Effectiveness of Anti-Tumor Necrosis Factor Agents in the Treatment of Rheumatoid Arthritis: Observational Study

Anna D’Souza, PhD; Brian L. Meissner, PharmD, PhD; Boxiong Tang, MD, PhD; R. Scott McKenzie, MD; Catherine T. Piech, MBA

Objective: The efficacy of anti-tumor necrosis factor therapies in rheumatoid arthritis has been demonstrated in randomized clinical trials. The purpose of the present study was to evaluate the effectiveness of these agents for the treatment of rheumatoid arthritis in a real-world setting.

Method: This retrospective chart review included patients from 6 clinics in the United States. Eligibility criteria included age ≥18 years, diagnosis of rheumatoid arthritis, and having been initiated with anti-tumor necrosis factor therapy (i.e., adalimumab, etanercept, or infliximab) between January 1, 2002, and November 30, 2004. Patients were assessed for up to 2 years after therapy initiation. Primary outcomes of interest were improvements in 4 effectiveness measures—joint pain, joint swelling, joint stiffness, and fatigue. A total of 496 patients met the study’s inclusion criteria: 84 (16.9%) in the adalimumab group, 146 (29.4%) in the etanercept group, and 266 (53.6%) in the infliximab group.

Results: Improvement in 1 of the 4 effectiveness measures was documented in 36.8% (n = 25) who received adalimumab, in 47.7% (n = 62) of those who received etanercept, and in 48.7% (n = 115) of patients who received infliximab. The infliximab group was the only cohort to demonstrate significant improvements from baseline in joint pain, joint swelling, and joint stiffness. The adalimumab group had significant improvement in joint pain (P = .004). No significant change in fatigue scores was reached with any of these agents.

Conclusion: In the real-world setting of patients with rheumatoid arthritis, anti-tumor necrosis factor therapy shows significant improvements in joint pain, joint swelling, and joint stiffness, although there are differences in effectiveness in the 4 measures among the 3 agents assessed in this study.

Significant limitations have been noted in the literature as to outcomes of clinical trials and observational studies of anti-TNF therapy. Clinical trial inclusion and exclusion criteria specify select patient populations, which may not necessarily reflect the diversity of patient and disease characteristics in real-world clinical scenarios. Therefore, real-world studies can offer unique insights over controlled clinical trials.

Although there is greater diversity of patient and disease characteristics in real-world studies, there may be bias, because patients receive various treatments. Nevertheless, healthcare decision makers are increasingly relying on studies of real-world outcomes for decisions related to coverage and reimbursement. Many studies have evaluated the effectiveness of anti-TNF therapies in patients with RA, but few real-world studies have...
evaluated the clinical effectiveness of adalimumab, etanercept, and infliximab. The objective of this analysis was to assess the effectiveness of these agents in improving joint pain, joint swelling, joint stiffness, and fatigue among patients with RA.

Methods

We conducted a retrospective, observational chart review using data from 6 rheumatology clinics across the United States, and evaluated treatment effectiveness among a sample of patients with RA who were treated with anti-TNF therapy (ie, adalimumab, etanercept, or infliximab). The study period was from January 1, 2001, to November 30, 2006. For each patient, the observation period included a 1-year period before the index date, which was the date of the first anti-TNF prescription or administration, and a 2-year follow-up period after the index date (Figure). Baseline status was assessed during the year before the index date. Data were collected for a maximum of 2 years after the index date, or until discontinuation of anti-TNF therapy, whichever occurred first.

Patient selection was based on the following inclusion criteria: age ≥ 18 years; diagnosis of RA; and therapy initiated with adalimumab, etanercept, or infliximab between January 1, 2002, and November 30, 2004, with no documentation of previous anti-TNF therapy. Patients were required to be under the care of the participating physician or physician group, or have complete documentation (related to RA) from another provider for a minimum of 12 months before the index date and up to 24 months after the index date (no minimum follow-up period was required).

Patients with a diagnosis of psoriasis, ulcerative colitis, Crohn's disease, or ankylosing spondylitis were excluded from the study, because of the differences in anti-TNF dosage recommendations. In addition, patients were excluded if they had a history of anti-TNF therapy or any of the following diagnoses any time before the index date—malignancy (eg, lymphoma, solid organ malignancy), infection (eg, tuberculosis, bacterial pneumonia), seizures, demyelinating disorders (eg, multiple sclerosis), heart failure, or drug-induced lupus.

Participating clinics identified and provided charts for all patients meeting the study inclusion criteria. Patient privacy was maintained through the use of unique alphanumeric values for each patient in compliance with the Health Insurance Portability and Accountability Act (HIPAA). The study protocol was approved by the New England Institutional Review Board.

