DAS of 28 joints (DAS-28) and serum angiogenic factors, although they did not include the power Doppler ultrasonography (PDUS) score of the finger joints of the hand, which are frequently affected by RA.

We focused on the association of the PDUS score including the MCP joints with serum biomarkers as well as clinical disease activity. We found that the PDUS score of 24 synovial sites at 12 joints reflects the clinical disease activity and serum biomarkers. Second, for convenience, we reduced the number of joints to six synovial sites at six joints and also found that the PDUS score of six synovial sites at six joints is clearly correlated with the clinical disease activity and serum biomarkers.

Materials and methods

RA patient and healthy control samples

Twenty-two RA patients were selected to be enrolled in the present study from the Unit of Translational Medicine, Department of Immunology and Rheumatology, Graduate School of Biomedical Sciences, Nagasaki University. All of the patients fulfilled the 1987 criteria of the ACR for RA [4]. The patients underwent clinical, laboratory and PDUS evaluation on the same day. We also collected serum samples from eight healthy controls without musculoskeletal disorder. Subjects' written consent was obtained according to the Declaration of Helsinki, and the design of the work was approved by the Institutional Review Board of Nagasaki University.

Clinical and laboratory assessment

Clinical evaluation was performed by Japan College of Rheumatology-certified rheumatologists (A.K. and K.E.), who were blinded to the PDUS findings. Disease activity was evaluated by DAS-28, simplified disease activity index (SDAI) and clinical disease activity index (CDAI). In using DAS-28, we followed the criteria set by the European League against Rheumatism (EULAR), and in using CDAI and SDAI, we followed the method recommended by Smolen and colleagues [5].

The following laboratory variables were assessed: RF (Dade Behring, Marburg, Germany; cut-off value, 14 IU/ml), anti-CCP antibodies (DIASTAT Anti-CCP, Axis-Shield, Dundee, UK; cut-off value, 4.5 U/ml), CRP (Eiken Chemical Co. Ltd, Tokyo, Japan), ESR, VEGF (Quantikine, R&D Systems, Abingdon, UK), MMP-3 (Daiichi Pure Chemicals, Fukuoka, Japan), MMP-9 (Biotrak ELISA System, GE Healthcare, USA) and tissue inhibitor of metalloproteinases-1 (TIMP-1; Biotrak ELISA System, GE Healthcare, USA). Clinical disease activity as well as serum variables were evaluated on the same day as US examination.

US examination

Each patient underwent a US assessment by a Japan College of Rheumatology-certified rheumatologist (S.K.), who was blinded to the clinical and laboratory findings. Images from all the examinations were stored, and the US scoring reliability was examined by assessing 24 synovial sites in randomly selected patients at the end of the study. This assessment was carried out by Japan College of Rheumatology-certified rheumatologists (S.K., N.I., K.F. and T.O.) with consensus. A systematic multiplanar grey scale (GS) and PD examination of 12 joints was performed with the same scanner (TOSHIBA AplioXG; Toshiba Medical Systems Corporation, Tochigi, Japan) using a multifrequency linear transducer (12 MHz) according to the EULAR guidelines [6]. The US score included the following 24 synovial sites at 12 joints: bilateral elbows (anterior and posterior recess), wrists (dorsal and carpal recess), second and third MCP joints (dorsal and palmar recess), knees (suprapatellar and lateral parapatellar recess) and ankles (anterior tibiotalar recess, medial tendon sheaths and lateral tendon sheaths). Signs of OA were not detected by US and X-ray in the examined joints.

The IA, tenosynovial and intrabursal PD signals were graded on a semi-quantitative scale from 0 to 3 (Grade 0 = absence, no synovial flow; Grade 1 = mild, $\leqslant 3$ isolated singles; Grade 2 = moderate, >3 isolated singles or confluent signal in less than half of the synovial area; Grade 3 = marked, signals in more than half of the synovial area). These scores corresponded to the maximum score for PD signals obtained from any of the synovial sites evaluated at each joint, as documented by Naredo *et al.* [2]. The sum of the PD signal scores obtained from each joint was used as the PDUS score, as reported by Naredo *et al.* [2]. The 12-joint (12j)-PDUS score was the sum of the scores of the above 24 synovial sites at 12 joints. In an attempt to expand the convenience of ultrasonography in clinical practice, we have chosen six synovial sites from six joints including the bilateral wrists (dorsal recess) and second and third MCP joints (dorsal recess). The six-joint (6j)-PDUS score was the sum of the six synovial sites.

Statistical analyses

Within-group comparisons were made using the Mann–Whitney *U*-test. Correlations were assessed with Spearman's correlation coefficient test. The overall significance level for statistical analysis was 5% (two-sided). $P < 0.05$ was considered statistically significant.

Results

Patient characteristics

The demographic and clinical characteristics of 22 RA patients (5 males and 17 females) are as follows. The mean (S.D.) (range) age of the patients was 64 (9) (48–81) years. The mean (S.D.) (range) of disease duration was 2.3 (2.5) (0.25–10) years, which corresponded to relatively early-stage disease. RF and anti-CCP antibodies were positive in 15 (68.2%) and 18 (81.8%) patients, respectively. They received either synthetic DMARDs ($n = 14$), a combination of synthetic DMARDs plus TNF inhibitor ($n = 1$) or TNF inhibitor monotherapy ($n = 1$). Six patients were not treated with DMARDs. The mean tender joint counts (TJCs), swollen joint counts (SJCs), ESR, CRP, DAS-28, SDAI and CDAI were 9.2, 8.0, 58.8, 2.38, 5.69, 30.6 and 28.3, respectively, which indicate that

patients with relatively high disease activity were included in the present study.

Twelve j-PDUS scores and serum biomarkers

The median (range) of PDUS scores was 13.5 (1-39). Serum VEGF, MMP-3, MMP-9 and TIMP-1 were significantly higher in RA patients than in healthy controls—the mean levels of serum biomarkers; RA patients *vs* healthy controls (*P*-value, Mann–Whitney *U*-test)—VEGF; 695 *vs* 308 pg/ml ($P < 0.0001$), MMP-3; 185 *vs* 30 U/l ($P < 0.001$), MMP-9; 1962 *vs* 55 pg/ml ($P < 0.0001$) and TIMP-1; 214 *vs* 160 pg/ml ($P < 0.05$).

The correlations of 12j-PDUS scores with disease activity and serum biomarkers

The correlation of DAS-28 with SDAI or CDAI was extremely high, indicating that the physical examination was well performed (Table 1). The 12j-PDUS scores from 24 synovial sites were significantly positively correlated with TJC, SJC, ESR, CRP, DAS-28, SDAI, CDAI, serum VEGF, MMP-3 and TIMP-1, whereas they were not correlated with serum MMP-9 (Table 2). In particular, DAS-28 ($r = 0.72$, $P < 0.001$) and serum VEGF ($r = 0.62$, $P < 0.01$) strongly correlated with PDUS scores.

The correlations between clinical disease activity and serum biomarkers are shown in Table 1. All serum biomarkers correlated with inflammatory markers such as CRP and ESR. With regard to angiogenic factors, VEGF correlated well with the variables other than MMP-9 or TIMP-1 (Table 1).

Six j-PDUS scores can be an alternative for 12j-PDUS scores

The 6j-PDUS scores were strongly correlated with 12j-PDUS scores ($r = 0.92$, $P < 0.0001$). Accordingly, 6j-PDUS scores were significantly correlated with TJC ($r = 0.50$, $P < 0.05$), SJC ($r = 0.44$, $P < 0.05$), ESR ($r = 0.57$, $P < 0.05$), DAS-28 ($r = 0.67$, $P < 0.01$), SDAI ($r = 0.55$, $P < 0.05$), CDAI ($r = 0.54$, $P < 0.05$) and serum VEGF ($r = 0.62$, $P < 0.01$), whereas they were not correlated with serum MMP-3, MMP-9 and TIMP-1. Although these associations were slightly weaker than those with 12j-PDUS scores, the tendencies of 6j-PDUS and 12j-PDUS scores were very similar to each other.

Discussion

We have verified additional information regarding PDUS scores in patients with RA. First, our data included the small MCP joints. Since the second and third MCP joints are considered to be important areas for radiographic imaging of RA, as reported by Naredo *et al.* for US [2] and by OMERACT for MRI [7], our present data may reinforce the utility of PDUS. Although our present study includes relatively elderly patients, signs of OA were not detected in the examined joints, indicating that our results reflect rheumatoid inflammatory change.

Second, we have chosen other biomarkers. MMP-9 is important for the budding of endothelial cells, and TIMP-1 is an inhibitor of MMP-9; both are elevated in serum as well as in the synovial tissues of RA [8, 9]. As suspected, TIMP-1 was correlated with PDUS score and several other biomarkers, although its correlation was weaker than that of VEGF. MMP-9 tended to correlate with PDUS score; however, it did not reach statistical significance. Since the budding of endothelial cells is an early step in angiogenesis, MMP-9 may be important in the early phase of rheumatoid synovitis. The selection of very early-stage RA may be necessary to identify any association of MMP-9 with PDUS score.

Third, we have assessed SDAI and CDAI in the present study. The present study has revealed a clear correlation of PDUS score with SDAI and CDAI, although that of DAS-28 ESR was better. These data reinforce the validity of PDUS for the measurement of the disease activity of RA.

Fourth, for better clinical availability, we have reduced the number of sites examined by US to only six sites of the wrist and finger joints. These methods are simple and can save time that would be spent on scanning. Since the correlation of disease activity and PDUS was weaker than those with 24 synovial sites, further studies with larger numbers of patients should be necessary.

Recent investigations have found that the presence of the PDUS signal is a better predictor of further radiographic joint destruction than DAS-28 [10, 11].

Table 1 Correlations between disease activity and serum biomarkers

	DAS-28	SDAI	CDAI	ESR	CRP	VEGF	MMP-3	MMP-9
SDAI	0.93***							
CDAI	0.93***	0.99***						
ESR	0.64**	0.44*	0.43					
CRP	0.64**	0.64**	0.59**	0.71**				
VEGF	0.59**	0.54*	0.51*	0.62**	0.70**			
MMP-3	0.61**	0.58**	0.55*	0.57**	0.68**	0.60**		
MMP-9	0.27	0.23	0.18	0.43*	0.49*	0.29	0.39	
TIMP-1	0.39	0.42	0.37	0.52*	0.71**	0.58**	0.69**	0.38

Correlations were assessed with Spearman's correlation coefficient test. *$P < 0.05$, **$P < 0.01$, ***$P < 0.0001$.

TABLE 2 Correlations of PDUS score with disease activity and serum biomarkers

Characteristics	r	P-value
Tender joints	0.52	0.017
Swollen joints	0.48	0.028
ESR	0.62	0.005
CRP	0.47	0.03
DAS-28	0.72	<0.001
SDAI	0.6	0.006
CDAI	0.6	0.006
Serum VEGF levels	0.62	0.005
Serum MMP-3 levels	0.47	0.03
Serum MMP-9 levels	0.38	0.08
Serum TIMP-1 levels	0.54	0.014

Correlations were assessed with Spearman's correlation coefficient test.

Therefore, it is very important to search the variables that correlate with PDUS score. Among the biomarkers and clinical disease activity indices in the present study, however, DAS-28 was the strongest variable that correlates with PDUS scores. VEGF was best in the biomarkers although it was weaker than that of DAS-28. These data may suggest that comprehensive analysis is necessary to identify the best biomarkers to reflect the severity of the PDUS score.

In conclusion, PDUS, especially when focused on the area of the wrist and finger joints, is an excellent tool for the evaluation of disease activity in patients with RA. Our six-site evaluation method can be adequately tolerated in clinical practice.

Rheumatology key messages

- Standard as well as simplified PDUS scores reflected clinical disease activity and serum variables, including angiogenic factors.
- Our six-site evaluation method can be adequately tolerated in clinical practice.

Disclosure statement: The authors have declared no conflicts of interest.

References

1 Joshua F, Edmonds J, Lassere M. Power Doppler ultrasound in musculoskeletal disease: a systematic review. Semin Arthritis Rheum 2006;36:99–108.

