Longer term benefits of treating rheumatoid arthritis: Assessment of radiographic damage and physical function in clinical trials

V. Strand

ABSTRACT

Evidence from randomized controlled trials with newly approved DMARD therapies have proven efficacy by American College of Rheumatology [ACR] response criteria, Disease Activity Scores [DAS] and radiographic measures of disease progression. Treatment over 2 years duration results in clinically meaningful improvements in physical function, by the Health Assessment Questionnaire [HAQ] and health related quality of life [HRQOL]. Using the medical outcomes short form 36 [SF-36]. Changes in HAQ are evident within one month, maximal at 3-6 months, and sustained over 24 months, reflected by improvements in social functioning, role emotional and the general health profile as well as physical domains of SF-36. Trials with the new DMARDs, as well as MTX, indicate that long term benefits in radiographic damage and physical function can be inferred from treatment data over 12 months. "Successful" patients who continue to do well will derive benefit from treatment for as long as 3 to 5 years. Recent RCTs indicate that combination therapy, initiated together, offers more improvement in radiographic progression and physical function than monotherapy, although the trial data cannot yet tell us which DMARD may be most beneficial in a given patient. Once treatment with any one agent – biologic or synthetic – is initiated, the addition of a second agent should occur rapidly, if for example active disease persists 8 or 12 weeks later, without waiting for documented treatment failure. If this treatment paradigm is followed, along with regular assessments of radiographic damage and physical function, then patients may more likely derive long term clinical benefit than with traditional approaches.

Introduction

Randomized controlled trials (RCTs) with recently approved therapies in rheumatoid arthritis (RA) have proven efficacy by American College of Rheumatology (ACR) response criteria, Disease Activity Scores (DAS) and radiographic measures of disease progression. Importantly, 2 year data were required by the US Food and Drug Administration for demonstration of longer term clinical benefit, by patient reported measures of physical function and health related quality of life (HRQOL).

Based on these requirements, RCTs with the newly approved DMARD therapies have demonstrated clinically meaningful improvements even in patients with long-standing disease who had Health Assessment Questionnaire (HAQ) scores reflecting large and potentially irreversible impairments in physical function. Together, disease relevant and generic self report measures of physical function and HRQOL reflect an individual patient's ability to engage in usual activities of daily living, including 'shopping', walking as much as a mile or engaging in sports. Although it was previously widely believed that progressive loss of physical function was inevitable in RA, and the best effect of treatment would be to 'stabilize' physical function, it has subsequently been shown in clinical trials that improvements in HAQ scores are evident within 1 month, maximal by 6 months and sustained thereafter for as long as 24 months' treatment. These improvements are reflected in clinically meaningful changes in the mental as well as physical domains of the HRQOL.

Even if statistically significant compared with placebo, mean or median changes from baseline in a treatment group are not necessarily clinically meaning-
ful or readily applicable to a clinical practice setting. The "minimum clinically important difference" (MCID), the amount of improvement perceptible to a patient and considered clinically meaningful, has been proposed in order to better understand the clinical applicability of data from RCTs. Improvements of 33% to 36% over baseline (or 18% greater than placebo) are thought to represent MCID (1, 2). Statistical analyses of RCT data linked degrees of improvement in global assessments of disease activity to reported changes in physical function, defining MCID as –0.22 for the HAQ disability index in patients receiving parenteral therapy, even with ‘failed’ background therapy, generally discontinue participation early, or require rescue therapy, thus ‘depleting’ the comparison group. The validity of intention to treat, last value carried forward (ITT, LOCF) analyses in these settings is questionable, as they cannot reflect true treatment comparisons.

Dropout rates have differed across trials, based on the selection of active as well as comparator treatments. Despite separate blinded assessors of efficacy and safety, rapid clinical improvements in patients receiving parenteral therapy with biologic agents have generally resulted in differential early dropout rates in patients receiving control (synthetic) treatment, with higher retention rates in those receiving biologic therapies. Based on ethical issues, data monitoring and institutional review boards have required unblinding in several RCTs before all patients completed 24 months of treatment (6, 7). Regardless, not all patients who continue protocol participation are ‘responders’ by whatever criteria selected, as ACR20% and DAS good and moderate response rates don’t generally exceed 75-85%.

