Comparative Effectiveness of Nonbiologic versus Biologic Disease-modifying Antirheumatic Drugs for Rheumatoid Arthritis

ESI MORGAN DEWITT, YANHONG LI, JEFFREY R. CURTIS, HENRY A. GLICK, JEFFREY D. GREENBERG, KEVIN J. ANSTROM, JOEL M. KREMER, GEORGE REED, KEVIN A. SCHULMAN, and SHELBY D. REED

ABSTRACT. *Objective*. To evaluate the comparative effectiveness of nonbiologic disease-modifying antirheumatic drugs (DMARD) versus biologic DMARD (bDMARD) for treatment of rheumatoid arthritis (RA), using 2 common analytic approaches.

Methods. We analyzed change in Clinical Disease Activity Index (CDAI) scores in patients with RA enrolled in a US-based observational registry from 2001 to 2008 using multivariable (MV) regression and propensity score (PS) matching. Among patients who initiated treatment with a nonbiologic DMARD (n = 1729), we compared patients who switched to, or added, another nonbiologic (n = 182) or a bDMARD (n = 342) at 5, 9, and 24 months after treatment change.

Results. Both analytic approaches showed that patients switching to or adding another nonbiologic DMARD demonstrated improvement across 9 and 24 months (both p < 0.001). Both approaches also demonstrated greater improvement in CDAI among recipients of bDMARD relative to a second nonbiologic DMARD at 5 months (p < 0.02). The MV regression approach upheld these results at 9 and 24 months (p < 0.03). In contrast, the PS-matching approach did not show a sustained advantage with bDMARD at these later timepoints, possibly because of lower statistical power and/or lower baseline disease activity in the PS-matched cohort.

Conclusion. Patients in both treatment groups generally experienced lower CDAI scores across time. Patients switching to bDMARD demonstrated greater improvement than patients switching to nonbiologic DMARD with both analytic approaches at 5 months. Relative advantages with bDMARD were observed at 9 and 24 months only with MV regression. These analyses provide a practical example of how findings in comparative effectiveness research can diverge with different methodological approaches. (First Release Jan 15 2013; J Rheumatol 2013;40:127–36; doi:10.3899/jrheum.120400)

Key Indexing Terms:

REGISTRIES   LONGITUDINAL STUDIES   RHEUMATOID ARTHRITIS

DISEASE-MODIFYING ANTIRHEUMATIC DRUGS   OUTCOMES

The study of observational data presents unique challenges compared to the study of experimental (i.e., randomized) data because of the inherent confounding by indication and a multitude of potential biases. Benefits of comparative effectiveness research in observational data often include the opportunity to (1) compare therapeutic approaches that would not be studied in head-to-head clinical trials; (2) include larger numbers of patients in the analysis to increase power; and (3) evaluate treatment effects over longer periods of time and in a wider array of patients than would have been enrolled in clinical trials[1]. These benefits amount to understanding the safety and effectiveness of treatments as used in "real-world" settings.

Given the marked differences in prices and potential

differences in clinical effectiveness between therapeutic classes for the treatment of rheumatoid arthritis (RA), we sought to compare the effectiveness of nonbiologic disease-modifying antirheumatic drugs (DMARD) with biologic DMARD [bDMARD; specifically tumor necrosis factor (TNF) antagonists] for treatment of RA in real-world clinical practice[2]. TNF antagonists are bDMARD approved by the US Food and Drug Administration for treatment of RA in patients who have insufficient improvement while taking a nonbiologic DMARD. To identify patients who experienced insufficient improvement on a nonbiologic DMARD, we adopted a "revealed preference" approach in which patients' first change in treatment from their initial nonbiologic DMARD was indicative of lack of effectiveness of the initial medication, acknowledging that this could have been due to a variety of factors including insufficient improvement, lack of tolerability, or preference[3]. The selection of our study cohort centered on the identification of patients at the time they switched from initial treatment with a nonbiologic DMARD to an alternative therapy: either a second nonbiologic DMARD or a bDMARD (i.e., TNF antagonist) with or without continued use of a nonbiologic DMARD.