2 Naredo E, Rodríguez M, Campos C *et al.* Validity, reproducibility, and responsiveness of a twelve-joint simplified power Doppler ultrasonographic assessment of joint inflammation in rheumatoid arthritis. Arthritis Rheum 2008; 59:515–22.

3 Kurosaka D, Hirai K, Nishioka M *et al.* Clinical significance of serum levels of vascular endothelial growth factor, angiopoietin-1, and angiopoietin-2 in patients with rheumatoid arthritis. J Rheumatol 2010;37:1121–8.

4 Arnett FC, Edworthy SM, Bloch DA *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988; 31:315–24.

5 Aletaha D, Nell VP, Stamm T *et al.* Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. Arthritis Res Ther 2005;7:R796–806.

6 Backhaus M, Burmester GR, Gerber T *et al.* Working Group for Musculoskeletal Ultrasound in the EULAR Standing Committee on International Clinical Studies including Therapeutic Trials. Guidelines for musculoskeletal ultrasound in rheumatology. Ann Rheum Dis 2001;60: 641–9.

7 Ostergaard M, Peterfy C, Conaghan P *et al.* OMERACT rheumatoid arthritis magnetic resonance imaging studies: core set of MRI acquisitions, joint pathology definitions, and the OMERACT RA-MRI scoring system. J Rheumatol 2003;30:1385–6.

8 Ram M, Sherer Y, Shoenfeld Y. Matrix metalloproteinase-9 and autoimmune diseases. J Clin Immunol 2006;26:299–307.

9 Tchetverikov I, Ronday HK, Van El B *et al.* MMP profile in paired serum and synovial fluid samples of patients with rheumatoid arthritis. Ann Rheum Dis 2004;63:881–3.

10 Brown AK, Conaghan PG, Karim Z *et al.* An explanation for the apparent dissociation between clinical remission and continued structural deterioration in rheumatoid arthritis. Arthritis Rheum 2008;58:2958–67.

11 Naredo E, Möller I, Cruz A *et al.* Power Doppler ultrasonographic monitoring of response to anti-tumor necrosis factor therapy in patients with rheumatoid arthritis. Arthritis Rheum 2008;58:2248–56.

Mod Rheumatol (2013) 23:36–43
DOI 10.1007/s10165-012-0628-7

ORIGINAL ARTICLE

Musculoskeletal ultrasonography assists the diagnostic performance of the 2010 classification criteria for rheumatoid arthritis

Shin-ya Kawashiri · Takahisa Suzuki · Akitomo Okada · Satoshi Yamasaki · Mami Tamai · Hideki Nakamura · Tomoki Origuchi · Akinari Mizokami · Masataka Uetani · Kiyoshi Aoyagi · Katsumi Eguchi · Atsushi Kawakami

Received: 1 February 2012 / Accepted: 2 March 2012 / Published online: 3 April 2012

Abstract
Objective We investigated whether musculoskeletal ultrasonography (MSKUS) assists the diagnostic performance of the 2010 rheumatoid arthritis (RA) classification criteria.
Methods Sixty-nine early arthritis patients were consecutively enrolled. None of the patients had been treated. In MSKUS of bilateral wrist and finger joints from 22 sites, the findings obtained by gray-scale and power Doppler (PD) assessment were graded on a semiquantitative scale from 0 to 3. Plain magnetic resonance imaging (MRI) of both wrist and finger joints was also examined. Diagnosis of RA was defined by the initiation of disease-modifying antirheumatic drugs within the first 3 months. The diagnostic performance of the patients was evaluated at entry using 2010 RA classification criteria in conjunction with MSKUS.
Results The indispensable MSKUS finding for differentiating RA was the presence of a PD grade 2 or 3 that was superior to 2010 RA classification criteria or MRI-proven bone edema. We propose that the decision tree algorithm of 2010 RA classification criteria with PD grade 2 or 3 reveals the best discriminative ability.
Conclusion MSKUS, especially with a strong PD signal, is very useful to assist the diagnostic performance of the 2010 RA classification criteria in the early recognition of RA.

Keywords Rheumatoid arthritis · 2010 RA classification criteria · Ultrasonography · Power Doppler · MRI

Abbreviations

ACPA	Anticyclic citrullinated peptide antibody
ACR	American College of Rheumatology
CRP	C-reactive protein
DMARDs	Disease-modifying antirheumatic drugs
ESR	Erythrocyte sedimentation rate
EULAR	European League Against Rheumatism
Gd-DTPA	Gadolinium-diethylenetriamine pentaacetic acid
GS	Gray-scale
IP	Interphalangeal
MCP	Metacarpophalangeal

S. Kawashiri (✉) · T. Suzuki · A. Okada · S. Yamasaki · H. Nakamura · A. Kawakami
Department of Immunology and Rheumatology, Unit of Translational Medicine, Nagasaki University Graduate School of Biomedical Sciences, 1-7-1 Sakamoto, Nagasaki 852-8501, Japan
e-mail: shin-ya@hotmail.co.jp

M. Tamai
Center for Health and Community Medicine, Nagasaki University, Nagasaki, Japan

T. Origuchi
Department of Health Sciences, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

A. Mizokami
Department of Internal Medicine, Nagasaki Municipal Hospital, Omura, Japan

M. Uetani
Department of Radiological Sciences, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

K. Aoyagi
Department of Public Health, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

K. Eguchi
Department of Internal Medicine, Sasebo City General Hospital, Sasebo, Japan

MMP	Matrix metalloproteinase
MRI	Magnetic resonance imaging
MSKUS	Musculoskeletal ultrasonography
NPV	Negative predictive value
PD	Power Doppler
PIP	Proximal interphalangeal
PPV	Positive predictive value
RA	Rheumatoid arthritis
RF	Rheumatoid factor
SJC	Swollen joint counts
T2T	Treat to target
TJC	Tender joint counts

Introduction

Early diagnosis and the treat to target (T2T) strategy are now indispensable for managing rheumatoid arthritis (RA) [1]. Application of the T2T strategy using the tight control approach in patients with RA, especially those with early-stage RA, has been shown to improve RA outcomes [1, 2]. Thus, the early recognition of RA is a great benefit in managing patients with RA. The 1987 American College of Rheumatology (ACR) classification criteria for RA [3] are not designed for early classification of RA. Consequently, to identify patients with erosive arthritis early, a task force of experts from both the ACR and the European League Against Rheumatism (EULAR) derived new classification criteria [4]. These new criteria, the 2010 RA classification criteria, have been verified to classify patients early as having RA more efficiently than the 1987 criteria; however, a substantial population is not still classified as having RA, even by the 2010 RA classification criteria [4].

Although physical examination is still the gold standard by which to identify the presence of arthritis [4], it has come to be apparent that modern imaging techniques such as musculoskeletal ultrasonography (MSKUS) and magnetic resonance imaging (MRI) are more sensitive than physical examination for detecting joint injury in patients with RA, especially early-stage RA [5–9]. MSKUS is well tolerated and can image a large number of joints at multiple time points over a relatively short period of time [10, 11]. Varying kinds of joint injury, including synovitis, tenosynovitis, and bone erosion, can be recorded by gray-scale (GS) and power Doppler (PD) [5–8, 10–13]. We recently reported the utility of PD to reflect clinical disease activity as well as serum biomarkers in patients with RA [14].

We speculated that the detection sensitivity for synovitis would be increased if MSKUS was routinely incorporated into clinical practice for patients with early arthritis. The objective of the study reported here was to evaluate whether the findings of MSKUS, in comparison with MRI, assist the diagnostic performance of the 2010 RA classification criteria.

Materials and methods

Patients

Sixty-nine early arthritis patients suspected of having RA were consecutively recruited. Patients who could be classified as non-RA at first visit were excluded. In addition, we excluded patients who had experience with disease-modifying antirheumatic drugs (DMARDs), including biologics and glucocorticoids. All patients were recruited from the Unit of Translational Medicine, Department of Immunology and Rheumatology, Graduate School of Biomedical Sciences, Nagasaki University, and the Department of Internal Medicine, Nagasaki Municipal Hospital, from May 2010 through February 2011. The duration from the appearance of symptoms to entry into the study in these 69 patients was <1 year. Patients gave their informed consent to be subjected to the protocol that was approved by the Institutional Review Board of Nagasaki University. This study was a prospective single-center observational study. Follow-up periods were at least 6 months.

Clinical and laboratory assessment

A clinical diagnosis of RA was comprehensively made by Japan College of Rheumatology (JCR)-certified rheumatologists (AK, HN, SY, and KE) using clinical histories, physical findings, blood tests including rheumatoid factor (RF) (Dade Behring, Marburg, Germany; cutoff value, 14 IU/ml), anticyclic citrullinated peptide antibodies (ACPA) (DIASTAT Anti-CCP, Axis-Shield, Dundee, UK; cutoff value, 4.5 U/ml), C-reactive protein (CRP) (Eiken Chemical Co., Ltd., Tokyo, Japan), erythrocyte sedimentation rate (ESR), matrix metalloproteinase 3 (MMP-3) (Daiichi Pure Chemicals, Fukuoka, Japan), 2010 RA classification criteria, plain radiography, ultrasound (US) findings, and MRI findings. All patients underwent the examinations except for MRI every 1–3 months. If JCR-certified rheumatologists introduced DMARDs within the first 3 months according to the above information, patients are diagnosed as having RA. Therefore, not only 2010 RA classification criteria but other information, such as MSKUS, MRI, and clinical course, are actually involved in these processes.

US examination

Each patient underwent a US assessment on the same day as the clinical evaluation by a JCR-certified rheumatologist

(SK), who was blinded to the clinical and laboratory findings. Images from all the examinations were stored, and the US scoring reliability was examined in randomly selected patients at the end of the study. This assessment was carried out by JCR-certified rheumatologists (SK, TS, AO, and TO) with consensus. A systematic multiplanar GS and PD examination of 22 joints was performed with the same scanner (TOSHIBA AplioXG) using a multifrequency linear transducer (12 MHz). The US score comprised the following 22 joints: bilateral wrists (intracarpal, radiocarpal, and ulnocarpal recesses) and finger joints, including the first through fifth metacarpophalangeal (MCP) joints, the first interphalangeal (IP) joint, and the second to fifth proximal interphalangeal (PIP) joints (dorsal recess). Flexor tendons of fingers and six components of extensor tendons of wrists were scanned. All joint regions were sonographically examined in a standardized manner according to the European League Against Rheumatism (EULAR) guidelines [13]. These are the same sites at which MRI has been used to examine patients with early arthritis, as we previously described [9, 15]. US examination of each patient took about 30 min, including documentation.

Each joint was scored for GS and PD on a semiquantitative scale [16] (synovial hypertrophy in GS: grade 0 = absence, no synovial thickening; grade 1 = mild, minimal synovial thickening filling the angle between the periarticular bones without bulging over the line linking the tops of the bones; grade 2 = moderate, synovial thickening bulging over the line linking the tops of the periarticular bones but without extension to at least one bone diaphysis; grade 3 = marked, synovial thickening bulging over the line linking the tops of the periarticular bones and with extension to at least one of the bone diaphyses; PD signals: grade 0 = absence, no synovial flow; grade 1 = mild, single-vessel signals; grade 2 = moderate, confluent signal in less than half of the synovial area; grade 3 = marked, signals in more than half of the synovial area) from 0 to 3, and presence or absence of tenosynovitis was noted. Tenosynovitis is defined by abnormal hypoechoic or anechoic material with or without fluid inside the tendon sheath and with positive PD signals in two perpendicular planes [17]. These scores corresponded to the maximum score for GS and PD obtained from any of the synovial sites evaluated at each joint. The sums of the GS and PD scores obtained from each joint were used as the GS score and PD score (range 0–66), respectively.

