (87.8)</td>
<td>408 (83.6)</td>
<td>204 (83.6)</td>
</tr>
<tr>
<td><strong>Low-dose aspirin use, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (flare) efficiency scores, mean±s.d., mm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOMAC Pain</td>
<td>68 (16.3)</td>
<td>67.3 (17.5)</td>
<td>66.4 (16.4)</td>
</tr>
<tr>
<td>WOMAC Physical Function</td>
<td>66.6 (17.7)</td>
<td>66.1 (18.8)</td>
<td>64.9 (18.2)</td>
</tr>
<tr>
<td>PGADS</td>
<td>72.7 (17)</td>
<td>70.7 (17.8)</td>
<td>70.6 (17.7)</td>
</tr>
</tbody>
</table>
Non-responders. Most patients who were non-responders at Week 2 were non-responders at Week 12, with similar incremental increases at later weeks irrespective of treatment group (Fig. 1B). As seen in the responder groups, although agreement between early and late timepoints increased with time, the absolute number of non-responders was similar across timepoints, indicating that non-response was primarily determined by Week 2. The proportion of placebo non-responders was numerically greater than for active treatment groups at Week 2 and significantly greater at Weeks 4 and 8 \( (P < 0.05) \). There were also significantly more consistent non-responders with placebo than with active treatments \( (P < 0.01) \). There were no significant differences between etoricoxib and celecoxib in either responders or non-responders at any timepoint.

**Overall response agreement**

The proportion of patients remaining in a given response status (i.e. either responder or non-responder) was high (~73%) for Weeks 2 and 12, and increased slightly for the later timepoint comparisons (Fig. 2). The \( \kappa \)-values at Weeks 2 and 12 indicate moderate agreement for the etoricoxib and celecoxib groups, but low agreement for the placebo group (Table 2), consistent with the comparatively greater variability in response over time in the placebo group. Response status agreement between Weeks 4 or 8 and Week 12 is numerically higher than that between Week 2 and 12 for all treatment groups and indicates moderate to high agreement based on \( \kappa \)-values.

When results were analysed excluding patients with unknown response, there were relatively small increases in the proportion of responders and decreases in the proportion of non-responders, as expected. Regardless of approach, the patterns of response and \( \kappa \)-values were similar (data not shown).

The logistic regression with treatment and index joint (hip vs knee) as covariates showed that index joint did not significantly affect response agreement between early timepoints and Week 12 \( (P = 0.965 \text{ when treating undefined response as non-responder}; \ P = 0.932 \text{ when excluding undefined response}) \).
The primary purpose of this analysis was to determine whether early response to treatment with etoricoxib, celecoxib or placebo predicts later response, and if so, to examine at which timepoint a response is primarily established. Because OA symptoms may be variable over time, we also evaluated the consistency of response by assessing the proportion of patients who maintained the same variable over time, we also evaluated the consistency of response during the study period beyond 2 weeks, as few, if any, additional responders were identified at 4 or 8 weeks. Our findings corroborate a previous study suggesting that the NSAID efficacy is most apparent in the initial 4 weeks of treatment [23]. As reported in the primary results of our study population, the largest proportion of improvement from baseline (approximately –23 mm) in WOMAC Pain, WOMAC Physical function and PGADS was observed by Week 2, with only a small additional benefit (approximately –3 mm further decrease) at Week 12 [21]. In that sense, individual patient response parallels the overall study results: just as the greatest magnitude of improvement occurs by Week 2, the greatest number of individual responders is identified by Week 2.

It is important to note that, although our study and the study by Battisti et al. show that response is most likely to appear within the first 2 weeks of treatment, the overall effect of treatment may continue to change over time. For example, OA symptoms may increase, necessitating more aggressive interventions beyond NSAID treatment, or patients may discontinue therapy due to adverse effects, treatment failure, or patient preference.

While there were only relatively small differences in the proportion of patients remaining in the same response category at Week 2 and at Week 12 for the active and placebo groups, the difference in the $\kappa$-coefficients for agreement between these two timepoints was notable: 0.459 for etoricoxib and 0.449 for celecoxib, both indicating moderate agreement; and 0.357 for placebo, indicating low agreement. These results reflect a greater variability in placebo responses over time.