Data were collected using a standardized chart review form with a graphical user interface developed in Microsoft Access. Patient demographics and concomitant DMARD use were collected. For this study, concomitant DMARD therapy included hydroxychloroquine, leflunomide, sulfasalazine, azathioprine, penicillamine, and auranofin.

Response to therapy was evaluated for joint pain, joint swelling, joint stiffness, and fatigue. Because of intra- and intervariability in documentation among physicians, a data abstraction form was used so that the chart reviewer could document both qualitative and quantitative data.
An algorithm was developed to translate qualitative data into quantitative values. Because there were no relevant studies available to guide the development of the algorithm, a clinical and psychometric rationale was applied to develop an algorithm that could be consistently applied to the varying levels of physician documentation.

A visual analog scale (VAS; 0-10 cm) was the preferred method for evaluating all 3 joint assessments (ie, pain, swelling, stiffness) and fatigue. If documented VAS was not available, the following algorithm was used:

A. Categorical documentation was converted to a numerical value using midpoints of a 0-to-10 scale (median range: mild = 1.65 [1-3.3], moderate = 4.95 [3.4-6.6], severe = 8.25 [6.7-9.9])

B. Dichotomous documentation was converted to a numerical scale using the midpoints of a 1-to-10 scale (yes = 5, no = 0).

If more than 1 form of documentation was present for a patient visit, the VAS was the preferred type of documentation, followed by categorical and dichotomous documentation.

Baseline scores were determined through available documentation on the index date or, if not available, documentation from the closest encounter before the index date. The mean scores for joint pain, swelling, and stiffness, and fatigue were determined for the 2-year follow-up period. If a patient had multiple assessments, an average was derived.

The following criteria were used to define improvement from baseline:

A. Decrease in VAS of at least 1 point
B. Change in categorical documentation from “severe” to “moderate/mild” or from “moderate” to “mild”
C. Change in dichotomous documentation from yes to no.

Improvement was defined as any improvement during the 2-year follow-up period among any of the 3 documentation types. Overall improvement was based on improvement in any of the 4 effectiveness measures. Mean scores were based only on patients with documented information for each joint assessment at the given time point (baseline or follow-up period).

Basic descriptive statistics and univariate statistical testing were applied to this analysis. Analysis of variance or Kruskal-Wallis tests (as applicable) were applied to continuous variables to determine if differences existed among the 3 cohorts. If differences did exist (ie, $P < .05$), post-hoc multiple comparison tests (Tukey or Mann-Whitney) that controlled for the family-wise type 1 error rate were conducted. For the multiple comparison tests, $P < .017$ (Bonferroni correction) was considered statistically significant. Paired t-tests were conducted to determine if differences existed among the VAS scores at baseline and during the 2-year follow-up period within each cohort. Chi-square tests were conducted to assess differences between the 3 cohorts on categorical variables. All analyses were conducted using SAS version 9.1 (Cary, NC).

**Results**

This study enrolled 496 patients, including 84 patients (16.9%) in the adalimumab group, 146 (29.4%) in the etanercept group, and 266 (53.6%) in the infliximab group (Table 1). Overall, 74.0% of the patients were female, and the mean age (standard deviation [SD]) was 56.1 (14.8) years. The mean age was 52.6 years in the etanercept group and 57.6 years in the infliximab group ($P < .017$).

Of patients with documented ethnicity ($n = 217$), whites comprised 80.2% of the overall population. The mean (SD) length of disease duration before initiation of anti-TNF therapy was 3.9 (3.5) years, and the mean (SD) length of follow-up after the index date was 614.8 (156.0) days. Approximately 76% of patients received concomitant DMARD therapy, which was more common (79.7%) in the infliximab cohort than in the etanercept (68.5%) or adalimumab (75.0%) cohorts ($P < .040$). The average time that patients remained on anti-TNF therapy ranged from 367 days to 423 days (etanercept, 423 days; adalimumab, 371 days; infliximab, 367 days; $P = .126$).

**Overall Clinical Improvement**

There was wide variation in the type of documentation used to indicate patient improvement (Table 2). For example, 99% of the patients had documentation for joint pain, using a dichotomous approach (yes, pain exists; no, pain does not exist), and 29% had VAS documentation for this assessment.

In total, 12.5% ($n = 62$) of all patients did not have joint assessment data to evaluate clinical improvement during the follow-up period. Of the 434 patients who had documented information for at least 1 of the 4 effectiveness measures, 46.5% ($n = 202$) reported a clinical benefit, including 48.7% ($n = 115/236$) in the infliximab group, 47.7% ($n = 62/130$) in the etanercept group, and 36.8% ($n = 25/68$) in the adalimumab group.