MRI examination

Plain MRI of both wrists and finger joints were acquired using a 1.5-T system (Sigma, GE Medical Systems, Milwaukee, WI, USA) with an extremity coil, as we recently described [9, 15, 18, 19]. Fifty-four patients were examined by MRI within a week of their US evaluation. T1-weighted spin-echo (TR 450, TE 13) images and short-tau inversion recovery (STIR; TR 3000, TE 12, T1 160) images were acquired simultaneously. The images were evaluated for synovitis, bone edema, and bone erosion at 15 sites in each finger and wrist at the distal radioulnar joint, radiocarpal joint, midcarpal joint, first carpometacarpal joint, second through fifth carpometacarpal joints (together), first through fifth metacarpophalangeal joints (separately), and first through fifth proximal interphalangeal joints (PIP joints) separately (for a total of 30 sites in both hands), as we recently reported [9, 15, 18, 19].

Statistical analyses

Within-group comparisons were made using Mann–Whitney's U test and the χ^2 test (Fisher's exact probability test when appropriate). The overall significance level for statistical analysis was 5 % (two-sided). P values <0.05 were considered statistically significant.

Results

Patient characteristics and diagnoses

The demographic and clinical characteristics of 69 patients are shown in Table 1. Thirty-seven patients (53.6 %) were diagnosed as having RA. Synthetic DMARDs were introduced within the first 3 months to these 37 patients. The initial treatments were methotrexate in 35 patients, sulfasalazine in one, and tacrolimus in one. Thirty-two patients (46.4 %) were diagnosed with other diseases (non-RA) during the follow-up periods, although they could not be classified as non-RA at entry. The diagnoses of these patients were osteoarthritis ($n = 8$), undifferentiated arthritis/arthropathy ($n = 7$), Sjögren syndrome ($n = 4$), polymyalgia rheumatica ($n = 2$), limited-type systemic sclerosis ($n = 2$), tenosynovitis ($n = 2$), reactive arthritis ($n = 1$), polymyositis ($n = 1$), immunoglobulin (Ig)G_4-related disease ($n = 1$), sarcoidosis ($n = 1$), adult T-cell leukemia (ATL), familial Mediterranean fever ($n = 1$), and phalangeal microgeodic syndrome ($n = 1$). The mean disease duration was approximately 4 months in both RA and non-RA patients. The swollen joint counts ($p = 0.0104$) and CRP ($p = 0.0003$) and ESR ($p = 0.0009$) values were higher in RA patients than in non-RA patients, but the tender joint counts were not different. The seropositive rates of RF (70.3 %, $p = 0.0002$) and ACPA (62.2 %, $p < 0.0001$) were significantly higher in RA than in non-RA patients. Patients with high MMP-3 were also predominantly distributed in the RA group (48.6 %, $p = 0.0432$).

Comparison of MSKUS findings between RA and non-RA patients

The MSKUS findings in RA and non-RA patients are shown in Table 2. The rates at which GS grade ≥1 ($p = 0.0005$), GS grade ≥2 ($p < 0.0001$), GS grade = 3 ($p < 0.0001$), PD grade ≥1 ($p < 0.0001$), and PD grade ≥2 ($p < 0.0001$) were present at any joint were significantly higher in RA than in non-RA patients. However, GS grade ≥1, GS grade ≥2, and PD grade ≥1 also occurred in non-RA patients, as 23 (71.9 %), 12 (37.5 %), and ten (31.3 %) patients were positive for the above grades, respectively, out of 32 non-RA patients. The occurrence of PD grade = 3 was specific to RA patients; however, it was only found in four of 37 RA patients (10.8 %). Both GS and PD scores were significantly higher in RA than in non-RA patients. The frequency of findings of tenosynovitis was prominent in the RA group, but the difference from the frequency in the non-RA group was not significant. Bone erosions were specifically detected in RA patients; however, the rate was not high (18.9 %, $p = 0.0094$). Accordingly, PD grade ≥2 at any joint is considered to be most important MSKUS findings in RA patients.

Comparison of plain MRI findings between RA and non-RA patients

The plain MRI findings in RA and non-RA patients are also shown in Table 2. As most patients with RA expressed symmetrical synovitis that was also found in non-RA patients, we could not find statistical significance in this result. As suspected, bone edema was significantly distributed in the RA group compared with the non-RA group; however, that was not so remarkable compared with MSKUS findings. Patients with MRI-proven bone erosion tended to be distributed in the RA group, but the difference did not reach statistical significance ($p = 0.0838$).

Laboratory data, MSKUS findings, MRI findings, and 2010 RA classification criteria for the diagnosis of RA

Sensitivity, specificity, and accuracy of laboratory data, MSKUS findings, MRI findings, and 2010 RA classification criteria are shown in Table 3. The presence of ACPA was the most specific laboratory data distributed in patients with RA. Surprisingly, the presence of MSKUS findings, especially the presence of PD grade 2 or 3 at any joint, was very specific in

Table 1 Demographic, clinical, and laboratory characteristics at baseline

	RA ($N = 37$)	Non-RA ($N = 32$)	P value
Age (years[a])	53.6 ± 17.2	54.5 ± 12.5	NS
Female/male (n)	28/9	26/6	NS
Durations of symptom (months[a])	4.0 ± 3.0	3.7 ± 2.9	NS
≥1.5 months/<1.5 months	31/6	24/8	NS
Tender joint counts (n[a])	7.9 ± 7.6	5.6 ± 6.9	NS
Swollen joint counts (n[a])	5.6 ± 6.9	3.4 ± 6.3	0.0104
CRP			
Positive/negative	24/13	8/24	0.0009
Value (mg/dl[a])	1.29 ± 2.94	0.40 ± 1.09	0.0003
ESR			
Positive/negative	27/10	11/21	0.0013
Value (mm/h[a])	32.2 ± 24.5	18.0 ± 20.6	0.0009
CRP and/or ESR			
Positive/negative	31/6	13/19	0.0002
RF			
Positive/negative	26/11	8/24	0.0002
Titers: >×3/≤×3	17/20	3/29	0.0083
ACPA			
Positive/negative	23/14	2/30	1.4×10^{-6}
Titers: >×3/≤×3	23/14	1/31	2.8×10^{-7}
IgM-RF and/or ACPA			
Positive/negative	27/10	9/23	0.0002
Titers: >×3/≤×3	23/14	4/28	2.5×10^{-5}
MMP-3			
Positive/negative	18/19	8/24	0.0432

Within-group comparisons were assessed with Mann–Whitney's U test and the χ^2 test (Fisher's exact probability test when appropriate)

NS not significant, *RF* rheumatoid factor, *ACPA* anti-CCP antibody, *MMP-3* matrix metalloproteinase-3

[a] Mean ± standard deviation

Table 2 Ultrasonography and MRI findings at baseline

	RA ($N = 37$)	Non-RA ($N = 32$)	P value
MSKUS			
Gray-scale			
Grade ≥1 presence/absence	37/0	23/9	0.0005
Grade 2 or 3 presence/absence	33/4	12/20	6.9×10^{-6}
Grade 3 presence/absence	21/16	1/31	1.9×10^{-6}
Total GS score (0–66)[a]	9.4 ± 7.6	3.7 ± 4.0	0.0001
Power Doppler			
Grade ≥1 presence/absence	34/3	10/22	1.7×10^{-7}
Grade 2 or 3 presence/absence	30/7	2/30	5.1×10^{-10}
Grade 3 presence/absence	4/33	0/32	0.0764
Total PD score (0–66)[a]	4.2 ± 3.7	0.6 ± 1.1	9.7×10^{-9}
Tenosynovitis			
Presence/absence	21/16	6/26	0.0013
Bone erosion			
Presence/absence	7/30	0/32	0.0094
	RA ($N = 32$)	Non-RA ($N = 22$)	P value
MRI			
Symmetrical synovitis			
Presence/absence	28/4	16/6	NS
Bone edema			
Presence/absence	15/17	4/18	0.0300
Bone erosion			
Presence/absence	9/23	2/20	0.0838

Within-group comparisons were assessed with Mann–Whitney's U test and the χ^2 test (Fisher's exact probability test when appropriate)

RA rheumatoid arthritis, *MSKUS* musculoskeletal ultrasonography, *GS* gray-scale, *PD* power Doppler, *MRI* magnetic resonance imaging, *NS* not significant

[a] Mean ± standard deviation

RA. If we considered patients to have RA in cases in which MSKUS findings showed PD grade 2 or 3, the diagnostic performance of MSKUS for RA had sensitivity 81.1 %, specificity 93.8 %, positive predictive value (PPV) 93.8 %, negative predictive value (NPV) 81.1 %, and accuracy 87.0 %. The 2010 RA classification criteria classified RA with sensitivity 59.5 %, specificity 87.5 %, PPV 84.6 %, NPV 65.1 %, and accuracy 72.5 %, suggesting that the presence of PD grade 2 or 3 may have been more specific than the 2010 RA classification criteria. In accordance with data shown in Table 2, MRI-proven bone edema could not differentiate RA from non-RA compared with PD grading.

We tried to combine 2010 RA classification criteria with the PD grade 2 or 3 rule for the clinical diagnosis of RA, and the results are shown in Fig. 1. We initially applied 2010 RA classification criteria, and if the patients did not fulfill those criteria, the PD grade 2 or 3 rule was introduced. We found that this decision tree can differentiate patients more efficiently than can the PD grade 2 or 3 rule alone.

Discussion

The authors of previous assessments of the performance of the 2010 RA classification criteria have usually tried to identify patients with RA as those who were treated with DMARDs within the first year of the follow-up period [20–23]. As of this writing, the 2010 RA classification criteria were published last year and are going to be applied in the clinical field of rheumatology. Rheumatologists tend to start DMARDs earlier in patients who are expected to develop erosive arthritis. Therefore, in this study, we considered patients to have RA if their physicians had started DMARDs within the first 3 months. This clinical setting may clarify more definitely which patients should be considered to have RA for the purpose of applying the T2T strategy that has come to be widely recommended.

Diagnostic performance of the 2010 RA classification criteria in this study was fairly good, with both the specificity and PPV around 85 %. As this was a prospective investigator-initiated clinical study, physicians were able to choose the treatment at every visit according to the clinical status of patients fulfilling the 2010 RA classification criteria. Thus, ithe score according to the 2010 RA classification criteria at each visit may be directly involved in the physician's decision, which associated with the increment of specificity and PPV of the 2010 RA classification criteria. However, the levels of other components, such as sensitivity, NPV, and accuracy, were not high, indicating that additional procedures may be necessary to assist the

Table 3 Performance of laboratory data, ultrasonography findings, and 2010 rheumatoid arthritis (RA) classification criteria

	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
Laboratory data					
CRP (positive)	64.9	75.0	75.0	64.9	69.6
ESR (positive)	73.0	65.6	71.1	67.7	69.6
RF (positive)	70.3	75.0	76.5	68.6	72.5
ACPA (positive)	62.2	93.8	92.0	68.2	76.8
MMP-3 (positive)	48.6	75.0	69.2	55.8	60.9
MSKUS					
Gray-scale; grade ≥1	100	28.1	61.7	100	66.7
Gray-scale; grade 2 or 3	89.2	62.5	73.3	83.3	76.8
Gray-scale; grade 3	56.8	96.9	95.5	66.0	75.4
Power Doppler; grade ≥1	91.9	68.8	77.3	88.0	81.2
Power Doppler; grade 2 or 3	81.1	93.8	93.8	81.1	87.0
Power Doppler; grade 3	10.8	100	100	49.2	52.2
Tenosynovitis (positive)	56.8	81.3	77.8	61.9	68.1
Bone erosion (positive)	18.9	100	100	51.6	56.5
MRI					
Symmetrical synovitis (positive)	87.5	27.3	63.6	60.0	63.0
Bone edema (positive)	46.9	81.8	78.9	51.4	61.1
Bone erosion (positive)	28.1	90.9	81.8	46.5	53.7
2010 RA classification criteria	59.5	87.5	84.6	65.1	72.5

RF rheumatoid factor, *ACPA* anti-CCP antibody, *MMP-3* matrix metalloproteinase-3, *MSKUS* musculoskeletal ultrasound, *PPV* positive predictive value, *NPV* negative predictive value, *MRI* magnetic resonance imaging

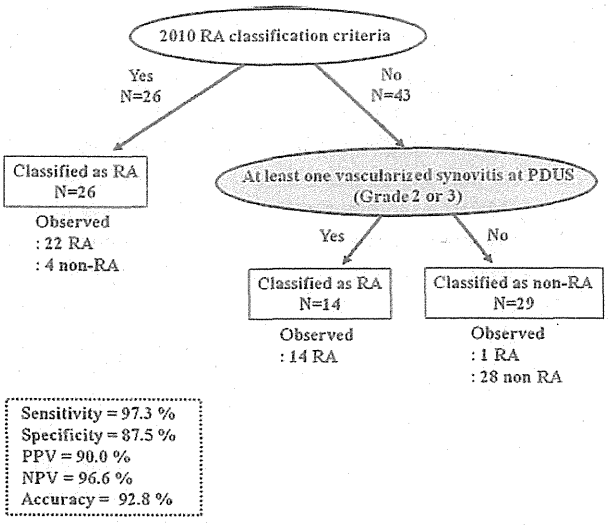

Fig. 1 Decision tree algorithm for diagnosis of early arthritis patients by 2010 rheumatoid arthritis (RA) classification criteria in conjunction with power Doppler PD grade 2 or 3; 2010 RA criteria were initially applied to 69 patients. If the patients fulfilled the criteria, they were classified as having RA (26 patients). PD grade 2 or 3 rule was applied for the remaining 43 patients. This tree algorithm classified patients as having RA at sensitivity 97.3 %, specificity 7.5 %, positive predictive value (PPV) 90.0 %, negative predictive value (NPV) 96.6 %, and accuracy 92.8 %

diagnostic performance of the 2010 RA classification criteria.