In general, lower dropout rates reflect patient expectations as well as treatment benefit. Table I compares completion rates in placebo, placebo + background therapy and active controlled trials over 12 to 24 months participation. Completion rates are higher when placebo treatment is superimposed on failed background therapy; higher still when all patients know they are receiving active therapy, be it synthetic or biologic, as demonstrated in the recent Trial of Etanercept and Methotrexate with Radiographic Patient Outcomes (TEMPO) and the Active controlled Study of Patients receiving Infliximab for RA of Early onset (ASPIRE) trials (8-11). Of interest, across RCTs, 24 month completion rates in methotrexate (MTX) treatment groups range from 42% (US301, placebo controlled), to 52% (TEMPO), 56% (MN302/4) and 59% [Etanercept in Early RA (ERA)]. In comparison, 24 month completion rates with biologic therapies range from 55-68% [Anti TNF Trial in RA with Concomitant Therapy (ATRACT) trial with background MTX, placebo controlled] to 71% (TEMPO) and 74% (ERA) (5-9).

In view of these limitations, long term data collected in the context of RCTs are best examined to determine whether initial benefit, evident at 12 months, is ‘sustained over 24 months in those continuing treatment.’ As it is no longer considered ethical to continue blinded treatment after 6 or 12 months unless all patients have been offered rescue therapy, analyses are best restricted to intra-group rather than across group comparisons (12). Sustained benefit over 3 to 5 years’ treatment has been reported with all the new DMARDs, in patients originally enrolled in RCTs who continued active therapy thereafter. Short term trials can reflect long term treatment effects but data are obviously limited, as these represent ‘successful’ cohorts of patients ‘preselected’ for tolerability and response to the therapy in question (13). Long-term treatment registries can also offer valuable information, but are based on the more limited collection of data, and may be biased according to the pre-selection of patient populations.

**Long term improvements in progression of radiographic damage in randomized controlled trials**

Leflunomide, etanercept, infliximab and adalimumab have been shown to inhibit radiographic progression of erosions and joint space narrowing (JSN), and anakinra to delay structural damage in patients with active RA (14, 15). Mean changes in total Sharp scores over 6 and 12 months with active treatment in these RCTs were within a limited range, from -0.71 to 2.24 points, significantly less than placebo and numerically less than the estimated anticipated yearly progression rates. Twelve months of treatment with methotrexate, leflunomide, and etanercept, as well as...
infliximab and adalimumab, with background methotrexate therapy, resulted in median change scores of 0.00. The extrapolation of radiographic data from clinical trials to clinical practice is complex. Without head to head comparisons, it is difficult to compare data across different protocol populations: patients with early versus established disease, aggressive versus non-progressive disease, or those receiving mono-therapy versus combination with failed DMARD treatment. Baseline radiographic damage appears to be the best predictor of future progression, and in individual patients and patient populations may significantly influence the response to an effective therapy (16-18). In contrast, patients who do not develop radiographic damage in the first several years of active disease are less likely to do so later. As summarized by Scott et al., 6 prospective series have reported that 39-73% of patients evaluated within their first year of disease subsequently develop erosions (19).

As recently reviewed, changes in total Sharp scores indicate that the beneficial effects of treatment on progression of erosions are more pronounced during the second 6 and 12 months of therapy, whereas the slope of the line reflecting changes in JSN scores remains unchanged from 0-6 to 12 and 24 months. These are reflected in graphic displays of changes in total Sharp scores over 6, 12 and 24 months in leflunomide, ERA, ATTRACTION, DE019, ASPIRE and TEMPO trials (20). It is unclear whether earlier effects on JSN are evident prior to erosions, or whether these observations reflect the limitations of radiographic assessment. Additional studies in patients with early disease, utilizing ultrasound and magnetic resonance imaging should help to clarify this matter.