**Joint Pain**

Overall, 150 of the 429 patients (35.0%) had an improvement from baseline in joint pain scores during the 2-year follow-up (Table 3). The adalimumab and infliximab groups reported significant improvements from baseline in joint pain scores ($P = .004$ and $P < .001$, respectively). The infliximab group had the highest percentage of patients with documented improvement in
Joint pain (36.9%) compared with the etanercept group (36.4%) and the adalimumab group (25.3%), although the difference was not significant.

**Joint Swelling**

Approximately 35% (122 of 348) of patients had an improvement from baseline in joint swelling scores. Among all patients, mean (SD) joint swelling scores significantly improved from baseline during the 2-year follow-up period (3.2 [1.9] vs 3.0 [1.8], P = .004). In the infliximab group, the mean (SD) score decreased from 3.2 (1.9) at baseline to 3.0 (1.8) during the follow-up period (P <.001). There were no significant differences between the mean scores at baseline and during the follow-up period in the adalimumab or the etanercept groups. The percentage of patients with improvement in joint swelling was highest in the infliximab group (38.0%), followed by the etanercept group (33.7%), and the adalimumab group (26.8%); however, these differences were not significant.

**Joint Stiffness**

Among all patients, 107 of 380 patients (28.2%) had an improvement in joint stiffness score, and a significant
### Table 3: Improvement in Clinical Effectiveness Measures

<table>
<thead>
<tr>
<th>Clinical measure</th>
<th>Adalimumab</th>
<th>Etanercept</th>
<th>Infliximab</th>
<th>Total</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Joint pain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Baseline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients, n&lt;sup&gt;a&lt;/sup&gt;</td>
<td>68</td>
<td>132</td>
<td>233</td>
<td>433</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) score</td>
<td>4.2 (2.5)</td>
<td>4.4 (2.2)</td>
<td>4.3 (2.2)</td>
<td>4.3 (2.2)</td>
<td>.867</td>
</tr>
<tr>
<td><strong>Follow-up period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients, n&lt;sup&gt;a&lt;/sup&gt;</td>
<td>67</td>
<td>129</td>
<td>233</td>
<td>429</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) score</td>
<td>4.0 (2.4)</td>
<td>4.2 (2.3)</td>
<td>4.0 (2.1)</td>
<td>4.1 (2.2)</td>
<td>.730</td>
</tr>
<tr>
<td>P (vs baseline)</td>
<td>.004</td>
<td>.208</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Patients with improvement, n (%)</td>
<td>17 (25.3)</td>
<td>47 (36.4)</td>
<td>86 (36.9)</td>
<td>150 (35.0)</td>
<td>.200</td>
</tr>
<tr>
<td><strong>Joint swelling</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Baseline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients, n&lt;sup&gt;a&lt;/sup&gt;</td>
<td>57</td>
<td>98</td>
<td>205</td>
<td>360</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) score</td>
<td>3.3 (1.7)</td>
<td>3.2 (1.9)</td>
<td>3.2 (1.9)</td>
<td>3.2 (1.9)</td>
<td>.880</td>
</tr>
<tr>
<td><strong>Follow-up period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients, n&lt;sup&gt;a&lt;/sup&gt;</td>
<td>56</td>
<td>92</td>
<td>200</td>
<td>348</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) score</td>
<td>3.3 (1.8)</td>
<td>3.0 (1.9)</td>
<td>3.0 (1.8)</td>
<td>3.0 (1.8)</td>
<td>.456</td>
</tr>
<tr>
<td>P (vs baseline)</td>
<td>.795</td>
<td>.523</td>
<td>&lt;.001</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>Patients with improvement, n (%)</td>
<td>15 (26.8)</td>
<td>31 (33.7)</td>
<td>76 (38.0)</td>
<td>122 (35.1)</td>
<td>.284</td>
</tr>
<tr>
<td><strong>Joint stiffness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Baseline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients, n&lt;sup&gt;a&lt;/sup&gt;</td>
<td>63</td>
<td>115</td>
<td>214</td>
<td>392</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) score</td>
<td>3.8 (2.0)</td>
<td>3.8 (1.8)</td>
<td>3.8 (1.8)</td>
<td>3.8 (1.9)</td>
<td>.986</td>
</tr>
<tr>
<td><strong>Follow-up period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients, n&lt;sup&gt;a&lt;/sup&gt;</td>
<td>60</td>
<td>111</td>
<td>209</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) score</td>
<td>3.6 (2.2)</td>
<td>3.9 (2.0)</td>
<td>3.6 (1.8)</td>
<td>3.7 (1.9)</td>
<td>.351</td>
</tr>
<tr>
<td>P (vs baseline)</td>
<td>.124</td>
<td>.520</td>
<td>.001</td>
<td>.022</td>
<td></td>
</tr>
<tr>
<td>Patients with improvement, n (%)</td>
<td>15 (25.0)</td>
<td>33 (29.7)</td>
<td>59 (28.2)</td>
<td>107 (28.2)</td>
<td>.806</td>
</tr>
<tr>
<td><strong>Fatigue</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Baseline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients, n&lt;sup&gt;a&lt;/sup&gt;</td>
<td>29</td>
<td>63</td>
<td>111</td>
<td>203</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) score</td>
<td>5.0 (1.9)</td>
<td>5.1 (1.8)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4.4 (2.0)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4.7 (1.9)</td>
<td>.031</td>
</tr>
<tr>
<td><strong>Follow-up period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients, n&lt;sup&gt;a&lt;/sup&gt;</td>
<td>19</td>
<td>55</td>
<td>95</td>
<td>169</td>
<td></td>
</tr>
<tr>
<td>Mean (SD) score</td>
<td>4.8 (2.1)</td>
<td>5.2 (1.7)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4.2 (2.0)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4.6 (2.0)</td>
<td>.009</td>
</tr>
<tr>
<td>P (vs baseline)</td>
<td>.631</td>
<td>.153</td>
<td>.296</td>
<td>.290</td>
<td></td>
</tr>
<tr>
<td>Patients with improvement, n (%)</td>
<td>5 (26.3)</td>
<td>21 (38.2)</td>
<td>26 (27.4)</td>
<td>52 (30.8)</td>
<td>.348</td>
</tr>
</tbody>
</table>

