Comparative Efficacy of Tumor Necrosis Factor-α Inhibitors in Ankylosing Spondylitis: A Systematic Review and Bayesian Network Metaanalysis

Runsheng Wang, Abhijit Dasgupta, and Michael M. Ward

ABSTRACT. Objective. To compare the efficacy of 6 tumor necrosis factor–α inhibitors (TNFi) in treatment of ankylosing spondylitis (AS) at 12 weeks and 24 weeks.

Methods. We performed a systematic literature review of randomized controlled trials of TNFi in patients with active AS. We included trials that reported efficacy at 10 to 14 weeks (12-week analysis) and at 24 to 30 weeks (24-week analysis). We used Bayesian network metaanalysis (NMA) to compare their relative efficacy to improve the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Bath Ankylosing Spondylitis Functional Index (BASFI), and C-reactive protein (CRP) level.

Results. We included 20 trials of 6 TNFi, with 43 treatment arms and 3220 participants. All TNFi were significantly better than placebo in reducing BASDAI and BASFI at 12 weeks and 24 weeks; all but certolizumab pegol (CZP) were statistically better than placebo in reducing CRP at 12 weeks; all but CZP and infliximab-dyyb (IFX biosimilar) were significantly better than placebo in reducing CRP at 24 weeks. IFX was superior to the other TNFi in decreasing BASDAI at 12 weeks, but not at 24 weeks. Excluding 1 open-label trial, there were no differences among TNFi.

Conclusion. Based on this NMA of clinical trials, IFX was superior to other TNFi in reducing BASDAI at 12 weeks, but sensitive to inclusion of an open-label trial, and its efficacy was diminished at 24 weeks. The analysis was limited by few direct comparison trials. Further study of relative safety and longterm effectiveness will help inform the choice of TNFi in treating active AS. (First Release January 15 2018; J Rheumatol 2018;45:481–90; doi:10.3899/jrheum.170224)

Key Indexing Terms:
ANKYLOSING SPONDYLITIS
COMPARATIVE EFFECTIVENESS RESEARCH
TUMOR NECROSIS FACTOR-Α
NETWORK METAANALYSIS

From Columbia University Medical Center, New York; National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), National Institutes of Health (NIH), Bethesda, Maryland, USA.

Funding provided by Intramural Research Program, NIAMS, and NIH. RW is a recipient of the Rheumatology Research Foundation Scientist Development Award. The content is the responsibility of the authors and does not necessarily represent the official views of the NIH.

Tumor necrosis factor-α inhibitors (TNFi) have been widely used as a second-line therapy when patients with ankylosing spondylitis (AS) have persistent symptoms despite treatment with nonsteroidal antiinflammatory drugs (NSAID)1. Six different TNFi have been approved for the treatment of AS, including adalimumab (ADA), certolizumab pegol (CZP), etanercept (ETN), golimumab (GOL), infliximab (IFX), and IFX-dyyb (IFX biosimilar). Although they share the same mechanism of action, they are structurally different and have varying efficacy in other conditions in the spondyloarthritis family, including uveitis and inflammatory bowel disease2. It remains unclear whether all TNFi are equally efficacious in relieving the symptoms and signs of active AS. In clinical practice, physicians and patients may favor a particular TNFi over others based on convenience, comorbidities, or cost, rather than a comparison of relative efficacy.

To date, only 2 head-to-head trials of TNFi in AS have been conducted: 1 of ETN versus IFX, and the other of IFX versus IFX-dyyb3,4. In the absence of direct comparisons, indirect comparisons of ≥ 2 medications can be made through a common comparator using network metaanalysis (NMA).
Previous NMA that used the Assessments in Ankylosing Spondylitis 20% response criteria (ASAS20) as the outcome did not detect any difference in efficacy among TNFi\(^5,6,7,8\). However, dichotomous measures such as the ASAS20 are less sensitive than continuous measures in detecting a difference among medications, in part because such measures ignore any differences in efficacy beyond the ASAS20 threshold\(^9\). In our study, we used 3 continuous measures: Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)\(^10\), Bath Ankylosing Spondylitis Functional Index (BASFI)\(^11\), and C-reactive protein (CRP) level as primary outcomes to compare the relative efficacy of 6 TNFi in treatment of active AS.

**MATERIALS AND METHODS**

**Literature search.** The study protocol was registered at PROSPERO (registration number CRD42014014228). We searched PubMed, EMBASE, Scopus, and the Cochrane Database for published randomized controlled trials (RCT) of TNFi in AS through March 31, 2016, in all languages. Searches were performed by a medical informatician, and search terms are summarized in Supplementary Table 1 (available with the online version of this article). We further manually searched reference lists of review articles. Two authors (RW and MMW) reviewed the search results for eligible studies based on selection criteria. Disagreement was resolved by discussion.