As each of the patient populations differ in baseline disease characteristics and radiographic damage, another way to compare inhibition of disease progression is to compare the percentage of patients with negative or 0 changes in total Sharp scores. Because radiographic damage progresses both by erosions and JSN, and radiographs are read in blinded random order, scores ≤ 0 do not necessarily imply healing of erosions or retention of cartilage, but do indicate the absence of additional damage. Table II presents the percentage of patients without radiographic progression over 2 years treatment with the new DMARDs, as absolute change scores ≤ 0; or < 0.5 (for etanercept RCTs) (14, 21, 8, 9).
Assessment of radiographic damage and physical function in clinical trials / V. Strand

Table II. Percentage of patients with negative or no change in total (composite); scores at 12 and 24 months.

<table>
<thead>
<tr>
<th>Patients with “no progression” at 12 and 24 mos.</th>
<th>US301*</th>
<th>ATTRACTION *</th>
<th>MN 302/4*</th>
<th>ERA*</th>
<th>DEXA9</th>
<th>TEMPO*</th>
<th>ASPIRE*</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNIF MTX</td>
<td>INF + MTX</td>
<td>LNIF MTX</td>
<td>MTX ETN</td>
<td>ADA</td>
<td>MTX</td>
<td>ETN + MTX</td>
<td>INF3* MTX</td>
</tr>
<tr>
<td>3 mglkg q8wk q4wk</td>
<td>10 mglkg q8wk q4wk</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patients evaluated [n]</th>
<th>71</th>
<th>66</th>
<th>71</th>
<th>71</th>
<th>77</th>
<th>66</th>
<th>149</th>
<th>169</th>
<th>132</th>
<th>122</th>
<th>153</th>
<th>212</th>
<th>212</th>
<th>218</th>
<th>228</th>
<th>307</th>
<th>303</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Baseline scores</td>
<td>23</td>
<td>23</td>
<td>79</td>
<td>71</td>
<td>67</td>
<td>76</td>
<td>25</td>
<td>25</td>
<td>12</td>
<td>13</td>
<td>68-70</td>
<td>22</td>
<td>27</td>
<td>22</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Baseline scores = 0 [%]</td>
<td>17%</td>
<td>24%</td>
<td>0</td>
<td>0</td>
<td>NR</td>
<td>0</td>
<td>10%</td>
<td>15%</td>
<td>13%</td>
<td>16%</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

| Patients + change ≤ 0 [%] at 12 mos [%] | 61% | 73% | 48% | 49% | 48% | 67% | 39% | 53% | 56% | 62% | 54% | 57% | 68% | 80% | 45% | 58% | 59% |
| Patients + change ≤ 0 [%] at 24 mos [%] | 58% | 59% | 43% | 47% | 41% | 54% | 32% | 44% | 51% | 63% | 67% | NR | NR | NR | NR | NR | NR |

*a” no change” defined as < 0.5

study end point were available, were evaluated by Van der Heijde et al. (22). Overall, 128 of the original 824 patients were included; mean disease duration was 5.1 years, and the mean leflunomide treatment duration was 4.3 years. The mean change from baseline in the total score was 8.6 (of 440 by total Sharp/van der Heijde scores) with a yearly adjusted rate of 0.5, compared with 7.9 and 4.9, respectively, before leflunomide treatment duration was 4.3 years, and the mean leflunomide treatment duration was 4.3 years. The mean change from baseline in the total score was 8.6 (of 440 by total Sharp/van der Heijde scores) with a yearly adjusted rate of 0.5, compared with 7.9 and 4.9, respectively, before leflunomide treatment. Radiographic progression rates improved in 92/128 (72%) and deteriorated in 21/128 (16%) patients. In 42 (33%) patients with total scores > 0 at baseline, no increase in total Sharp scores were evident after leflunomide treatment. Genovese et al. briefly reported sustained radiographic benefit in patients continuing treatment with 25 mg etanercept in the ERA trial over 3 years; those who increased from 10 to 25 mg or who switched from MTX or added etanercept to MTX showed additional radiographic benefit (23).