We are aware of only one other study in OA using NSAIDS that evaluated overall response as a function of early response at the patient level. Battisti et al. [19] pooled results from four OA trials comparing rofecoxib to celecoxib, acetaminophen and placebo, or to nabumetone and placebo. They found that 74% of the responders and 76% of the non-responders [defined by patient global assessment of response to therapy (PGART)] at Day 6 remained in their response category at Week 6 (Pearson’s $r = 0.675$ for the correlation between Day 6 and Week 6 response). Their results were consistent across studies and treatment groups and are similar to the results of our analysis [19].

Current OA guidelines recommend NSAID trial periods of 4 weeks [13, 14]. However, our results, as well as those from Battisti et al. [19] suggest that NSAID response is likely to occur within the first 2 weeks of treatment. Moreover, our results show that there appears to be little benefit in extending an NSAID trial period beyond 2 weeks, as few, if any, additional responders were identified at 4 or 8 weeks. Our findings corroborate a previous study suggesting that the NSAID efficacy is most apparent in the initial 4 weeks of treatment [23]. As reported in the primary results of our study population, the largest proportion of improvement from baseline (approximately –23 mm) in WOMAC Pain, WOMAC Physical Function and PGADS was observed by Week 2, with only a small additional benefit (approximately –3 mm further decrease) at Week 12 [21]. In that sense, individual patient response parallels the overall study results: just as the greatest magnitude of improvement occurs by Week 2, the greatest number of individual responders is identified by Week 2.

It is important to note that, although our study and the study by Battisti et al. show that response is most likely to appear within the first 2 weeks of treatment, the overall effect of treatment may continue to change over time. For example, OA symptoms may increase, necessitating more aggressive interventions beyond NSAID treatment, or patients may discontinue therapy due

### Table 2. $\kappa$-Coefficients (95% CI) for the agreement between early timepoints and Week 12

<table>
<thead>
<tr>
<th>Week</th>
<th>Etoricoxib 30 mg</th>
<th>Celecoxib 200 mg</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.459 (0.378, 0.541)</td>
<td>0.449 (0.369, 0.529)</td>
<td>0.357 (0.234, 0.480)</td>
</tr>
<tr>
<td>4</td>
<td>0.546 (0.469, 0.624)</td>
<td>0.526 (0.450, 0.603)</td>
<td>0.567 (0.457, 0.677)</td>
</tr>
<tr>
<td>8</td>
<td>0.650 (0.580, 0.720)</td>
<td>0.676 (0.610, 0.745)</td>
<td>0.697 (0.603, 0.791)</td>
</tr>
</tbody>
</table>

$\kappa$-Values were calculated based on analysis treating unknown response as non-responders.

### Discussion

The primary purpose of this analysis was to determine whether early response to treatment with etoricoxib, celecoxib or placebo predicts later response, and if so, to examine at which timepoint a response is primarily established. Because OA symptoms may be variable over time, we also evaluated the consistency of response by assessing the proportion of patients who maintained the same response category at all studied timepoints.

We found that response status at Week 12—both responder and non-responder—was primarily established by Week 2, when it was established in ~73% of the patients. Response status agreement between Weeks 8 and 12 increased to ~85%, which is to be expected, as variability in response decreases with smaller time intervals between the early and later timepoint. Importantly, although agreement did increase from Weeks 2 and 12 to Weeks 4 or 8 and Week 12, the overall number of responders or non-responders at a given timepoint remained quite stable, further indicating that response status was primarily determined by Week 2. Thus, while overall response status agreement increased, few, if any, additional patients became new responders or non-responders after Week 2.

We also found that patients receiving etoricoxib or celecoxib who were responders at Week 2 were significantly more likely to remain responders at Week 12 compared with patients receiving placebo. Conversely, patients receiving placebo who were non-responders at Week 2 were numerically more likely to remain non-responders at Week 12 than patients receiving active treatment. Furthermore, they were also significantly more likely to remain non-responders between Weeks 4 and 12 and Weeks 8 and 12, indicating some ongoing conversion from non-response to response in the active groups, although absolute numbers were fairly small.
to adverse experiences. Indeed, duration of use or time to discontinuation have previously been used to evaluate and differentiate NSAID efficacy [24–26] and may be useful measures of a medication’s overall utility (i.e. its effectiveness) because they take into account both efficacy and tolerability.