In this regard, we focused on MSKUS, as it is more sensitive and reliable than clinical examination for detecting joint injury in patients with RA [5–8]. Synovitis, tenosynovitis, and bone erosion are the major joint injuries that are frequently found in patients with RA examined by MSKUS [5–8, 10–13]. GS determines the hypertrophy of synovial tissues, whereas PD identifies vascularity [5–8, 10–13]. In our study, PD grade, especially grade 2 or 3, was highly specific in patients with RA. These data are consistent with the previous findings that the synovial vascularity determined by PD reflects RA disease activity more efficiently than do GS findings [24, 25]. The levels of statistical components were even better than those of the 2010 RA classification criteria, indicating that the presence of severe and active synovial inflammation detected by PD may deeply affect physicians' decisions to start DMARDs.

Although the US examiner was always blinded to the clinical and laboratory findings of patients in this study, physicians could take into consideration the results of US for the choice of DMARDs at each point. Therefore, it could also be said that PD overestimates the presence of RA and thus influences the initiation of or choice of DMARDs that was directly associated with our data. As for

MRI, the presence of bone edema is thought to be the most suitable indicator for a clinical diagnosis of RA. These results are consistent with our previous report that bone edema is able to predict the development of RA that fulfills the 1987 classification criteria from patients with early arthritis more efficiently than symmetrical synovitis and bone erosion [15]. As physicians judge patients as having RA based on findings of not only MSKUS but also MRI, we could state that PD grade 2 or 3 is superior to bone edema on plain MRI for making a clinical diagnosis of RA. If we obtain gadolinium-diethylenetriamine pentaacetic acid (Gd-DTPA)-enhanced MRI instead of plain MRI, bone edema may be more significantly involved in RA diagnosis. In our previous study, we found that the detection sensitivity of bone edema on plain MRI is 30 % less than that with Gd-DTPA-enhanced MRI [15]. We therefore propose a tree algorithm for clinical RA diagnosis that combines the 2010 RA classification criteria and PD, as shown in Fig. 1. This kind of approach can also be applied in patients with spondyloarthropathy, indicating that Amor's criteria in conjunction with vascularized enthesis bring good results [26]. Accordingly, our data identify that the tree algorithm shown in Fig. 1 can classify more patients as having RA at a high discriminative value compared with the 2010 RA classification criteria or PD alone, supposing more patients received the chance of early introduction of DMARDs. Our data may also indicate that the combination of physical examination and serology with a sensitive imaging technique, such as MSKUS, is the best way to identify erosive disease early. Filer et al. [7] reported that a combination of Leiden score, but not the 2010 RA classification criteria, with MSKUS-proven synovitis improved in clinical RA diagnosis. Our data may follow that result. Long-term follow-up and larger studies are warranted to confirm that MSKUS, especially PD, in combination with the 2010 RA classification criteria, is valuable for early identification of patients with erosive RA

Conflict of interest None.

References

1. Smolen JS, Aletaha D, Bijlsma JW, Breedveld FC, Boumpas D, Burmester G, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis. 2010;69:631–7.
2. Smolen JS, Landewé R, Breedveld FC, Dougados M, Emery P, Gaujoux-Viala C, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. Ann Rheum Dis. 2010;69:964–75.
3. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988;31:315–24.
4. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum. 2010;62:2569–81.
5. Naredo E, Bonilla G, Gamero F, Uson J, Carmona L, Laffon A. Assessment of inflammatory activity in rheumatoid arthritis: a comparative study of clinical evaluation with grey scale and power Doppler ultrasonography. Ann Rheum Dis. 2005;64:375–81.
6. Salaffi F, Filippucci E, Carotti M, Naredo E, Meenagh G, Ciapetti A, et al. Inter-observer agreement of standard joint counts in early rheumatoid arthritis: a comparison with grey scale ultrasonography—a preliminary study. Rheumatology (Oxford). 2008;47:54–8.
7. Filer A, de Pablo P, Allen G, Nightingale P, Jordan A, Jobanputra P, et al. Utility of ultrasound joint counts in the prediction of rheumatoid arthritis in patients with very early synovitis. Ann Rheum Dis. 2011;70:500–7.
8. Saleem B, Brown AK, Keen H, Nizam S, Freeston J, Karim Z, et al. Disease remission state in patients treated with the combination of tumor necrosis factor blockade and methotrexate or with disease-modifying antirheumatic drugs: a clinical and imaging comparative study. Arthritis Rheum. 2009;60:1915–22.
9. Tamai M, Kawakami A, Iwamoto N, Kawashiri SY, Fujikawa K, Aramaki T, et al. Comparative study of the detection of joint injury in early-stage rheumatoid arthritis by MRI of wrist and finger joints and physical examination. Arthritis Care Res (Hoboken). 2011;63:436–9.
10. Østergaard M, Pedersen SJ, Døhn UM. Imaging in rheumatoid arthritis–status and recent advances for magnetic resonance imaging, ultrasonography, computed tomography and conventional radiography. Best Pract Res Clin Rheumatol. 2008;22:1019–44.
11. Naredo E, Collado P, Cruz A, Palop MJ, Cabero F, Richi P, et al. Longitudinal power Doppler ultrasonographic assessment of joint inflammatory activity in early rheumatoid arthritis: predictive value in disease activity and radiologic progression. Arthritis Rheum. 2007;57:116–24.
12. Wakefield RJ, Gibbon WW, Conaghan PG, O'Connor P, McGonagle D, Pease C, et al. The value of sonography in the detection of bone erosions in patients with rheumatoid arthritis: a comparison with conventional radiography. Arthritis Rheum. 2000;43:2762–70.
13. Backhaus M, Burmester GR, Gerber T, Grassi W, Machold KP, Swen WA, Working Group for Musculoskeletal Ultrasound in the EULAR Standing Committee on International Clinical Studies including Therapeutic Trials, et al. Guidelines for musculoskeletal ultrasound in rheumatology. Ann Rheum Dis. 2001;60:641–9.
14. Kawashiri SY, Kawakami A, Iwamoto N, Fujikawa K, Satoh K, Tamai M, et al. The power Doppler ultrasonography score from 24 synovial sites or 6 simplified synovial sites, including the metacarpophalangeal joints, reflects the clinical disease activity and level of serum biomarkers in patients with rheumatoid arthritis. Rheumatology (Oxford). 2011;50:962–5.
15. Tamai M, Kawakami A, Uetani M, Takao S, Arima K, Iwamoto N, et al. A prediction rule for disease outcome in patients with undifferentiated arthritis using magnetic resonance imaging of the wrists and finger joints and serologic autoantibodies. Arthritis Rheum. 2009;61:772–8.
16. Szkudlarek M, Court-Payen M, Jacobsen S, Klarlund M, Thomsen HS, Østergaard M. Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis. Arthritis Rheum. 2003;48:955–62.
17. Backhaus M. Ultrasound and structural changes in inflammatory arthritis: synovitis and tenosynovitis. Ann N Y Acad Sci. 2009;1154:139–51.

18. Kita J, Tamai M, Arima K, et al. Delayed treatment with tumor necrosis factor inhibitors in incomplete responders to synthetic disease-modifying anti-rheumatic drugs shows an excellent effect in patients with very early rheumatoid arthritis with poor prognosis factors. Mod Rheumatol. 2011 (Epub ahead of print).
19. Kita J, Tamai M, Arima K, et al. Treatment discontinuation in patients with very early rheumatoid arthritis in sustained simplified disease activity index remission after synthetic disease-modifying anti-rheumatic drug administration. Mod Rheumatol. 2011 (Epub ahead of print).
20. van der Linden MP, Knevel R, Huizinga TW, van der Helm-van Mil AH. Classification of rheumatoid arthritis: comparison of the 1987 American College of Rheumatology criteria and the 2010 American College of Rheumatology/European League Against Rheumatism criteria. Arthritis Rheum. 2011;63:37–42.
21. Kaneko Y, Kuwana M, Kameda H, Takeuchi T. Sensitivity and specificity of 2010 rheumatoid arthritis classification criteria. Rheumatology (Oxford). 2011;50:1268–74.
22. Britsemmer K, Ursum J, Gerritsen M, van Tuyl L, van Schaardenburg D. Validation of the 2010 ACR/EULAR classification criteria for rheumatoid arthritis: slight improvement over the 1987 ACR criteria. Ann Rheum Dis. 2011;70:1468–70.
23. Cader MZ, Filer A, Hazlehurst J, de Pablo P, Buckley CD, Raza K. Performance of the 2010 ACR/EULAR criteria for rheumatoid arthritis: comparison with 1987 ACR criteria in a very early synovitis cohort. Ann Rheum Dis. 2011;70:949–55.
24. Joshua F, Edmonds J, Lassere M. Power Doppler ultrasound in musculoskeletal disease: a systematic review. Semin Arthritis Rheum. 2006;36:99–108.
25. Freeston JE, Wakefield RJ, Conaghan PG, Hensor EM, Stewart SP, Emery P. A diagnostic algorithm for persistence of very early inflammatory arthritis: the utility of power Doppler ultrasound when added to conventional assessment tools. Ann Rheum Dis. 2010;69:417–9.
26. D'Agostino MA, Aegerter P, Bechara K, Salliot C, Judet O, Chimenti MS, et al. How to diagnose spondyloarthritis early? Accuracy of peripheral enthesitis detection by power Doppler ultrasonography. Ann Rheum Dis. 2011;70:1433–40.

Use of data from multiple registries in studying biologic discontinuation: challenges and opportunities

K. Yoshida[1,2], H. Radner[1,3], A. Kavanaugh[4], Y.-K. Sung[1,5], S.-C. Bae[5], M. Kishimoto[6], K. Matsui[2], M. Okada[6], S. Tohma[7], M.E. Weinblatt[1], D.H. Solomon[1]

[1]*Division of Rheumatology, Brigham and Women's Hospital, Boston, MA, USA;*
[2]*Dept. of Rheumatology, Kameda Medical Center, Kamogawa, Japan;*
[3]*Dept. of Medicine III, Division of Rheumatology, Medical University Vienna, Vienna, Austria;*
[4]*Division of Rheumatology, University of California San Diego, La Jolla, California, United States;*
[5]*Dept. of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, South Korea;*
[6]*Division of Allergy and Rheumatology, St. Luke's International Hospital, Chuo-ku, Tokyo, Japan;*
[7]*Dept. of Rheumatology, Clinical Research Center for Allergy and Rheumatology, Sagamihara National Hospital, Sagamihara, Japan.*

Kazuki Yoshida, MD
Helga Radner, MD.
Arthur Kavanaugh, MD
Yoon-Kyoung Sung, MD, MPH, PhD
Sang-Cheol Bae, MD, MPH, PhD
Mitsumasa Kishimoto, MD, PhD
Kazuo Matsui, MD
Masato Okada, MD
Shigeto Tohma, MD, PhD
Michael E. Weinblatt, MD
Daniel H. Solomon, MD, MPH

Please address correspondence to:
Kazuki Yoshida, MD,
Division of Rheumatology,
Brigham and Women's Hospital,
75 Francis Street,
Boston, MA 02115, USA
E-mail: kazukiyoshida@mail.harvard.edu

Received on August 5, 2013; accepted in revised form on August 22, 2013.