Our study was exempted from ethics review by the US National Institutes of Health Office of Human Subjects Research.

**Selection criteria.** We included RCT that evaluated the efficacy of TNFi in adult patients with AS, compared to placebo or to a different TNFi, at 10 to 16 weeks, or at 24 to 30 weeks. AS was defined in the trials by the modified New York criteria\(^12\). To enhance homogeneity, we excluded studies of axial spondyloarthritis, unless a subgroup analysis of patients with AS was reported. TNFi include ADA, CZP, ETN, GOL, IFX, and IFX-dyb. We included studies irrespective of whether they allowed concomitant use of NSAID, corticosteroids, and disease-modifying antirheumatic drugs (DMARD). We excluded studies that were reported only as an abstract.

**Data extraction and assessment of bias.** Data extraction was performed independently by 2 reviewers (RW and MMW). Any discrepancies were resolved by discussion. We extracted features of the study design, characteristics of participants, and relevant outcome measures. The primary efficacy measures were changes from baseline in the BASDAI, BASFI, and CRP. We extracted the mean change score and its SD, or calculated the change from baseline and final scores. When only medians and ranges were reported, we imputed means and SD using standard methods\(^13\). Authors of the original articles or study sponsors were contacted for additional data when needed. Missing SD were imputed using SD of other included studies\(^13\). Intention-to-treat data were collected whenever available.

To assess study quality, we used the Cochrane Collaboration tool for assessment of risk of bias\(^13\). Each study was evaluated on 6 domains (i.e., sequence generation, allocation concealment, blinding, incomplete outcome data, selective reporting, and other sources of bias), and rated as low, unclear, or high risk.

**Statistical analysis.** We performed Bayesian NMA to quantify the relative efficacy of each drug using a random effects model under the assumption of consistency. Bayesian NMA allows the indirect comparison of 2 drugs based on the observed direct effects. For example, the relative effect of drug A and B is the difference of the relative effects of drug A and C and the relative effects of drug B and C; if these direct comparisons are available. We grouped studies that reported outcomes at 10 to 16 weeks for the 12-week analysis, and studies that reported outcomes at 24 to 30 weeks for the 24-week analysis. Bayesian NMA was performed for each outcome at these 2 timepoints. The relative effect size was presented as the mean difference (MD) with 95% credible intervals (CrI). Outcomes of open-label trials may differ from those of blinded trials; on the other hand, open-label trials closely mimic the real-life experience. Therefore, we performed 2 analyses: the first included a single open-label study, and a second analysis excluded this single open-label study.

We assessed the absolute model fit by the overall residual deviance (Dbar)\(^14\). Dbar of each drug should approximate the total number of trial arms included in the metaanalysis when the model fits the data well. We assessed heterogeneity among the trial results using Higgins I\(^2\), which measures the percent of variability in effect estimates that is a result of heterogeneity rather than sampling error\(^15\). Lower I\(^2\) indicates less heterogeneity. To estimate the effect of heterogeneity as a result of differences in initial AS activity among trial participants, we performed metaregression that adjusted for the weighted mean baseline value of each outcome; the BASDAI and BASFI analyses were also adjusted for mean baseline CRP. Because TNFi trials were performed over a span of 15 years, we examined whether there was a drift in placebo responses over time (owing to possibly greater expectations of benefit in later trials), which could affect the direct and indirect comparisons among TNFi.

All analyses were performed using R (version 3.3.1), the R package gmeta (version 0.8.1), and the Markov Chain Monte Carlo engine JAGS (version 4.2)\(^16,17,18\).

**RESULTS**

**Literature review.** We identified 402 articles through a systematic literature search, and 3 additional articles from the reference lists of previous reviews. There were 20 studies included after applying the inclusion and exclusion criteria, consisting of 18 placebo controlled trials and 2 head-to-head comparison trials. A flow diagram that illustrates the study selection process is in Supplementary Figure 1 (available with the online version of this article). A total of 43 trial arms and 3220 patients were included. A summary of study characteristics is presented in Table 1\(^3,4,15–32\). The sample sizes ranged from 40 to 356. Mean age of study participants ranged from 27.4 to 48.0 years, and the mean durations of AS were from 6.8 to 23.0 years. The range of patients who were HLA-B27–positive was 72%–96.2%. Mean baseline BASDAI scores ranged from 5.5 to 6.9 cm (possible range 0–10) on the visual analog scale, mean baseline BASFI scores ranged from 3.2 to 6.7 cm (possible range 0–10) on the visual analog scale, and mean baseline CRP values were from 11 mg/l to 33 mg/l. Fourteen studies reported concomitant use of DMARD, including methotrexate (MTX). Five studies did not permit concomitant use of DMARD, all of which were trials of IFX. One study did not report information on use of DMARD.