Long term improvements in physical function and HRQOL

Recent RCTs have primarily compared treatment with experimental agents to placebo when added to ‘failed’ background therapy with MTX or other DMARDs – in patients with long disease duration having failed multiple DMARDs (24). Mean and median improvements in HAQ scores well exceeded -0.22 or MCID (6, 15, 25-28). Reported improvements were maximal by 6 months and sustained over 12 to 24 months treatment, and compared favorably with changes observed following monotherapy with etanercept, leflunomide and MTX in patients with early and later disease, and adalimumab in those with longstanding RA (5, 15, 29-32). Prior to the introduction of the new DMARDs, longitudinal series reported progressive deterioration in physical function or, at best, stabilization with standard of care (including MTX). Summarizing data from 12 studies Scott et al. report average increases in HAQ scores of 0.033/year (19). They describe a ‘j-shaped curve’ plotting disability (by HAQ score) versus disease duration. It is well known that patients with early disease report more impairment in physical function, which is dramatically improved when the first DMARD is initiated (33). Therefore comparisons of reported improvements in HAQ scores across treatment groups should account for baseline disease characteristics in the protocol populations. Patient reported outcomes of pain, physical function and global disease activity best differentiate active from placebo therapy (34, 35). When comparing changes from baseline in HAQ scores at 6 and 12 months in recent RCTs, mean improvements in patients who received active treatment (34) ranged from -0.25 to -0.80, compared with placebo groups where the mean change scores do not meet or exceed an MCID of -0.22. However, as noted above, it is not possible to compare changes across trials and therapies, given differences in baseline scores and disease duration in each protocol population. Subtraction of changes reported with placebo from those with active treatment within each protocol offers a simple ‘correction’ for these differences, and demonstrates comparable improvements with the TNFα inhibitors as well as MTX and leflunomide. Requiring mean improvements in HAQ scores which exceed MCID can offer a reasonable way to detect whether a treatment is effective.

The percentage of patients reporting improvement meeting or exceeding MCID offers another way to compare data across protocols, although neither method accounts for change relative to baseline scores. The ASPIRE trial enrolled 1,051 patients with a mean disease duration of 0.6 months; 66-72% were DMARD naïve; baseline HAQ scores were 1.5 (10,11). Mean improvements in HAQ scores over 30 to 54 weeks of treatment ranged from -0.75 to -0.79 across treatments, with 76% of patients reporting improvements which met or exceeded MCID in both the infliximab+MTX groups compared with 65% receiving MTX alone. This compares with 74-84% of leflunomide-treated and 69-78% of MTX-treated patients over 2 years of treatment in US301 and MN302/4, where the disease duration and baseline HAQ scores were 6.7 and 3.2 years, and 1.2 and 1.5, respectively (5). The TEMPO trial enrolled 682 RA patients with 6.6 years disease duration who had failed a mean of 2.3 DMARDs; baseline HAQ scores were 1.7 to 1.8 (8). Mean improve-
ments from baseline in the HAQ scores at 12 and 24 months were -0.60 and -0.61 with MTX; -0.70 and -0.70 with etanercept, compared with -1.0 and -1.01 with etanercept + MTX. (9) Importantly, the improvement in HAQ scores over 12-24 months indicates more than an additive treatment effect when combination therapy is initiated simultaneously.

A generic measure of HRQOL, the Medical Outcomes Survey Short Form 36 (SF-36) was first incorporated in RA clinical trials as part of the US301 study comparing leflunomide with MTX, and was subsequently studied in ATTRACT, ERA and phase III adalimumab trials. All have demonstrated that treatment related changes in HAQ scores are closely reflected in improvements in HRQOL, social functioning, role emotional and the general health profile, as well as physical domains. Patients receiving treatment over 12 and 24 months in US301, ATTRACT and

---

**Figure 1.**

Panels A and B (scales 0-80): Baseline and median or mean changes in SF-36 domain scores are presented, over 24 months treatment in ATTRACT and US301 RCTs. Baseline scores are denoted by the lines; change scores at 24 months are added to baseline values to demonstrate final, post treatment values, which can be compared to data in age and gender matched normative populations.