<sup>a</sup>Values in this row indicate the number of patients with data during period of assessment.

<sup>b</sup>Indicates that the pairwise difference between etanercept and infliximab is significant (P < .017).

**NOTE:** Scores in each of the outcomes measures (joint pain, swelling, and stiffness, and fatigue) ranged from 0 (no symptoms) to 10 (severe symptoms). SD indicates standard deviation.
There is no standardized measure to assess outcomes in routine clinical practice; therefore, physicians typically utilize several measures.

Nearly half of the patients included in this study reported an improvement in joint pain, joint swelling, joint stiffness, or fatigue during the 2-year follow-up period. The infliximab group was the only cohort to demonstrate significant improvements from baseline in 3 (joint pain, joint swelling, and joint stiffness) of the 4 clinical effectiveness measures. None of the cohorts demonstrated a significant change from baseline in the fatigue score, although numerical improvements were noted.

There is no standardized measure to assess outcomes in routine clinical practice; therefore, physicians typically utilize several measures, including laboratory tests, radiographic scores, joint counts, measures of functional status, global measures, and patient self-reported questionnaires.16,17 Indeed, wide variability has been demonstrated regarding the frequency with which different outcome measures are documented by providers.17

The present analysis assessed variable outcome measures with the development of an algorithm to integrate intra- and intervariability among providers reporting clinical status in patients with RA. A number of tools are currently available (eg, Simplified Disease Activity Index; Clinical Disease Activity Index; Disease Activity Score, including 28 joints).18,19 However, further exploration of a validated practical tool that could be adopted for widespread use to monitor RA treatment response is warranted, followed by exploration of the incentives needed to improve such data collection.

An advantage of this study is that it provides real-world outcomes associated with anti-TNF therapy for drugs that have already demonstrated efficacy and safety in controlled clinical trials. The 2-year follow-up period allowed for the assessment of long-term outcomes. However, because this analysis was designed to evaluate anti-TNF agents, other biologic therapies used for RA— including anakinra (Kinereit),20 abatacept (Orencia),21 and rituximab (Rituxan)22—were not assessed.

Randomized clinical trials and observational studies in routine clinical practice are important for the evaluation of therapies and have unique advantages and limitations. Although not as scientifically rigorous, real-
world studies can complement the information gained from clinical trials. Specifically, one observational cohort study of the effectiveness of anti-TNF therapy in patients who would have been ineligible for a clinical trial demonstrated that the majority of such patients would benefit from these treatments.\textsuperscript{23} Well-designed observational studies can be invaluable in gaining further understanding of response to therapy.