The overall study quality was moderate to high (Supplementary Figure 2, available with the online version of this article). One study (5%) was an open-label trial, and therefore was graded as high risk for bias in blinding of participants, personnel, and outcome assessment. Five studies (25%) had high risk of bias as a result of selective reporting because of missing data, but most were provided by trial investigators or sponsors on inquiry. One study (5%) reported covariate-adjusted mean values instead of raw means, and was considered unclear risk for other bias.
Networks of evidence and comparison to placebo. Eighteen trials (including 39 arms, 2900 participants) were included in the analysis of relative efficacy at 12 weeks. The networks of comparisons for the BASDAI, BASFI, and CRP are presented in Figure 1A-C. Model fit was good (Supplementary Table 2, available with the online version of this article). All TNFi were significantly more efficacious than placebo in reducing BASDAI and BASFI scores [relative effect size for BASDAI reduction range from −2.66 to −1.45 mean difference (MD); for BASFI reduction from −1.99 to −1.05 MD], and all but CZP were significantly better than placebo in decreasing CRP (relative effect size from

<table>
<thead>
<tr>
<th>First Author</th>
<th>Year</th>
<th>Drug and Dose</th>
<th>Enrollees, Length, Age, Male, %</th>
<th>HLA-B27, %</th>
<th>Disease Duration, yrs</th>
<th>Baseline BASDAI, cm</th>
<th>Baseline BASFI, cm</th>
<th>Baseline CRP, mg/l</th>
<th>MTX use, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van der Heijde</td>
<td>2006</td>
<td>ADA 40 mg QOW</td>
<td>208 24 41.7 (11.7) 75.5 78.4 11.3 (10.0) 6.3 (1.7) 5.2 (2.2) 18 (22)</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hu</td>
<td>2012</td>
<td>ADA 40 mg QOW</td>
<td>26 12 28.2 (6.9) 92.3 96.2 7.4 (5.7) 5.9 (1.4) 3.7 (2.1) 25 (23)</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huang</td>
<td>2014</td>
<td>ADA 40 mg QOW</td>
<td>229 12 30.1 (8.7) 80.8 95.6 8.1 (6.0) 6.0 (1.4) 4.3 (2.3) 22 (24)</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van der Heijde</td>
<td>2005</td>
<td>IFX 5 mg/kg</td>
<td>201 24 40.0 (10.9) 81.8 81.8 12.1 (8.7) 6.5 (1.6) 5.6 (2.2) 23 (24)</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis</td>
<td>2003</td>
<td>GOL 50 mg Q4W</td>
<td>138 24 38.0 (12.6) 73.9 81.8 11.0 (9.6) 6.5 (1.6) 5.0 (2.4) 18 (18)</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calin</td>
<td>2004</td>
<td>ETN 50 mg QW</td>
<td>25 24 31.9 (9.2) 76.0 91.0 16.4 (8.3) 6.5 (1.2) 5.4 (1.8) 24 (21)</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van der Heijde</td>
<td>2005</td>
<td>IFX 5 mg/kg</td>
<td>28 12 41.0 (11.1) 78.6 88.6 13.8 (10.3) 6.6 (2.1) 6.5 (1.8) 23 (13)</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis</td>
<td>2003</td>
<td>GOL 50 mg QW</td>
<td>115 24 29.6 (7.5) 82.6 94.8 7.7 (4.7) 6.2 (1.4) 4.4 (2.3) 23 (30)</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giardina</td>
<td>2010</td>
<td>IFX 5 mg/kg</td>
<td>25 24 32.6 (6.8) 80.0 96.0 15.7 (6.5) 6.6 (1.1) 5.6 (1.5) 23 (11)</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huang</td>
<td>2014</td>
<td>CZP 400 mg QW</td>
<td>65 24 41.0 (10.8) 72.3 81.5 8.8 (5.4) 6.5 (1.7) 5.6 (2.3) 23 (30)</td>
<td>NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tam</td>
<td>2014</td>
<td>GOL 50 mg Qmonth</td>
<td>20 24 35.6 (9.9) 90.0 88.5 13.0 (10.5) 6.6 (1.5) 5.2 (2.6) 18 (21)</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis</td>
<td>2003</td>
<td>IFX 5 mg/kg</td>
<td>25 24 36.8 (8.0) 80.0 96.0 15.4 (10.6) 6.5 (1.2) 6.1 (0.9) 25 (12)</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis</td>
<td>2003</td>
<td>IFX 5 mg/kg</td>
<td>25 24 36.8 (8.0) 80.0 96.0 15.4 (10.6) 6.5 (1.2) 6.1 (0.9) 25 (12)</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis</td>
<td>2003</td>
<td>IFX 5 mg/kg</td>
<td>25 24 36.8 (8.0) 80.0 96.0 15.4 (10.6) 6.5 (1.2) 6.1 (0.9) 25 (12)</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis</td>
<td>2003</td>
<td>IFX 5 mg/kg</td>
<td>25 24 36.8 (8.0) 80.0 96.0 15.4 (10.6) 6.5 (1.2) 6.1 (0.9) 25 (12)</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis</td>
<td>2003</td>
<td>IFX 5 mg/kg</td>
<td>25 24 36.8 (8.0) 80.0 96.0 15.4 (10.6) 6.5 (1.2) 6.1 (0.9) 25 (12)</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis</td>
<td>2003</td>
<td>IFX 5 mg/kg</td>
<td>25 24 36.8 (8.0) 80.0 96.0 15.4 (10.6) 6.5 (1.2) 6.1 (0.9) 25 (12)</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis</td>
<td>2003</td>
<td>IFX 5 mg/kg</td>
<td>25 24 36.8 (8.0) 80.0 96.0 15.4 (10.6) 6.5 (1.2) 6.1 (0.9) 25 (12)</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis</td>
<td>2003</td>
<td>IFX 5 mg/kg</td>
<td>25 24 36.8 (8.0) 80.0 96.0 15.4 (10.6) 6.5 (1.2) 6.1 (0.9) 25 (12)</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis</td>
<td>2003</td>
<td>IFX 5 mg/kg</td>
<td>25 24 36.8 (8.0) 80.0 96.0 15.4 (10.6) 6.5 (1.2) 6.1 (0.9) 25 (12)</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASFI: Bath Ankylosing Spondylitis Functional Index; CRP: C-reactive protein; ADA: adalimumab; CZP: certolizumab; ETN: etanercept; GOL: golimumab; IFX: infliximab; IFX-dyrb: infliximab-dyrb (IFX biosimilar); MTX: methotrexate; DMARD: disease-modifying antirheumatic drugs; PBO: placebo; N/A: not available; NR: not reported; QOW: every other week; QW: once per week; Q2W: every 2 weeks; Q4W: every 4 weeks; Q6W: every 6 weeks; Qmonth: once per month.
CRP results were not available for IFX-dyyb at 12 weeks. I² values were 2.25, 4.53, and 7.46 for the BASDAI, BASFI, and CRP models, respectively, indicating low heterogeneity.