**Panel A:** Improvements reported by patients receiving placebo + MTX, infliximab 3mg q 8 weeks + MTX, and 10 mg q 8 weeks + MTX are shown, representing minimum and maximum changes from baseline in active treatment groups receiving infliximab.

**Panel B:** Improvements reported by patients receiving leflunomide or MTX are shown.

Panels C and D (scales 0-40): Median or mean changes from baseline over 24 months treatment are depicted from ATTRACT and US301 RCTs. Values for MCID: improvements of 5-10 points in domain scores are indicated by dotted lines.

**Panel C:** Improvements reported by patients receiving infliximab 3mg q 8 weeks + MTX, and 10 mg q 8 weeks + MTX are shown, representing minimum and maximum changes from baseline in active treatment groups receiving infliximab.

**Panel D:** Improvements reported by patients receiving leflunomide or MTX are shown.

Despite differing baseline domain scores, active treatment over 24 months resulted in numerical improvements which met or exceeded MCID. In physical domains, changes with active treatment are most evident in those with lowest scores at baseline.
ERA reported changes in SF-36 domains and physical component summary scores which met or exceeded MCID (5, 6, 37). Figures 1 and 2 present baseline and change scores in SF-36 domains following treatment with adalimumab, etanercept, infliximab, leflunomide and MTX (5, 15, 32, 37, 38). Improvements show that many patients report HRQOL scores after treatment which approach or equal those in age and gender matched populations without arthritis (5, 36).

Whether patients had early or long disease duration, baseline physical component scores (PCS) were remarkably similar across trials: 23.9 – 25.8 in ATTRACT; 30.2 – 30.9 in US301; 28.0 – 29.2 in ERA; and 28.5 – 29.1 in protocol DE019. Mean and median improvements with active treatment over 12 and 24 months resulted in increases in PCS scores from more than 2 standard deviations (SDs) below to within 1-2 SDs of normative values of 50; the magnitude of changes well exceeded MCID (2.5 – 5.0 points) in all trials. These changes are reflected by reductions in the percentage of patients reporting common physical limitations, as reported after treatment with MTX.

**Figure 2.**

**Panels A and B (scales 0-80):** Baseline and mean changes in SF-36 domain scores are presented, over 12 months treatment in ERA and DE019 RCTs. Baseline scores are denoted by the lines; change scores at 12 months are added to baseline values to demonstrate final, post treatment values, which can be compared to data in age and gender matched normative populations.

**Panel A:** Improvements reported by patients receiving etanercept or MTX are shown.

**Panel B:** Improvements reported by patients receiving adalimumab 40 mg qw weekly + MTX vs placebo + MTX (with rescue) are shown.

**Panels C and D (scales 0-40):** Mean changes from baseline over 12 months treatment are depicted from ERA and DE019 RCTs. Values for MCID: improvements of 5-10 points in domain scores are indicated by dotted lines.

**Panel C:** Improvements reported by patients receiving etanercept 25 mg or MTX are shown.

**Panel D:** Improvements reported by patients receiving adalimumab 40 mg qw weekly + MTX or placebo + MTX (with rescue) are shown.

Despite differing baseline domain scores, active treatment over 12 months resulted in numerical improvements which met or exceeded MCID. In physical domains, changes with active treatment are most evident in those with lowest scores at baseline.
etanercept and leflunomide (Table III). The impressive changes in PCS scores alone do not reflect the full range of improvement in HRQOL which occurs when treatments positively impact physical function in patients with active RA. As with the HAQ disability index, reported improvements are maximal within 6 months and sustained over 12 to 24 months of continuing therapy. These findings confirm that both HAQ and SF-36 are sensitive to change and accurately reflect treatment associated improvements. Data reflected by SF-36 are 'nice but not necessary', yet offer a means to measure the impact of RA upon other aspects of HRQOL, and compare to other chronic diseases, as well as facilitating economic analyses (39).