\textbf{Limitations}

Consistent with observational research, the limitations of this study include missing data and inconsistencies with documentation. Lack of standardized outcome measures in clinical practice was confirmed by the results and limited the quantitative and qualitative nature of this study, thereby requiring the development of an algorithm to capture variability in assessment of clinical outcomes among providers. As such, this approach has not been validated. Although the study was designed to evaluate patients who were newly initiated with anti-TNF therapy, it is possible that patients might have received anti-TNF treatments before the pre-index assessment period.

Well-designed observational studies can be invaluable in gaining further understanding of response to therapy.

Notable limitations of the study are inherent in the nature of observational studies. There exists the possibility of confounding by indication associated with lack of randomization, because other unmeasured factors might have influenced the selection of one anti-TNF agent over another, and variable reasons for changes in therapy may influence the outcomes. Also, the retrospective nature of the analysis prevented the collection of patient-reported outcomes. In addition, quantification of tender or swollen joints and radiographs to assess joint damage were not available.

These factors may have also contributed to the lack of significant improvements among patients who received etanercept in this analysis. Although these patients did not show significant improvements in the clinical response measures used in this study, etanercept has been shown to be effective for treating RA in randomized, placebo-controlled clinical trials.\textsuperscript{46} This discrepancy may also be a result of the different methods used in measuring treatment response.

\textbf{Conclusion}

Data from this analysis demonstrate the 2-year clinical effectiveness of anti-TNF therapy in patients with RA. This real-world study shows improvements in joint pain, joint swelling, joint stiffness, and fatigue. Overall clinical improvement was reported in 36.8% of the adalimumab group, 47.7% of the etanercept group, and 48.7% of the infliximab group over the 2-year follow-up period. These results are beneficial in that they provide confirmation that benefits seen in clinical trials are being translated into actual practice. Future observational research, with more uniform patient and provider assessments during the course of anti-TNF therapy, is warranted.

\textbf{Acknowledgments}

The authors thank Rebecca E. Clemente, PhD, and Robert Achenbach of Centocor Ortho Biotech, Inc, for editorial support.

\textbf{Disclosure Statement}

This study was supported by funding from Centocor Ortho Biotech, Inc.

\textbf{References}

Effectiveness of Anti-TNF Agents in Rheumatoid Arthritis

STAKEHOLDER PERSPECTIVE

Effectiveness of Anti-TNFs in Patients with RA, and Coverage Considerations

Payers: Over the past several years, rheumatoid arthritis (RA) has become a difficult condition to manage from a managed care standpoint. Drugs for the treatment of RA are among the few medication classes to keep a double-digit trend year after year and have become some of the largest cost contributors on a percentage of per-member-per-month basis.

Anti-tumor necrosis factor (TNF) agents discussed in this article were the first medications for RA that broke $20,000 for 1 year of treatment. Recently, several anti-TNF biologics have been introduced to the market that have different modes of action but similar efficacy and similar cost. For these reasons, RA drugs have become a prominent concern for payers, including health insurance companies and employer groups. However, very little comparative data are available for these agents. Until such data become available, coverage decisions will need to be based solely on placebo-controlled trials, postmarketing claims-based or chart-based retrospective analysis, and cost.

In 2008, the American College of Rheumatology (ACR) released its recommendations for the use of both nonbiologic disease-modifying antirheumatic drugs (DMARDs) and biologics.1 The recommendations emphasize the importance of starting therapy with traditional DMARDs, which may be much more cost-effective than starting immediately with a biologic. Another ACR recommendation is using a concomitant DMARD when a biologic is needed. The majority of clinical trials show a better response with biologics when used in combination with a DMARD such as methotrexate.

However, several retrospective claims analyses have shown the combination use rate to be lower than might be expected (based on efficacy results from pivotal clinical trials). Another challenge has been the need to encourage providers to prescribe combination therapy in support of available evidence-based recommendations.

Drug manufacturers will likely hesitate to conduct comparative trials for their products. This means more reliance on retrospective analysis or “real-world” data to determine if enough data are available to recommend coverage of one biologic over another for first-line treatment, after failure of a different biologic, as well as the extent or length of coverage for these products.

Patients: The past decade has given patients with RA several novel biologic options for the treatment of this painful, progressing condition. The variety of options also gives patients the opportunity to receive therapy at their provider’s office on a regular basis, or the option to receive a self-injectable therapy from their pharmacy or specialty pharmacy.

Besides the difficult decision of determining the most appropriate therapy based on efficacy and safety, is the necessity of cost-sharing. Patients need to be insurance-savvy to realize not only which products may be covered but also the type and amount of cost-sharing that is required. Medical versus pharmacy benefit differences may also create challenges to understanding optimal coverage opportunities.


Matthew Mitchell, PharmD, MBA
Manager, Pharmacy Services
SelectHealth, Salt Lake City, UT