Eleven trials (including 24 arms, 2083 participants) were included in the analysis of relative efficacy at 24 weeks. The network of each comparison is presented in Supplementary Figure 3A–C (available with the online version of this article). The fit of the 24-week models was also good (Supplementary Table 2). All TNFi were significantly more efficacious than placebo in reducing BASDAI and BASFI scores (relative effect sizes for BASDAI range from −3.04 to −1.48 MD, for BASFI from −1.96 to −1.23 MD), and all TNFi except CZP and IFX-dyyb were superior to placebo in decreasing CRP (relative effect size from −1.30 to −0.69 MD; Figure 3). I² values were 7.39, 2.64, and 2.35 for the BASDAI, BASFI, and CRP models, respectively, again indicating low heterogeneity.

There was no evidence of substantial drift in placebo responses over calendar time (data not shown).

**Comparisons among TNFi at 12 weeks.** We conducted 2 analyses for paired comparison between TNFi at 12 weeks: 1

---

**Figure 1.** Network of treatment comparison at 12 weeks for (A) BASDAI, (B) BASFI, and (C) CRP. The size of the node corresponds to the number of total trial participants. Direct comparisons are linked with a line; the line thickness is proportional to the number of trials that assessed the comparison. BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASFI: Bath Ankylosing Spondylitis Functional Index; CRP: C-reactive protein.
Figure 2. Relative effect sizes compared to placebo at 12 weeks for BASDAI, BASFI, and CRP. Relative efficacies are represented as MD with 95% CrI, and are listed in the right column. Black lines represent the analysis with an open-label trial; grey lines represent the analysis without the open-label trial. A negative value means greater reduction compared with placebo.

BASDAI:
- Adalimumab: Better -1.55 (-2.15, -0.95) Worse -1.54 (-1.92, -1.15)
- Certolizumab: Better -1.45 (-2.32, -0.59) Worse -1.45 (-2.03, -0.86)
- Etanercept: Better -1.51 (-1.96, -1.04) Worse -1.76 (-2.04, -1.44)
- Golimumab: Better -1.55 (-2.21, -0.92) Worse -1.47 (-1.89, -1.13)
- Infliximab: Better -2.66 (-3.32, -1.89) Worse -1.95 (-2.61, -1.28)
- Infliximab_dyyb: Better -2.67 (-3.87, -1.31) Worse -1.94 (-2.82, -1.09)