Our study has a few limitations. First, this was neither a pre-specified analysis nor was it designed or powered to assess agreement between responses at Week 2 to responses at Week 12. However, given a similar analysis that prospectively addressed correlation of early response and long-term response in AS [27], our sample size of 1207 patients is likely sufficiently robust. Second, we limited our analysis to the 12-week placebo-controlled portion of the original studies. Because the placebo groups in Part I of the original study were switched to active treatment in Part II (Weeks 12–26), the 12-week data would show the true treatment effect, whereas the 26-week data could have had a confounding effect, with patients in all groups more likely to remain in the study in order to receive active drug after 12 weeks. This approach would, if anything, likely decrease the numbers of non-responders across all groups. Alternatively, we could have excluded the placebo group altogether and followed only the patients who received active treatment in Part I through Part II. Although the active group results are the most clinically relevant, because there is a sizable placebo response in OA NSAID studies [22, 28–31], we felt that the placebo group should be included in the analysis in order to provide a more accurate picture. In our study as in prior NSAID studies, the placebo benefit was sizeable. Third, as noted above, patients with unknown response status were treated as non-responders to account for the high dropout rate in the placebo group. Although the analysis excluding patients with unknown response confirmed these results, it is possible that patients who dropped out may have responded had they remained in the study. However, it is likely that these patients would experience similar patterns of response or non-response as those included in the analyses. Fourth, this analysis was limited to data pooled from two trials because of their identical, concurrent nature, and because it was a direct extension of previous analyses [22]. Whether these results are generalizable to the larger OA population, to patients with other rheumatological diseases requiring NSAIDs, or to otherwise healthy individuals who require as-needed NSAIDs, would necessitate further investigation. Fifth, as noted, previous studies have not found a consistent relationship between patient characteristics and NSAID response [15–18]. As such, we did not stratify our findings by features such as age, gender or ARA class. It is possible that there is an interaction between early response and demographics. Finally, the earliest timepoint available for this analysis was at Week 2. It is possible that response is established earlier than Week 2, as demonstrated by Battisti et al. [19]: future studies evaluating daily patient-level data during the first 2 weeks of treatment would provide further insight into when response is primarily established.

In summary, we found the majority of patients classified as OMERACT-OARSI responders at 12 weeks were classified as responders after 2 weeks of treatment, and those patients who do not respond after 2 weeks were unlikely to respond thereafter. These data suggest that clinicians may be able to limit NSAID trial periods to less than the currently recommended 4 weeks in their patients. The question of when treatment should be adjusted due to lack of response in clinical practice should be left to clinical judgement, taking into account patients’ disease course and severity.

Rheumatology key messages

- Early response to NSAIDs is highly predictive of later response.
- NSAID response is primarily determined within the first 2 weeks of treatment.

Acknowledgements

All authors are responsible for the work described in this article, and they contributed to the conception, data interpretation and drafting of the manuscript and/or revising the manuscript for important intellectual content. All authors provided final approval of the version to be published. The authors would like to thank Qian Dong, MS, of Merck & Co., Inc. for computational assistance and Martha Vollmer, MA, of Merck & Co., Inc. for her assistance with the preparation of this manuscript. No writing assistance was provided. The original studies and the present analysis were supported by Merck & Co., Inc. The sponsor’s roles included study design, data collection, analysis and interpretation, performing analysis for the manuscript, assisting in manuscript development and the decision to submit the manuscript for publication.

Funding: This study was supported by Merck & Co.

Disclosure statement: S.S.S. is an employee of Merck & Co. and holds stock options in this company. C.O.B. has served as a clinical trial investigator for Merck, Pfizer, Novartis and McNeill; and as a consultant to Merck, Novartis and McNeill. A.M.T. is an employee of Merck & Co. Inc., owns stock and holds stock options in Merck. H.W. is an employee of Merck & Co., Inc., owns stock and holds stock options in the Company.

References