Correlation between clinical improvements and radiographic changes

Once well controlled data documented radiographic inhibition of disease progression, relatively poor correlations between clinical and radiographic responses further confounded the clinical interpretation of data from these recent RCTs (15). As Scott et al. elegantly demonstrated in both metaanalyses of longitudinal studies, measures of radiographic damage and impairment in physical function are not closely correlated until patients have a disease duration of 8-15 years (19, 40). Weak correlations between ACR response rates and decreases in C-reactive protein (CRP) levels have been reported in several RCTs; others have emphasized beneficial treatment effects on radiographic progression even in patients without apparent clinical responses (6, 41-43).

Clinically, patients may be responding to therapy, yet their radiographs may nonetheless demonstrate disease progression (44). This was a lesson learned decades ago, when the 'gold standard' therapy of MTX was first adopted. Whether pathophysiological processes underlying the development of JSN and erosions are different, differ according to disease course, or reflect the limitations of radiographic assessment, these long-term data indicate that patient reported measures of physical function and imaging of joint damage should be conducted simultaneously on a regular basis.

Conclusion

Trials with the new DMARDS, as well as MTX, indicate that long term benefits in radiographic damage and physical function can be inferred from treatment data over 12 months. "Successful" patients who continue to do well will derive benefit from treatment for as long as 3 to 5 years. Recent RCTs indicate that combination therapy, initiated together, offers more improvement in radiographic progression and physical function than monotherapy, although the trial data cannot yet tell us which DMARD may be most beneficial in a given patient. Once treatment with any one agent – biologic or synthetic – is initiated, the addition of a second agent should occur rapidly if, for example, active disease persists 8 or 12 weeks later, without waiting for documented treatment failure. If this treatment paradigm is followed, along with regular assessments of radiographic damage and physical function, then patients may more likely derive long term clinical benefit than with traditional approaches.

References


Table III. Treatment changes in % of patients reporting limitations by SF-36.

<table>
<thead>
<tr>
<th></th>
<th>ERA: 12 months</th>
<th>US301: 24 months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rx w/ MTX or ETN Rx w/ MTX Rx w/ LEF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BL 12 mos. BL 24 mos. BL 24 mos.</td>
<td></td>
</tr>
<tr>
<td>Limitations walking 1 block</td>
<td>65.4% 35.9%</td>
<td>44.5% 38.4% 44.5% 17.3%</td>
</tr>
<tr>
<td>Limitations climbing 1 flight of stairs</td>
<td>74.7% 43.3%</td>
<td>66.9% 43.2% 66.9% 27.7%</td>
</tr>
<tr>
<td>Difficulty performing at work</td>
<td>89.6% 52.9%</td>
<td>88.5% 67.5% 88.5% 47.2%</td>
</tr>
</tbody>
</table>
31. KEYSTONE EC, KAVANAUGH AF, PEREZ JL, SPENCER-GREEN G: Adalimumab plus methotrexate provides sustained improvements in physical function over 2 years of treatment in patients with rheumatoid arthritis. *Ann Rheum Dis* 2004; 63; S278.
36. STRAND V, PINCUS T: The Health Assessment Questionnaire (HAQ) provides a single effective measure to discriminate active from placebo treatment in randomized controlled trials (RCTs) in rheumatoid arthritis (RA). *Arthritis Rheum* 2003; 48: 3656 (LB 6).
41. BOERS M, for the COBRA STUDY GROUP: Demonstration of response in rheumatoid arthritis patients who are non-responders according to the American College of Rheumatology 20% criteria: the paradox of beneficial treatment effects in non-responders in the ATTRACT trial. *Arthritis Rheum* 2001; 44: 2703-4.
42. KEYSTONE E, HAN C, KEENAN GF, VAN DER HEIJDE D, HARRIGAN M: Infliximab plus methotrexate prevents structural damage progression in rheumatoid arthritis patients which was independent of clinical response. *Arthritis Rheum* 2001; 44: S81.