BASFI:
- Adalimumab: Better -1.44 (-1.85, -0.96) Worse -1.46 (-1.81, -1.06)
- Certolizumab: Better -1.05 (-1.73, -0.38) Worse -1.05 (-1.65, -0.46)
- Etanercept: Better -1.43 (-1.76, -1.11) Worse -1.54 (-1.83, -1.25)
- Golimumab: Better -1.57 (-2.05, -1.07) Worse -1.57 (-1.99, -1.14)
- Infliximab: Better -1.99 (-2.47, -1.42) Worse -1.53 (-2.11, -0.95)
- Infliximab_dyyb: Better -1.81 (-2.68, -0.78) Worse -1.34 (-2.22, -0.44)

CRP:
- Adalimumab: Better -1.19 (-1.76, -0.57) Worse -1.18 (-1.76, -0.56)
- Certolizumab: Better -0.71 (-1.62, 0.18) Worse -0.71 (-1.60, 0.18)
- Etanercept: Better -1.35 (-1.83, -0.86) Worse -1.35 (-1.83, -0.86)
- Golimumab: Better -1.19 (-1.83, -0.56) Worse -1.19 (-1.83, -0.56)
- Infliximab: Better -1.57 (-2.33, -0.86) Worse -1.57 (-2.32, -0.87)

95% credible intervals
Figure 3. Relative effect sizes compared to placebo at 24 weeks for BASDAI, BASFI, and CRP. Relative efficacies are represented as MD with 95% CrI, and are listed in the right column. Black lines represent the analysis with an open-label trial; grey lines represent the analysis without the open-label trial. A negative value means greater reduction compared with placebo. BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASFI: Bath Ankylosing Spondylitis Functional Index; CrI: credible interval; CRP: C-reactive protein; MD: mean difference.
that included a single open-label study, and a second analysis that excluded this single open-label study. In the analysis that included the open-label study, IFX was significantly more efficacious in reducing BASDAI than ADA (relative effect size \(-1.1\) MD 95% CrI \(-2.0\) to \(-0.01\)), CZP (relative effect size \(-1.2\) MD 95% CrI \(-2.3\) to \(-0.02\)), ETN (relative effect size \(-1.2\) MD 95% CrI \(-1.8\) to \(-0.4\)), and GOL (relative effect size \(-1.1\) MD 95% CrI \(-2.0\) to \(-0.1\); Table 2A, below diagonal). IFX was also significantly better in reducing BASFI than CZP (relative effect size \(-1.0\) MD 95% CrI \(-1.7\) to \(-0.03\); Table 2B, below diagonal). However, there were no significant differences among TNFi in the paired comparison of changes in CRP at 12 weeks (Table 2C, below diagonal). Biosimilar IFX-dyyb had MD similar to that of IFX in reducing BASDAI, consistent with the result of the head-to-head trial between the 2 drugs. However, it had wider CrI, likely because it was assessed in only 1 trial.

In the analysis that excluded the open-label trial, IFX was not more efficacious than other TNFi in decreasing BASDAI (Table 2A–C, above diagonals).

In the metaregression model, when adjusted for baseline BASDAI and baseline CRP, IFX remained superior to CZP, ADA, and ETN in reducing BASDAI (Supplementary Table 3, available with the online version of this article). In addition, IFX-dyyb was significantly more efficacious than ETN in reducing BASDAI. When adjusted for baseline BASFI and baseline CRP, IFX was superior to CZP and ETN in BASFI reduction. No significant difference was detected in CRP changes.

**Comparisons among TNFi at 24 weeks.** The advantage of IFX seen at 12 weeks was not present in the 24-week analysis. At 24 weeks, no TNFi was significantly more efficacious than others in reducing BASDAI, BASFI, or CRP (Tables 3A–C). IFX-dyyb had numerically a higher reduction in BASFI and baseline CRP, IFX was superior to CZP and ETN in reducing BASDAI (Supplementary Table 3, below diagonals).

### Table 2A. Comparative efficacy of BASDAI changes at 12 weeks. The relative effect size is presented as mean difference with 95% CrI.