Clin Exp Rheumatol 2013; 31 (Suppl. 78): Sxx-Sxx.

Key words: rheumatoid arthritis, anti-rheumatic agents, remission, discontinuation

Competing interests: see page 4.

ABSTRACT

Many studies have been conducted concerning discontinuation of biologic disease-modifying anti-rheumatic drugs (DMARD), but mainly in trial settings which result in limited generalisability. Registry studies can complement the current literature of biologic DMARD discontinuation by providing more generalisable information. However, it may be necessary to combine registries to increase power and provide more diverse patient populations. This increased power could provide us information about risk and benefits of discontinuing biologic DMARD in typical clinical practice. However, use of multiple registries is not without challenges. In this review, we discuss the challenges to combining data across multiple registries, focusing on biologic discontinuation as an example. Challenges include: 1) generalisability of each registry; 2) new versus prevalent users designs; 3) outcome definitions; 4) different health care systems; 5) different follow up intervals; and 6) data harmonisation. The first three apply to each registry, and the last three apply to combining multiple registries. This review describes these challenges, corresponding solutions, and potential future opportunities.

Prior biologic discontinuation studies

Many studies concerning discontinuation of biologic disease-modifying anti-rheumatic drugs (DMARD) in rheumatoid arthritis (RA) patients have been conducted to date (1, 2). In our previous review summarising 14 such studies (1), we classified them into three groups: (a) randomised controlled trials, in which discontinuation and continuation strategies were randomly assigned; (b) single arm prospective studies of discontinuation, in which patients were prospectively recruited for biologic discontinuation; and (c) long-term extension of efficacy trials, in which patients who discontinued biologic DMARDs were observed. Many of these studies were conducted in rather specialised settings that may not be fully representative of typical clinical practice. In addition, patients from clinical trials can differ in important ways from general clinic populations, such as disease activity and presence of comorbidities that may impact the success of discontinuation of therapy. The current evidence would be complemented with information gained from more generalisable sources, such as registries.

Definition of registries

One paper (3) stated that the term *registry* is often loosely used to mean "any database storing clinical information collected as a byproduct of patient care", and defined a medical data registry as "system functioning in patient management or research, in which a standardised and complete dataset including associated follow-up is prospectively and systematically collected for a group of patients with a common disease or therapeutic intervention".

In the "User's Guide" published by Agency for Healthcare Research and Quality (AHRQ) (4), registry was defined as "an organised system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes". Others have defined registries as "longitudinal observational cohorts, typically

prospective, which enroll patients with a specific purpose; it could either be drug- or disease-based, or both" (5).

For practical purposes, we define a *registry* as a longitudinal follow-up database consisting of clinical data collected as a byproduct of usual care. By "usual care", we mean typical clinical practice where treatment decisions are made by patients and physicians rather than predefined study protocols.

Registries enroll subjects based on a particular disease, condition, or exposure (4), Product registries, health services registries, disease or condition registries, and combinations of these are examples. In the case of biologic discontinuation studies, both biologic DMARD registries (*product* registries) and RA registries (*disease* registries), preferrably enrolling consecutive patients (6), can be utilised.

Studies combining multiple registries

The introduction of biologic DMARDs has led to increased interest in use of registries in studying real-life long-term effectiveness and safety of these agents (5), since randomised controlled efficacy trials do not provide sufficient answers to these questions due to the restrictive nature of their inclusion criteria and follow-up (7-9). Combining multiple databases together can improve power and has been used in studying rare diseases, rare exposures, and rare outcomes; for example, a rare neurodevelopmental disorder (10) and rare environmental exposures, such as infrequently applied pesticides can be well studied in combined registries (11). In rheumatology, the European Collaborative Registries for the Evaluation of Rituximab in rheumatoid arthritis (CERERRA) initiative for rituximab use in daily practice in Europe is an example (12). This study addressed the effectiveness of rituximab using 10 European cohorts, resulting in a large patient sample (n = 2019), which would not have been possible in any one of these registries or countries alone. Comparing across registries may also be used to reveal regional or national differences in diseases and treatment practice. Similarly, the increased power from multiple registries is useful for biologic DMARD discontinuation studies because the numbers of eligible patients, *i.e.* those who have discontinued biologic DMARDs in good disease control, are expected to be fewer in typical practice than trials in which discontinuation is systematically assigned. Nevertheless, when using data from combined registries, we are faced with several challenges; some of them are challenges to all registries (challenges 1–3 below) and some are methodological complexities specific to combining registries (challenges 4–6 below).

Challenge 1.
Generalisability of each registry

Generalisability as a particular strength of registry studies is dependent on the source population from which the registry enrolls subjects and how these subjects are enrolled. If the source population is not the typical RA patient on a biologic DMARD results will not be generalisable. The representativeness of the biologic DMARD users in a given registry is dependent on how these subjects compare to the population of biologic DMARD users in the country. Some registries contain (almost) all biologic DMARD users in a given country, for example the British Society for Rheumatology Biologics Register (BSRBR) (13). Registries that are not directly required by the health care system usually enroll patients from one or several participating institutions (or practices) and may capture patients associated with rheumatologists involved with research, not representative of all rheumatology practice. Unless the sample of patients is truly random, there is the potential for bias in the acquisition of patients that could impact the results. These points must be examined before claiming the generalisability of information obtained from the registry. To ensure generalisability, nationally (or internationally) representative registries that enroll wide range of patients at multiple centres are preferable (14-16).

Challenge 2.
New users *vs.* prevalent users designs

When studying comparative effectiveness of two active agents, choosing new users of both agents is important for ensuring exchangeability (17). Biologic DMARD registries are usually comprised of new users of biologic DMARDs, as the UK's BSRBR (13) or postmarketing surveillance registries in Japan (18-21). In contrast, disease-based RA registries may enroll prevalent RA cases already using biologic DMARDs. If the enrolment date of patients is after the initiation of a biologic DMARD, information prior to initiation is often incomplete.

However, this is less problematic for discontinuation studies where the study index date is typically defined as the time at discontinuation of therapy. Sensitivity analysis comparing new users only design to both new users and prevalent users design is recommended if there is a suspicion that different baselines before use of biologic DMARDs might exist among prevalent users versus new users.

Challenge 3.
Outcome definition

Studying outcomes that are not directly related to the primary reason for which the registry was started can present challenges, as endpoints may not be collected in a direct manner. Biologic DMARD discontinuation study is usually not the primary reason for registries and thus the outcome determination may not be ideal. The definition of "failure of discontinuation" (the outcome of interest in biologic DMARD discontinuation studies) has not been standardised in previous non-registry studies. In our previous review, we examined how "failure of biologic DMARD discontinuation" was defined across various studies (1): all studies used increase in disease activity, and many included reuse of biologic DMARDs for the definition of failure in discontinuing biological DMARDs. Moreover, the thresholds of increase in disease activity varied, and there was no consensus on whether intensification of non-biologic DMARDs or glucocorticoids should constitute failure.

In a registry study, long intervals between study visits might obscure an increase of disease activity in between visits, thus, "failure of discontinuation"

could be missed by criteria that only use disease activity and biologic DMARD reuse (Fig. 1). This is primarily why intensification in non-biologic DMARDs and glucocorticoids should be regarded as a sign of failure. The thresholds for failure should be determined in such a way that they are comparable across registries if deemed feasible after considering national differences in disease control (22).

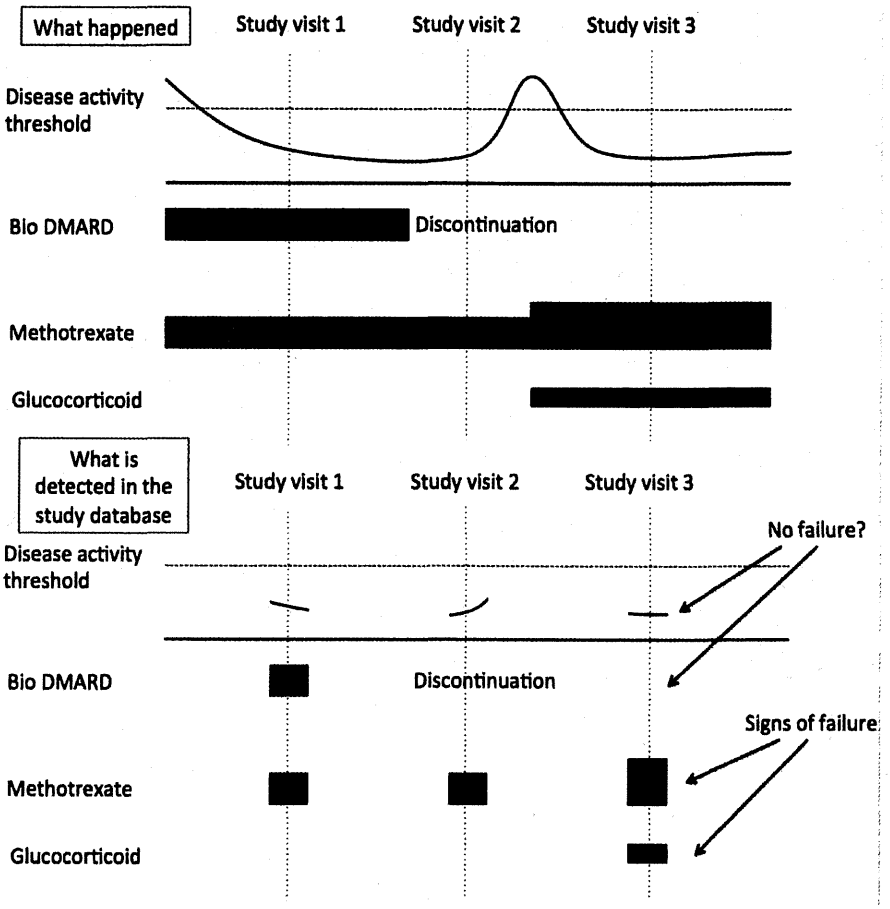

Fig. 1. In a registry study, not all clinical visits are necessarily captured as part of the study protocol. In this example, biologic discontinuation is detected at study visit 2. The definition with disease activity threshold violation and biologic DMARD reuse at study visits misses the failure that occurred between study visits 2 and 3. By including non-biologic treatment changes detected at study visit 3 as a failure criterion, the outcome of interest will be indirectly captured.

Challenge 4.
Health care system differences

Due to the rapid development of new biologic DMARDs and their high cost, different countries have different biologic DMARDs approved (drug lag) and also have different policies regarding biologic DMARDs use and reimbursement, resulting in varying access to biologic DMARDs (23). In some countries, biologic DMARDs are prescribed at the discretion of physicians and commonly used, for example, in the US, 43% of RA patients received biologic DMARDs in one study (24). Biologic prescription practice is more restricted by practice guidelines that are required by health insurance providers, in some European and Asian countries for example (16, 23, 25). In these settings, the users of biologic DMARDs are expected to differ. In more restrictive prescription setting, there may be fewer early RA patients compared to long-standing RA patients. Such patients may have different patterns of treatment response both before and after discontinuation of biologic DMARDs. Also, in some countries, including the USA, patients may pay directly for a portion of their drug costs (*i.e.* co-payment). This could impact their decision on whether and when to stop particular therapies. Finally, in the not-too-distant future, discontinuation of biologics will likely become incorporated into treatment recommendations and individual country guidelines, which will also have an effect on the data.