<table>
<thead>
<tr>
<th>Drug</th>
<th>ADA</th>
<th>CZP</th>
<th>ETN</th>
<th>GOL</th>
<th>IFX</th>
<th>IFX-dyyb</th>
<th>PBO</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADA</td>
<td>–</td>
<td>(-0.09)</td>
<td>(-0.3)</td>
<td>(-0.08)</td>
<td>(0.4)</td>
<td>(-1.5)</td>
<td>NA</td>
</tr>
<tr>
<td>CZP</td>
<td>(0.10)</td>
<td>(-0.3)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.4)</td>
<td>(-1.5)</td>
<td>(-1.5)</td>
</tr>
<tr>
<td>ETN</td>
<td>(0.04)</td>
<td>(-0.3)</td>
<td>(0.03)</td>
<td>(-0.3)</td>
<td>(0.5)</td>
<td>(-1.1)</td>
<td>(-1.4)</td>
</tr>
<tr>
<td>GOL</td>
<td>(0.0)</td>
<td>(-0.3)</td>
<td>(0.03)</td>
<td>(0.5)</td>
<td>(0.2)</td>
<td>(-1.2)</td>
<td>(-1.5)</td>
</tr>
<tr>
<td>IFX</td>
<td>(-1.1)</td>
<td>(-1.2)</td>
<td>(-0.1)</td>
<td>(-0.3)</td>
<td>(0.01)</td>
<td>(-1.9)</td>
<td>(-1.9)</td>
</tr>
<tr>
<td>IFX-dyyb</td>
<td>(-1.1)</td>
<td>(-1.2)</td>
<td>(-0.1)</td>
<td>(-0.3)</td>
<td>(0.01)</td>
<td>(-2.8)</td>
<td>(-1.1)</td>
</tr>
<tr>
<td>PBO</td>
<td>(1.5)</td>
<td>(1.5)</td>
<td>(2.7)</td>
<td>(2.7)</td>
<td>(2.7)</td>
<td>(2.7)</td>
<td>(2.7)</td>
</tr>
</tbody>
</table>

### Table 2B. Comparative efficacy of BASFI changes at 12 weeks. The relative effect size is presented as mean difference with 95% CrI.

<table>
<thead>
<tr>
<th>Drug</th>
<th>ADA</th>
<th>CZP</th>
<th>ETN</th>
<th>GOL</th>
<th>IFX</th>
<th>IFX-dyyb</th>
<th>PBO</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADA</td>
<td>–</td>
<td>(-0.4)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>NA</td>
</tr>
<tr>
<td>CZP</td>
<td>(0.4)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(-1.6)</td>
</tr>
<tr>
<td>ETN</td>
<td>(0.0)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(-1.4)</td>
</tr>
<tr>
<td>GOL</td>
<td>(0.1)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(-1.0)</td>
</tr>
<tr>
<td>IFX</td>
<td>(-0.6)</td>
<td>(-1.0)</td>
<td>(-0.4)</td>
<td>(-0.4)</td>
<td>(-0.4)</td>
<td>(-0.4)</td>
<td>(-1.0)</td>
</tr>
<tr>
<td>IFX-dyyb</td>
<td>(-0.4)</td>
<td>(-0.3)</td>
<td>(-0.2)</td>
<td>(-0.2)</td>
<td>(-0.2)</td>
<td>(-0.2)</td>
<td>(-1.3)</td>
</tr>
<tr>
<td>PBO</td>
<td>(1.4)</td>
<td>(1.1)</td>
<td>(1.6)</td>
<td>(2.0)</td>
<td>(1.8)</td>
<td>(1.8)</td>
<td>(-0.6)</td>
</tr>
</tbody>
</table>

### Table 2C. Comparative efficacy of CRP changes at 12 weeks. The relative effect size is presented as mean difference with 95% CrI.

<table>
<thead>
<tr>
<th>Drug</th>
<th>ADA</th>
<th>CZP</th>
<th>ETN</th>
<th>GOL</th>
<th>IFX</th>
<th>IFX-dyyb</th>
<th>PBO</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADA</td>
<td>–</td>
<td>(-0.45)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>NA</td>
<td>(-1.2)</td>
</tr>
<tr>
<td>CZP</td>
<td>(0.5)</td>
<td>(0.6)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td>NA</td>
<td>(-0.7)</td>
</tr>
<tr>
<td>ETN</td>
<td>(-0.2)</td>
<td>(0.6)</td>
<td>(0.2)</td>
<td>(-0.2)</td>
<td>(0.2)</td>
<td>NA</td>
<td>(-1.3)</td>
</tr>
<tr>
<td>GOL</td>
<td>(-0.0)</td>
<td>(0.2)</td>
<td>(0.2)</td>
<td>(0.2)</td>
<td>(0.2)</td>
<td>NA</td>
<td>(-1.2)</td>
</tr>
<tr>
<td>IFX</td>
<td>(-0.4)</td>
<td>(-0.9)</td>
<td>(-0.9)</td>
<td>(-0.9)</td>
<td>(0.0)</td>
<td>NA</td>
<td>(-1.6)</td>
</tr>
<tr>
<td>IFX-dyyb</td>
<td>(-0.4)</td>
<td>(-0.9)</td>
<td>(-0.9)</td>
<td>(-0.9)</td>
<td>(0.0)</td>
<td>NA</td>
<td>(-1.6)</td>
</tr>
<tr>
<td>PBO</td>
<td>(1.2)</td>
<td>(0.7)</td>
<td>(1.3)</td>
<td>(1.2)</td>
<td>(1.2)</td>
<td>NA</td>
<td>(-1.0)</td>
</tr>
</tbody>
</table>