This could potentially cause a problem if pooling data, but it may also be possible to take advantage of these differences to compare different treatment strategies. Thus, if registries from different health care systems are to be studied, the guidelines regarding biologic DMARD prescription should be assessed for expected prescription pattern differences. If differences are sufficiently substantial so that biologic DMARD users in these registries are very different, direct pooling of data should be avoided. The focus should be a cross-registry comparison, which can potentially provide interesting "natural experiments" in different treatment strategies. Individual-patient level data meta-analysis using random effects models is another possibility in such a situation (26).

Challenge 5.
Different follow-up intervals

The intervals of follow-ups usually are different among registries. For example, some registries may have information on every physician visit while another may collect information on a less frequent basis (*i.e.* every 6 months, annually, etc). Therefore, when combining data across multiple registries, it is very likely that assessment timepoints vary. This can be further complicated by missing values, giving rise to unbalanced data even within each registry.

To overcome this, one can use the "least common denominator" approach by simply focusing on the longest of all available intervals; however, much data would be discarded through such an approach. A better approach may be to use analytical methods that can accommodate different intervals for individual patients, such as (generalised) linear mixed effect models for repeatedly measured binary outcomes (27) or extended Cox models for time-to-event outcomes (28), which can accommodate time-varying variables.

Table. Challenges and solutions for (multiple) registry studies of biologic DMARD discontinuation.

Challenge	Solution
Challenges faced by all registries	
1) Generalisability of registries	Check if the source population for the registry is typical population of biologic DMARD users.
2) New *vs.* prevalent users of biologic DMARDs	Prevalent users can be included as long as they are new to discontinuation. Sensitivity analysis is recommended.
3) Outcome definition	Changes in non-biologic DMARDs should be incorporated in a composite "failure of discontinuation" definition.
Methodological complexities specific to combining registries	
4) Different health care system	If registries are from very different health care systems with different utilisation patterns of biologic DMARDs, comparison rather than pooling is preferred.
5) Different follow up intervals	Analysis methods that can accommodate "unbalanced" longitudinal data with varying follow up intervals should be used.
6) Data harmonisation	Variables should be matched as individual variables (swollen joint count, etc.) rather than composite variables such as disease activity scores, if possible.

Challenge 6. Data harmonisation

One purpose of a biologic discontinuation study is to attempt to identify variables that can predict continued disease control after cessation of biologic DMARD. If we can predict which patients will be able to discontinue biologic DMARDs successfully, drug exposure and associated risks and costs might be reduced. Development of such a prediction model across registries would require matching of variables that have been measured differently. This process is often called "data harmonisation"; there is debate about requirements for data harmonisation (29).

The most robust way of harmonising variables is to design, prospectively, multiple registries with harmonisation in mind. This so-called "stringent approach to harmonisation" (29) requires collaboration before registries are started and would be very time-consuming. This will result in higher quality data, but may cancel out one benefit of registry studies, namely prompt access to data that can be utilised quickly.

The "flexible approach to harmonisation" (29), on the other hand, is an effort to match variables in previously collected data. For example, in the case of biologic DMARD discontinuation studies, one element of the composite outcome (see Challenge 3) is the disease activity. Different disease activity measures have been used in different registries, for example, Disease Activity Score 28 with erythrocyte sedimentation rate (DAS28-ESR) (15), DAS28 with C-reactive protein (DAS28-CRP) (30), and Clinical Disease Activity Index (CDAI) (14). These measures correlate well in biologic DMARD users (31), but the thresholds for remission and low disease activity have different characteristics depending on the measures used (32, 33). To overcome this challenge, the collection of each component of the composite scores (such as joint counts) might be useful to recalculate a desired composite score. If harmonising scores are difficult, one could also consider harmonising the disease activity categories (remission, low disease activity, etc.) or treatment response categories (34).

Discussion and future direction

Combining data from multiple registries may be useful to study outcomes as biological DMARD discontinuation. Nevertheless several potential challenges must be addressed, as we discussed above (summarised in Table I). Registry studies can give us insights into biologic DMARD discontinuation patterns and outcome in real-life practice settings, which could provide evidence that complements currently available evidence from trials.

Use of individual-patient level data (IPD) when combining multiple registries has certain strengths compared to aggregate patient data (APD) meta-analysis (26), which collects published studies and combine the aggregated results. Firstly, use of IPD enables more careful examination of the heterogeneity of the subjects. Secondly, it allows better adjustment for baseline differences using similar sets of variables across registries. Thirdly, some variable heterogeneity can be adjusted for by redefining variables using raw IPD. It is also possible to use random-effects meta-analytic techniques on this type of data. This approach is beneficial when the sample sizes of data sources are very different because small data sources may be overshadowed by larger ones if datasets are simply combined.

The use of multiple registries is not limited to biologic DMARD discontinuation studies. Research questions that require generalisable clinical information and large sample sizes can potentially gain advantages from combining datasets. Potential examples include studies of rare exposures, such as very newly introduced medications, or rare outcomes such as certain toxicities. In addition, cross-national comparisons using multiple registries can answer interesting health services questions, as well as providing natural experiments through treatment variation.

In conclusion, the use of multiple registry data studies could offer substantial opportunities for studying biologic DMARD discontinuation and beyond.

Funding

K. Yoshida's time at Brigham and Women's Hospital is funded by a scholarship from Kameda Medical Cente;. D.H. Solomon receives salary support from NIH-K24AR055989; S.C. Bae was supported by a grant from the Korea Healthcare Technology R&D Project, Ministry for Health and Welfare, Republic of Korea (A102065).

Competing interests

K. Yoshida has received honoraria and served as an instructor at a musculo-

skeletal ultrasonography workshop sponsored by Abbott Japan;
A. Kavanaugh has conducted sponsored clinical research for Abbott, Janssen, Amgen, BMS, Roche and UCB;
S.C. Bae was supported by a grant of the Korea Healthcare technology R&D Project, Ministry for Health and Welfare, Republic of Korea (A102065) and has received research grants from Abbott Ltd., Bristol Myers Squibb Pharmaceutical Ltd., Eisai Inc., GlaxoSmithKline Ltd., MSD Ltd. and Pfizer Inc.;
M.E. Weinblatt has received grant support from Bristol-Myers Squibb and serves in consultant roles for Amgen, Abbott, Janssen, Bristol-Myers Squibb, Roche, and UCB;
M. Kishimoto has received speaking fees, and/or honoraria from Santen Pharmaceutical, Mitsubishi Tanabe Pharma, Pfizer, and Abbott Japan;
M. Okada has received speaking fees, and/or honoraria from Santen Pharmaceutical, Mitsubishi Tanabe Pharma, Pfizer, and Abbott Japan;
S. Tohma has received research grants from Pfizer Japan Inc., Eisai Co., Ltd, and Chugai Pharmaceutical Co., Ltd.;
D.H Solomon receives salary support from institutional research grants from Eli Lilly, Amgen, and CORRONA, and also receives royalties from UpToDate, and serves in unpaid roles in studies funded by Pfizer and Novartis;
the other co-authors have declared no competing interests.

References

1. YOSHIDA K, SUNG Y-K, KAVANAUGH A *et al.*: Biologic discontinuation studies: a systematic review of methods. *Ann Rheum Dis* 2013.
2. NISHIMOTO N, AMANO K, HIRABAYASHI Y *et al.*: Drug free REmission/low disease activity after cessation of tocilizumab (Actemra) Monotherapy (DREAM) study. *Mod Rheumatol* 2013 May 3. [Epub ahead of print]
3. DROLET BC, JOHNSON KB: Categorizing the world of registries. *J Biomed Inform* 2008; 41: 1009-20.
4. GLIKLICH RE, DREYER NA eds. Registries for Evaluating Patient Outcomes: A User's Guide. 2nd ed. Rockville (MD): Agency for Healthcare Research and Quality (US); 2010.
5. ELKAYAM O, PAVELKA K: Biologic registries in rheumatology: Lessons learned and expectations for the future. *Autoimmunity Reviews* 2012; 12: 329-36.
6. MOSES LE: The series of consecutive cases as a device for assessing outcomes of intervention. *N Engl J Med* 1984; 311: 705-10.
7. PINCUS T, STEIN CM: Why randomized controlled clinical trials do not depict accurately long-term outcomes in rheumatoid arthritis: some explanations and suggestions for future studies. *Clin Exp Rheumatol* 1997; 15 (Suppl. 17): S27-38.
8. PINCUS T: Limitations of randomized clinical trials in chronic diseases: explanations and recommendations. *Adv Mind Body Med* 2002; 18: 14-21.
9. KREMER JM, GIBOFSKY A, GREENBERG JD: The role of drug and disease registries in rheumatic disease epidemiology. *Curr Opin Rheumatol* 2008; 20: 123-30.
10. GRILLO E, VILLARD L, CLARKE A *et al.*: Rett networked database: an integrated clinical and genetic network of Rett syndrome databases. *Hum Mutat* 2012; 33: 1031-6.
11. LEON ME, BEANE FREEMAN LE, DOUWES J *et al.*: AGRICOH: a consortium of agricultural cohorts. *Int J Environ Res Public Health* 2011; 8: 1341-57.
12. CHATZIDIONYSIOU K, LIE E, NASONOV E *et al.*: Effectiveness of disease-modifying antirheumatic drug co-therapy with methotrexate and leflunomide in rituximab-treated rheumatoid arthritis patients: results of a 1-year follow-up study from the CERERRA collaboration. *Ann Rheum Dis* 2012; 71: 374-7.
13. SILMAN A, SYMMONS D, SCOTT DGI, GRIFFITHS I: British Society for Rheumatology Biologics Register. *Ann Rheum Dis* 2003; 62 (Suppl. 2): ii28-29.
14. KREMER JM: The CORRONA database. *Clin Exp Rheumatol* 2005; 23: S172-7.
15. YAMANAKA H, TOHMA S: Potential impact of observational cohort studies in Japan on rheumatoid arthritis research and practice. *Mod Rheumatol* 2006; 16: 75-6.
16. SUNG Y-K, CHO S-K, CHOI C-B *et al.*: Korean Observational Study Network for Arthritis (KORONA): establishment of a prospective multicenter cohort for rheumatoid arthritis in South Korea. *Semin Arthritis Rheum* 2012; 41: 745-51.
17. RAY WA: Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003;158: 915-20.
18. TAKEUCHI T, TATSUKI Y, NOGAMI Y *et al.*: Postmarketing surveillance of the safety profile of infliximab in 5000 Japanese patients with rheumatoid arthritis. *Ann Rheum Dis* 2008; 67: 189-94.
19. TAKEUCHI T, HARIGAI M, INOKUMA S *et al.*: Postmarketing surveillance of tocilizumab for rheumatoid arthritis in japan interim analysis of 6424 patients. *Ann Rheum Dis* 2011; 70 (Suppl. 3): 610.
20. KOIKE T, HARIGAI M, INOKUMA S *et al.*: Postmarketing surveillance of safety and effectiveness of etanercept in Japanese patients with rheumatoid arthritis. *Mod Rheumatol* 2011; 21: 343-51.
21. KOIKE T, HARIGAI M, ISHIGURO N *et al.*: Safety and effectiveness of adalimumab in Japanese rheumatoid arthritis patients: postmarketing surveillance report of the first 3,000 patients. *Mod Rheumatol* 2012; 22: 498-508.
22. SOKKA T, KAUTIAINEN H, PINCUS T *et al.*: Disparities in rheumatoid arthritis disease activity according to gross domestic product in 25 countries in the QUEST-RA database. *Ann Rheum Dis* 2009; 68: 1666-72.
23. PUTRIK P, RAMIRO S, KVIEN TK *et al.*: Inequities in access to biologic and synthetic DMARDs across 46 European countries. *Ann Rheum Dis* 2013.
24. KIM SC, YELIN E, TONNER C, SOLOMON DH: Changes in use of disease modifying antirheumatic drugs for rheumatoid Arthritis in the U.S. for the period 1983-2009. *Arthritis Care Res* (Hoboken) 2013.
25. ORLEWSKA E, ANCUTA I, ANIC B *et al.*: Access to biologic treatment for rheumatoid arthritis in Central and Eastern European (CEE) countries. *Med Sci Monit* 2011; 17: SR1-13.
26. LYMAN GH, KUDERER NM: The strengths and limitations of meta-analyses based on aggregate data. *BMC Med Res Methodol* 2005; 5: 14.
27. FITZMAURICE GM, LAIRD WARE: Applied Longitudinal Analysis. Hoboken, NJ: Wiley; 2011.
28. KLEINBAUM DG, KLEIN M: Survival analysis. New York [etc.]: Springer; 2012.
29. FORTIER I, DOIRON D, BURTON P, RAINA P: Invited commentary: consolidating data harmonization--how to obtain quality and applicability? *Am J Epidemiol* 2011; 174: 261-4; author reply 265–266.
30. IANNACCONE CK, LEE YC, CUI J *et al.*: Using genetic and clinical data to understand response to disease-modifying anti-rheumatic drug therapy: data from the Brigham and Women's Hospital Rheumatoid Arthritis Sequential Study. *Rheumatology* (Oxford) 2011; 50: 40-6.
31. NISHIMOTO N, TAKAGI N: Assessment of the validity of the 28-joint disease activity score using erythrocyte sedimentation rate (DAS28-ESR) as a disease activity index of rheumatoid arthritis in the efficacy evaluation of 24-week treatment with tocilizumab: subanalysis of the SATORI study. *Mod Rheumatol* 2010; 20: 539-47.
32. INOUE E, YAMANAKA H, HARA M, TOMATSU T, KAMATANI N: Comparison of Disease Activity Score (DAS)28- erythrocyte sedimentation rate and DAS28- C-reactive protein threshold values. *Ann Rheum Dis* 2007; 66: 407-9.
33. LEE YC, CUI J, LU B *et al.*: Pain persists in DAS28 rheumatoid arthritis remission but not in ACR/EULAR remission: a longitudinal observational study. *Arthritis Res Ther* 2011; 13: R83.
34. ALETAHA D, MARTINEZ-AVILA J, KVIEN TK, SMOLEN JS: Definition of treatment response in rheumatoid arthritis based on the simplified and the clinical disease activity index. *Ann Rheum Dis* 2012; 71: 1190-6.

EXTENDED REPORT

A comparative effectiveness study of adalimumab, etanercept and infliximab in biologically naive and switched rheumatoid arthritis patients: results from the US CORRONA registry

Jeffrey D Greenberg,[1] George Reed,[2] Dennis Decktor,[3] Leslie Harrold,[2] Daniel Furst,[4] Allan Gibofsky,[5] Ralph DeHoratius,[3] Mitsumasa Kishimoto,[1] Joel M Kremer,[6] on behalf of the CORRONA Investigators

► Additional supplementary figure and tables are published online only. To view the files please visit the journal online (http://ard.bmj.com/content/71/7.toc).

[1]New York University Hopsital for Joint Diseases, New York, New York, USA
[2]University of Massachusetts Medical School, Worcester, Massachusetts, USA
[3]Janssen Services LLC, Horsham, Pennsylvania, USA
[4]University of California, Los Angeles, California, USA
[5]Hospital for Special Surgery, New York, New York, USA
[6]Albany Medical College, Albany, New York, USA

Correspondence to
Jeffrey D Greenberg, New York University Hospital for Joint Diseases, Department of Rheumatology, 301 East 17th Street, suite 1410, New York, NY 10003, USA; jeffrey.greenberg@nyumc.org

Received 25 January 2011
Accepted 29 November 2011
Published Online First 30 January 2012

ABSTRACT

Purpose To compare the effectiveness of anti-tumour necrosis factor (TNF) agents in biologically naive and 'switched' rheumatoid arthritis (RA) patients.

Methods RA patients enrolled in the CORRONA registry newly prescribed adalimumab (n=874), etanercept (n=640), or infliximab (n=728) were stratified based on previous anti-TNF use. Clinical effectiveness at 6, 12 and 24 months was examined using the modified American College of Rheumatology response criteria (mACR20/50/70) and achievement of remission (28-joint disease activity score (DAS28) and clinical disease activity index (CDAI)) in unadjusted and adjusted analyses. The persistence of anti-TNF treatment was examined using Cox proportional hazard models.

Results Among 2242 patients (1475 biologically naive, 767 switchers), mACR20, 50 and 70 responses were similar (p>0.05) for adalimumab, etanercept and infliximab at all time points, as were rates of CDAI and DAS28 remission (p>0.05). Response and remission outcomes were consistently inferior for switched versus biologically naive patients. The adjusted OR for achieving an mACR20 response was 0.54 (95% CI 0.38 to 0.76) in first-time switchers and 0.42 (95% CI 0.23 to 0.78) in second-time switchers versus biologically naive patients at 6 months. The adjusted OR for achieving DAS28 remission were 0.29 (95% CI 0.15 to 0.58) for first-time switchers and 0.26 (95% CI 0.08 to 0.84) for second-time switchers. Persistence was higher in biologically naive patients, for whom persistence was highest with infliximab.

Conclusions No differences in rates of drug response or remission were observed among the three anti-TNF. Infliximab was associated with greater persistence in biologically naive patients. Response, remission and persistence outcomes were diminished for patients who switched anti-TNF.

Over the past decade, anti-tumour necrosis factor (TNF) therapies have become the most frequently prescribed class of biological agents for the treatment of rheumatoid arthritis (RA) in the USA and Europe. Currently, there are five anti-TNF agents approved by the European Medicines Agency and the US Food and Drug Administration, with varying structures, dosing and pharmacokinetics. Despite these differences, they all block TNF, and two randomised clinical trial (RCT) meta-analyses of three commonly prescribed anti-TNF (adalimumab, etanercept and infliximab) concluded that the three anti-TNF demonstrated comparable efficacy.[1 2] However, these meta-analyses have been criticised, and their findings conflict with the results reported in two European registry studies demonstrating that adalimumab and etanercept users have better clinical responses than infliximab users.[3 4] Those reports originated from European countries with more restricted access to biological agents and dosage restrictions.

An important caveat to the application of anti-TNF RCT results to RA patients in the clinic is that the vast majority of the RCTs were conducted in biologically naive patients, particularly in those without a previous history of anti-TNF treatment. However, intraclass switching of anti-TNF agents is common in clinical practices in Europe and the USA.[5–8] Currently, there is inadequate evidence regarding the benefits of this strategy. As a result, switching patients to a different anti-TNF agent is restricted in certain European countries. Comparative effectiveness research using observational data sources has gained broader support in Europe and the USA across clinical disease areas.[9–11]

Comparative effectiveness studies using observational data from registries represent a promising alternative to RCT for comparing interventions and therapies between biologically naive patients and patients who switch anti-TNF.[11] This is important because rheumatologists in the USA and many European countries prescribe anti-TNF agents to RA patients with markedly lower disease activity than RCT populations.[12–14] Given that comparative effectiveness data for US-based cohorts are lacking, the aim of the present study was to compare the clinical effectiveness of specific anti-TNF agents and the strategy of intraclass switching in a large US cohort of RA patients using the Consortium of Rheumatology Researchers of North America (CORRONA) registry. In particular, we sought to compare composite rates of drug response and remission outcomes as well as the persistence of anti-TNF treatment over a 2-year period.

Table 1 Baseline characteristics of the study cohort stratified by previous exposure to anti-TNF and newly prescribed anti-TNF agent

	Biologic naive				First-time switchers*			
Characteristics	**ADA N=460**	**ETA N=480**	**INF N=535**	**p Value†**	**ADA N=311**	**ETA N=139**	**INF N=166**	**p Value†**
Demographics								
Women	78%	76%	72%	0.06	82%	79%	82%	0.72
Age (years)	55±12	54±13	61±13	<0.001	56±13	56±13	56±12	0.83
Healthcare coverage‡								
Private insurance	78%	81%	72%	0.04	79%	81%	74%	0.39
Medicare	27%	24%	45%	<0.001	34%	35%	34%	0.98
Medicaid	7%	9%	6%	0.30	9%	6%	5%	0.23
Clinical								
Duration of RA (years)	8.9±9.5	8.8±9.2	9.6±9.9	<0.001	12.7±9.7	10.6±10.0	11.8±9.4	0.09
Tender joint count	7.1±7.1	6.4±6.2	6.3±6.7	0.11	7.6±7.1	6.6±6.8	6.3±6.9	0.14
Swollen joint count	7.6±6.5	6.5±5.8	8.2±6.9	<0.001	6.7±6.3	6.9±6.5	7.4±7.00	0.57
Patient global assessment (0–100 mm)	41.2±27.5	40.1±24.7	38.7±24.9	0.34	44.4±25.3	42.9±27.3	38.7±25.5	0.09
Patient pain assessment (0–100 mm)	43.3±28.0	41.5±24.7	41.5±25.8	0.48	45.7±25.5	46.0±26.0	41.9±24.9	0.26
Physician global assessment (0–100 mm)	36.9±20.5	33.5±20.3	34.4±20.9	0.03	37.3±22.3	33.3±20.5	32.8±22.2	0.05
mHAQ score	0.5±0.5	0.5±0.5	0.4±0.5	0.11	0.6±0.5	0.6±0.5	0.4±0.4	0.01
ESR (mm/h)	25.7±23.3	24.2±19.8	28.2±23.2	0.19	28.9±23.2	28.1±23.5	28.2±22.0	0.96
DAS28	4.49±1.6	4.48±1.4	4.53±1.4	0.91	4.55±1.5	4.39±1.3	4.46 ±.6	0.79
CDAI	22.3±13.7	20.2±12.3	22.0±13.4	0.04	22.4±14.3	21.1±13.4	20.6±13.9	0.43
Disease activity per CDAI				0.15				0.69
High (>21)	21	22	22		23	21	25	
Moderate (>5–≤21)	37	41	34		33	39	36	
Low (>2.2–≤5)	42	37	44		44	40	39	
BMI	29.2±7.1	29.5±7.6	29.6±7.5	0.67	28.6±7.3	30.5±7.7	29.2±6.6	0.04
Disabled	11	11	10	0.79	24	12	17	0.01
Medication at entry								
Prednisone	35	33	33	0.80	35	35	33	0.81
Methotrexate	68	61	68	0.05	53	63	60	0.13
Methotrexate dose								
≤7.5 mg	22%	17%	28%		24%	15%	23%	
10–17.5 mg	43%	49%	38%		37%	48%	36%	
≥20 mg	35%	34%	35%		40%	36%	41%	
No of previous DMARD	0.7±1.0	0.7±1.0	0.7±1.0	0.73	2.1±1.4	1.5±1.3	1.8±1.3	<0.001

Data shown are percentages of patients or mean±SD.
*Second-time switchers, including 103 switched to adalimumab, 21 to etanercept and 27 to infliximab, are not included due to relatively small sample size.
†p Values are derived from analysis of variance for continuous measures and Fisher's exact test for dichotomous variables.
‡Categories are not mutually exclusive.
ADA, adalimumab; BMI, body mass index; CDAI, clinical disease activity index; DAS28, disease activity score employing 28-joint count; DMARD, disease-modifying antirheumatic drugs; ESR, erythrocyte sedimentation rate; ETA, etanercept; INF, infliximab; mHAQ, modified health assessment questionnaire; RA, rheumatoid arthritis; TNF, tumour necrosis factor.

METHODS

Data source

The CORRONA registry is an independent prospective observational cohort of patients with arthritis who are enrolled by participating rheumatologists in both academic and private practice sites. As detailed previously,[15 16] CORRONA is governed by a board of academically affiliated US rheumatologists. CORRONA has no governance or ownership ties to the pharmaceutical industry. CORRONA receives funding from multiple pharmaceutical manufacturers to support the registry.