Values below the diagonal represent the analysis with 1 open-label trial. Values above the diagonal represent the analysis without the open-label trial. Each cell represents a paired comparison. The columns represent the reference medication for each comparison, and the rows represent the comparators. A negative value means greater improvement by the comparator, indicating the comparator is more efficacious than the reference drug. A positive value means less improvement by the comparator, indicating the reference drug is more efficacious than the comparator. Cells in bold are statistically significant comparisons at p < 0.05. BASDAI: Bath Ankylosing Spondylitis Disease Activity Score; BASFI: Bath Ankylosing Spondylitis Functional Index; CrI: credible interval; CRP: C-reactive protein; ADA: adalimumab; CZP: certolizumab; ETN: etanercept; GOL: golimumab; IFX: infliximab; IFX-dyyb: infliximab-dyyb (IFX biosimilar); PBO: placebo.
in BASDAI and BASFI compared to other TNFi, and ADA had a numerically higher reduction in CRP compared to other TNFi. However, these comparisons were not statistically significant.

In the metaregression model, when adjusted for baseline BASDAI and CRP, IFX-dyyb remained numerically better than other TNFi in reducing BASDAI (Supplementary Table 4, available with the online version of this article). When adjusted for baseline BASFI and baseline CRP, ADA was numerically better than other TNFi in BASFI and CRP reduction. None of these differences were statistically significant.

**DISCUSSION**

In our systematic review and Bayesian NMA, we compared the relative efficacy of 6 different TNFi in the treatment of active AS, using BASDAI, BASFI, and CRP as outcome measures. We found that at 12 weeks, IFX was superior to ADA, CZP, ETN, and GOL in reducing BASDAI, and superior to CZP in reducing BASFI. These differences persisted in analyses that adjusted for baseline values of BASDAI, BASFI, and CRP, indicating that differences among trials in the activity of AS at enrollment did not account for this association. Qualitatively similar results were present for IFX-dyyb. We did not find differences among TNFi other than IFX in these outcomes, and found no differences among TNFi in reducing CRP levels. Responses were not different among TNFi in the 24-week analysis. We did not find differences among TNFi when we excluded an open-label trial.

The apparent earlier response to IFX and IFX-dyyb than to other TNFi may relate to the use of loading doses or to their intravenous method of administration, both of which are
unique to these 2 medications. However, the advantage of IFX at 12 weeks in BASDAI responses should be interpreted cautiously. A statistically significant difference does not necessarily translate into a clinically important difference. Further, given that AS is a chronic condition, the early symptom improvement may not be viewed as important as intermediate or longterm effects. Additionally, this early advantage was not evident in the analyses that excluded an open-label study, indicating sensitivity of this association to the results of Giardina, et al, which directly compared IFX and ETN. This sensitivity reflects the influence of trials with direct head-to-head comparisons in NMA. However, these results may assist in the choice of TNFi in clinical situations when prompt symptom responses are needed, although these situations are rare in AS. A report on comparisons of biological agents in treatment of severely active ulcerative colitis also concluded that IFX was more effective than ADA in induction therapy.

We attempted to address many potential biases. First, we chose continuous outcome measures rather than dichotomous outcomes, to maximize the potential to differentiate effects among medications. Second, in the process of data extraction, we contacted principal investigators and/or study sponsors for additional data, increasing the completeness of the dataset and decreasing the risk of bias as a result of incomplete reporting of data. Third, to address the heterogeneity among the studies, we tested several factors that could potentially influence the relative effects of different TNFi, including examination of changes in placebo responses over time, and estimation of associations with metaregression using baseline disease activity measures as covariates. In the analysis that excluded the open-label trial, IFX was numerically (but not statistically) more efficacious than other TNFi in reducing BASDAI, indicating that this aspect of study design may influence treatment effects on patient-reported outcomes. Given its low use in these trials, concomitant MTX was unlikely to influence the comparisons among TNFi.

The major limitation of our study is the small number of head-to-head trials. Two such trials were identified, and we could construct only 1 closed loop in the evidence network with both direct and indirect comparisons of 2 drugs available (IFX vs ETN). The other TNFi were compared either to placebo directly or through another TNFi, forming a star network. Because of this, we had to assume consistency in the analysis, which reduces confidence in the estimation.

Our study showed that IFX was somewhat more efficacious in reducing the BASDAI than several other TNFi in the short term, but this advantage was sensitive to the inclusion of an open-label trial and diminished at 24 weeks. IFX (or IFX-dyyb, which had similar effects) may therefore be conditionally preferred in the uncommon case in which a prompt symptom response is needed. The choice of TNFi in patients with AS may also be guided by the presence of specific comorbid conditions, such as inflammatory bowel disease or recurrent iritis. Without these considerations, more information on the relative safety and longterm effectiveness of TNFi will provide critical guidance on the choice of TNFi in the treatment of AS in clinical practice.

ACKNOWLEDGMENT
The authors thank study investigators, Amgen, Pfizer, and UCB for providing additional data upon request.

ONLINE SUPPLEMENT
Supplementary material accompanies the online version of this article.

REFERENCES
decision making 2: a generalized linear modeling framework for
pairwise and network meta-analysis of randomized controlled trials.
15. Higgins JPT, Thompson SG. Quantifying heterogeneity in a
16. R core team. R: A Language and Environment for Statistical
Computing. [Internet. Accessed November 16, 2017.] Available from:
www.R-project.org
17. van Valkenhoef G, Kuiper J, van Valkenhoef MG. Package ‘gemtc.’
[Internet. Accessed November 16, 2017.] Available from:
cran.r-project.org/web/packages/gemtc/gemtc.pdf
18. Plummer M. JAGS. A program for analysis of Bayesian graphical
models using Gibbs sampling. [Internet. Accessed November 16,
19. Salanti G, Kavoura FK, Ioannidis JP. Exploring the geometry of
20. van der Heijde D, Kivitz A, Schiff MH, Sieper J, Dijkmans BA,
Braun J, et al. Efficacy and safety of adalimumab in patients with
ankylosing spondylitis: results of a multicenter, randomized,
21. Maksymowych WP, Rahman P, Shojarina K, Olszynski WP,
Thomson GT, Ballal S, et al. Beneficial effects of adalimumab on
biomarkers reflecting structural damage in patients with ankylosing
significantly reduces inflammation and serum DKK-1 level but
increases fatty deposition in lumbar spine in active ankylosing
of adalimumab in Chinese adults with active ankylosing spondylitis:
results of a randomised, controlled trial. Ann Rheum Dis 2014;
73:587–94.
24. Landewe R, Braun J, Deodhar A, Dougdas M, Maksynowych WP,
Mease PJ, et al. Efficacy of certolizumab pegol on signs and
symptoms of axial spondyloarthritis including ankylosing
spondylitis: 24-week results of a double-blind randomised
25. Gorman JD, Sack KE, Davis JC Jr. Treatment of ankylosing
spondylitis by inhibition of tumor necrosis factor alpha. N Engl J
26. Davis JC Jr, Van Der Heijde D, Braun J, Dougados M, Cushman
Cleugh DO, et al. Recombinant human tumor necrosis factor receptor
(etanercept) for treating ankylosing spondylitis: a randomized,
27. Calin A, Dijkmans BA, Emery P, Hakala M, Kalden J,
Leirisalo-Repo M, et al. Outcomes of a multicentre randomised
clinical trial of etanercept to treat ankylosing spondylitis. Ann
28. van der Heijde D, Da Silva JC, Dougdas M, Geher P, van der
Horst-Bruinsma I, Juanaola X, et al. Etanercept 50 mg once weekly
is as effective as 25 mg twice weekly in patients with ankylosing
29. Barkham N, Coates LC, Keen H, Hensor E, Fraser A, Redmond A,
et al. Double-blind placebo-controlled trial of etanercept in the
prevention of work disability in ankylosing spondylitis. Ann Rheum
Efficacy of etanercept on rheumatic signs and pulmonary function
tests in advanced ankylosing spondylitis: results of a randomised
double-blind placebo-controlled study (SPINE). Ann Rheum Dis
2011;70:799–804.
al. Efficacy and safety of golimumab in patients with ankylosing
spondylitis: results of a randomized, double-blind,
placebo-controlled, phase III trial. Arthritis Rheum 2008;
58:3402–12.
effects of golimumab on subclinical atherosclerosis and arterial
stiffness in ankylosing spondylitis—a randomized,
efficacy of golimumab in Chinese patients with active ankylosing
spondylitis: 1-year results of a multicentre, randomized,
double-blind, placebo-controlled phase III trial. Rheumatology
Treatment of active ankylosing spondylitis with infliximab: a
35. Marzo-Ortega H, McGonagle D, Jarrett S, Haugeberg G, Hensor E,
O’Connor P, et al. Infliximab in combination with methotrexate in
active ankylosing spondylitis: a clinical and imaging study. Ann
Rheum Dis 2005;64:1568–75.
36. van der Heijde D, Dijkmans B, Geusens P, Sieper J, DeWoody K,
Williamson P, et al. Efficacy and safety of infliximab in patients
with ankylosing spondylitis: results of a randomized,
placebo-controlled trial (ASSERT). Arthritis Rheum 2005;
52:582–91.
37. Inman RD, Maksymowych WP; CANDLE Study Group. A
double-blind, placebo-controlled trial of low dose infliximab in