CORRONA data collection began in 2002; data collected to 11 March 2008 are included in the current analyses. Up to 2008, there were 83 sites across 33 states in the USA, and approximately 200 rheumatologists have enrolled a total of 19 902 patients, including 16 696 with RA. Approximately 22% of the sites were academic sites and 78% were private sites. The geographical distribution of patients in the registry across the USA was the northeast region 34%, midwest region 24%, south region 28% and west region 14%. Patients were enrolled into the CORRONA registry at the time of a routine clinic visit. Enrolment into the CORRONA registry remains active. Both patient and physician questionnaires are filled out during routine clinical encounters. Completed questionnaires are faxed or mailed to a central processing site. Approvals for data collection and analyses were obtained for academic and private practice sites from local and central institutional review boards, respectively.

Study population

Among the 16 696 patients with RA enrolled in the CORRONA registry, 2530 were newly prescribed an anti-TNF agent with at least one follow-up visit between 4 February 2002 and 11 March 2008. No disease activity or comorbidity exclusion criteria were required for RA patients enrolled into the consortium registry. For the purposes of this study, the 162 RA patients in remission at baseline, defined by a clinical disease activity index (CDAI)[17] score of 2.8 or less or a disease activity score based on 28 joints (DAS28) and erythrocyte sedimentation rate (ESR) less than 2.6 were excluded from the study population. Patients with a previous history of the use of a non-TNF agent (N=126) were also excluded, resulting in 2242 RA patients included in this analysis. Among these 2242 patients, 1475 were biologically naive, 616 were first switchers and 151 were switching to their second or more biological agent. A flowchart describing the study population in greater

Table 2 Crude response and remission rates at 6 and 12 months among adalimumab, etanercept and infliximab users in those who were biologically naive

	6 Months			12 Months		
	INF	ADA	ETA	INF	ADA	ETA
mACR response						
No of patients	230	235	222	182	190	178
mACR 20						
Responders (%)	26.5	30.6	37.4	26.9	26.8	31.5
Adjusted OR*	1.00	0.95 (0.60–1.50)	1.37 (0.94–1.99)	1.00	0.96 (0.56, 1.64)	1.35 (0.84, 2.18)
mACR50						
Responders (%)	14.3	19.6	26.6	20.3	17.4	20.8
Adjusted OR*	1.00	1.03 (0.52, 2.01)	1.75 (0.99, 3.09)	1.00	0.72 (0.46–1.13)	1.03 (0.62–1.70)
mACR70						
Responders (%)	9.6	10.2	9.9	12.1	12.1	11.8
Adjusted OR*	1.00	0.76 (0.41–1.42)	0.81 (0.42–1.56)	1.00	0.83 (0.46–1.49)	1.04 (0.61–1.78)
CDAI remission						
No of patients	254	249	242	199	202	189
Responders (%)	15.7	13.7	16.1	17.1	12.9	18.5
Adjusted OR†	1.00	0.83 (0.42–1.63)	1.18 (0.65–2.14)	1.00	0.69 (0.42–1.15)	1.15 (0.61–2.12)
DAS28 remission						
No of patients	103	107	116	71	75	72
Responders (%)	28.2	25.2	28.4	33.8	33.3	37.5
Adjusted OR†	1.00	0.72 (0.48–1.08)	0.95 (0.43–2.08)	1.00	0.89 (0.39–2.00)	1.01 (0.47–2.12)

*Adjusted for duration of RA, joint counts, patient global, age, mHAQ, disability, use of methotrexate and year of initiation.
†Adjusted for duration of RA, baseline disease activity, age, mHAQ, disability, use of methotrexate and year of initiation.
ADA, adalimumab; CDAI, clinical disease activity index; DAS28, disease activity score employing 28-joint count; ETA, etanercept; INF, infliximab; mACR, modified American College of Rheumatology.

detail can be found in supplementary figure S1 (available online only).

Measures and data collection

Data were collected during the study period from physician assessment and patient questionnaires completed during clinical encounters. Patients were followed as frequently as every 3 months. For this dataset, the mean time between visits was 4.7 months and the median time between visits was 3.8 months. Non-biological and biological disease-modifying antirheumatic drugs (DMARD), including anti-TNF agents, were recorded at the time of the clinical encounter. Data elements also documented at the time of a clinical encounter that are relevant to the current analysis included 28 tender and swollen joint counts, physician and patient global assessments of disease activity, patient assessment of pain, the modified health assessment questionnaire (HAQ) assessing physical function and ESR. Across the 2242 patients, data on tender and swollen joint counts were complete in 2210 (98.6%) patients. All components of the CDAI were completed for 2069 (92.3%) patients. Acute phase reactant data were recorded from laboratory tests obtained within 10 days of the clinical encounter, but collection of laboratory data was not mandated by the registry protocol. ESR values were available for 1210 (54%) patients. Insurance data was available for 73.5% of patients. Completeness was high for data required for the CDAI (>92%).

Drug exposure cohorts

Patients initiating adalimumab, etanercept and infliximab were stratified into one of three cohorts. Biologically naive patients initiating an anti-TNF agent were defined as patients with no lifetime history of treatment with anakinra, other anti-TNF agents, abatacept or rituximab. First-time switchers were defined as patients initiating an anti-TNF agent with a history of previous treatment with a different anti-TNF agent. Second-time switchers were defined as patients with a history of previous treatment with two different anti-TNF agents. Within each of the three cohorts, comparisons among the three individual anti-TNF agents (adalimumab, etanercept and infliximab) were performed.

Registry outcomes

Responsiveness to anti-TNF therapy was assessed using the modified American College of Rheumatology (ACR) 20, 50 and 70 response criteria without the requirement for an acute phase reactant to maximise the amount of patient data available for analysis. These measures have been previously defined and validated.[18 19] A modified ACR20 response required a 20% or greater improvement in tender and swollen joint counts, as well as in two or more of the following four ACR response components: physician global assessment, patient global assessment, patient global pain and modified HAQ. The modified ACR50 and 70 responses were calculated using the same criteria, but requiring at least 50% and 70% improvement, respectively. Disease remission outcomes were defined as a DAS28–ESR score less than 2.6[20] and a CDAI score, which does not require an acute phase reactant, of 2.8 or less.[17] Continuation or persistence of treatment with the newly prescribed anti-TNF agent was defined as the duration of time from anti-TNF initiation to discontinuation.

Statistical analysis

Patient clinical and demographic characteristics were compared within the three strata of previous anti-TNF exposure by specific agent. For continuous measures, means and SD were estimated and analysis of variance was used to assess the statistical significance of any differences among the groups. For dichotomous measures, percentages were estimated and Fisher's exact test was used to assess the significance of differences among groups.

For modified ACR20, 50 and 70 response, patients who discontinued the newly prescribed anti-TNF agent were categorised as

Table 3 Secondary analysis of response and remission rates in biologically naive patients: dose escalation imputed as non-response

	6 Months			12 Months		
	INF	ADA	ETA	INF	ADA	ETA
mACR response						
No of patients	230	235	222	182	190	178
mACR 20						
Responders (%)	25.7	28.9	37.4	23.6	25.2	31.5
Adjusted OR	1.00	0.97 (0.63–1.49)	1.50 (1.06–2.13)	1.00	1.03 (0.2–1.70)	1.60 (0.98–1.69)
mACR50						
Responders (%)	13.5	18.7	26.6	18.1	16.8	20.8
Adjusted OR	1.00	1.16 (0.64–2.12)	2.04 (1.24–3.35)	1.00	0.73 (0.47–1.15)	1.10 (0.65–1.86)
mACR70						
Responders (%)	9.1	9.8	9.9	11.0	11.6	11.8
Adjusted OR	1.00	1.04 (0.62–1.75)	1.10 (0.58–2.09)	1.00	1.03 (0.59–1.81)	1.07 (0.62–1.85)
CDAI remission						
No of patients	254	249	242	199	202	189
Responders (%)	15.4	12.9	16.1	16.1	12.4	18.5
Adjusted OR	1.00	0.78 (0.37–1.62)	1.19 (0.64–2.22)	1.00	0.69 (0.43–1.10)	1.20 (0.63–2.27)
DAS28 remission						
No of patients	103	107	116	71	75	72
Responders (%)	26.2	25.2	28.4	28.1	32.0	37.5
Adjusted OR	1.00	1.08 (0.74–1.58)	1.26 (0.68–2.33)	1.00	1.22 (0.64–2.35)	1.57 (0.91–2.72)

ADA, adalimumab; CDAI, clinical disease activity index; DAS28, disease activity score employing 28-joint count; ETA, etanercept; INF, infliximab; mACR, modified American College of Rheumatology.

non-responders (ie, no modified ACR20 50 or 70 response or no DAS28–ESR or CDAI remission) for any study visit after discontinuation, using intention-to-treat analyses with non-responder imputation approach as previously applied.[13 21] Unadjusted ACR response rates were determined at 6, 12 and 24 months following the start of the newly prescribed anti-TNF using 3-month time windows for capturing study visits. Unadjusted and adjusted OR comparing response rates among anti-TNF agents were estimated using multivariable logistic regression models and were reported with estimated 95% CI. Covariates associated with either anti-TNF agent selection or response to treatment were considered as possible confounders and included patient demographics, disease activity and severity measures, previous medication usage, history of comorbidities and years since anti-TNF agent initiation Sensitivity analyses were carried out applying a completer's analysis approach. Similar methodology was employed to assess remission based on the DAS28–ESR and CDAI cut points defined above.

Treatment persistence was estimated using survival analysis methods. Time from initiation to discontinuation of the anti-TNF or to last follow-up visit was estimated based on the initiation visit dates and discontinuation (or last follow-up) dates. Unadjusted Kaplan–Meier survival curves were estimated for each of the three study cohorts, as well as individually for anti-TNF agents within each cohort. Log rank tests were used to test the null hypothesis of no differences among the Kaplan–Meier survival curve estimates. Proportional hazard assumptions were assessed graphically by comparing survival curves estimated by Cox regression models and Kaplan–Meier estimates and by assessing the log–log survival plots. Cox proportional hazard regression models estimated unadjusted and adjusted HR of discontinuation.

For each of the study outcomes, comparisons were performed among the three cohorts (biologically naive, first-time switchers and second-time switchers), and among the three anti-TNF agents stratified within the biologically naive and first-time switcher cohorts. For the primary analysis of persistence, we used the visit dates of reported initiation and visit dates of reported discontinuation. An analysis was also carried out in which we used dates as described above for those who indicated starting or discontinuing 'at the visit' but for those indicating 'since last visit' we substituted the date halfway between visits with little change in results. Comparisons of the three anti-TNF agents among second-time switchers were not performed due to small sample sizes within this cohort. We also performed sensitivities that incorporated major changes in dose/frequency in the survival analyses and imputed non-response for major dose/frequency escalations. We distinguished high versus low dose/frequency for adalimumab as 40 mg weekly versus every 2 weeks, and for infliximab using the cutpoint of of greater than 6 mg/kg every 8 weeks or equivalent based on a previously published cutpoint.[22]

To allow comparison with other registries and RCT, crude response and remission rates were stratified on the basis of whether or not patients met the eligibility criteria from three major published controlled trials.[12] As the registry records 28-joint counts, we estimated 28-joint count equivalents for the RCT 66-joint count requirements based on the 28-joint validation methodology previously described.[12] For the 66-joint count threshold of six or more tender and swollen joints, we applied the estimated 28-joint count equivalent of four of more joints such that patients who were deemed RCT eligible had four or more swollen joints, four or more tender joints and 45 min or more of morning stiffness at the time of registry enrollment. Power calculations varied across study outcomes for 6-month modified ACR outcomes. In biologically naive patients, we had 93% power to detect an OR of 2.0. For DAS28 remission at 6 months we had 76% power to detect an OR of 2.25.

RESULTS

Demographic and clinical characteristics

The study population consisted of 2242 RA patients; 1475 patients were biologically naive before initiating anti-TNF therapy, 616 had switched to a second anti-TNF agent (termed 'first-time switchers') and 151 had switched to their third anti-TNF agent (termed 'second-time switchers'). The baseline