For 2 reasons, we chose to focus on patients who were undergoing a switch from an initial treatment course with a nonbiologic DMARD to an alternative treatment. First, this event represents an important juncture in treatment decision-making that is experienced by many patients with RA and their doctors. Second, focusing on patients switching from an initial treatment minimizes threats of bias and confounding encountered in observational studies of real-world clinical data, thereby increasing the validity of our findings. Including all patients with RA in an observational data set and dividing them into groups based on treatment with either a nonbiologic DMARD or a bDMARD would result in inclusion of patients with countless possible treatment courses over time, making it extremely difficult to disentangle the effects of any one treatment. The consequences of an "all-comer" analysis include lack of precision, confounding, and potential underestimation of benefit, because recent studies have shown that the initial therapeutic effect of a bDMARD is greater than that of subsequent bDMARD[4]. Although focusing on a narrow but well-defined timepoint was expected to reduce sample size, it creates an unbiased situation and allows for a more valid treatment comparison.

Having identified the population in which to conduct a comparative effectiveness analysis, we used 2 analytic approaches that are common in comparative effectiveness research. The first approach relied on multivariable regression to adjust for any potential differences in patient and site characteristics. The second relied on matching patients treated with nonbiologic DMARD and patients treated with bDMARD on the basis of clinical and demographic features, to develop like groups for comparison. We anticipated similar results with both analytic approaches.

MATERIALS AND METHODS

The study population consisted of adult rheumatology patients enrolled in a United States-based observational registry between October 2001 and June 2008 [Consortium of Rheumatology Researchers of North America (CORRONA)]. The registry collects clinical data every 4 to 5 months from physicians and patients at private practice and academic medical centers. Patients are contributed to the registry by 272 physicians from 100 study sites. Physicians and patients complete detailed case report forms providing information on demographics, medication history, medication use, clinical symptoms, tender and swollen joint counts, global assessments, and patient-reported outcomes. Sites are reimbursed for case report form completion[5].

To construct the study cohorts, we first included patients whose first use of any DMARD occurred during their participation in the registry (patients who at enrollment reported prior use of a DMARD were excluded), creating an "incident DMARD user" cohort (Figure 1). We then excluded patients whose first DMARD was a bDMARD. Patients with a history of cancer or heart failure were excluded because of relative contraindication to bDMARD use, to avoid potential selection bias in the nonbiologic DMARD group. Then, we identified patients who switched or added a different nonbiologic DMARD, to construct the "switch to nonbiologic DMARD" group. We constructed the "switch to bDMARD" group by identifying patients from the "incident DMARD user" cohort whose next switch in therapy included a bDMARD. Along with the bDMARD, patients could have discontinued or continued their initial nonbiologic DMARD or switched to another nonbiologic DMARD. The analysis of propensity score matching was conducted in propensity-score matched subsets of these 2 "full" cohorts.

For our analyses, the patient's treatment group was established on the basis of the first switch after the initial DMARD. Demographic and clinical characteristics of treatment groups were compared at baseline, defined as at the time of drug switch, with t-tests for continuous variables and chi-squared tests for categorical variables.

Modeling comparative effectiveness. Effectiveness was measured as change (increase or decrease) in Clinical Disease Activity Index (CDAI) scores over time. The CDAI is an index of patient-reported and physician-reported variables [sum of 28 tender and swollen joint counts, physician and patient global assessments each on a 0–10 visual analog scale (VAS) that measures disease activity][6]. The minimum clinically important improvement in CDAI reported by patients in a study by Aletaha, *et al* varies by baseline disease activity; for low, moderate, and high disease activity, these values are 1.8, 7.3, and 17.8, respectively[7]. The CDAI is more practical for use in clinical registries than measures such as the American College of Rheumatology 20/50/70 or the 28-joint Disease Activity Score, which require laboratory values that are not always measured routinely in clinical care[7]. We did not use the modified Health Assessment Questionnaire (mHAQ)[8,9,10], even though it was also available in CORRONA. The mHAQ measures physical disability, which can be less reflective of change over time[11]. Typical threshold CDAI values are used to represent remission (< 2.8), low disease activity (2.8 to < 10), moderate disease activity (> 10 to ≤ 22), and high disease activity (> 22)[12].

To better understand any short-term differences in treatment effect, we computed mean CDAI estimates for both treatment groups at the visit immediately prior to the baseline visit, the baseline visit when the drug switch was initiated, and the initial followup visit after the switch.

We evaluated comparative effectiveness at 5 months, 9 months, and 24 months. The 5-month timeframe was chosen to maximize the number of patients with CDAI from at least 1 visit after the baseline visit while minimizing the number of patients in the second nonbiologic DMARD cohort crossing over to treatment with a bDMARD. The longer timepoints
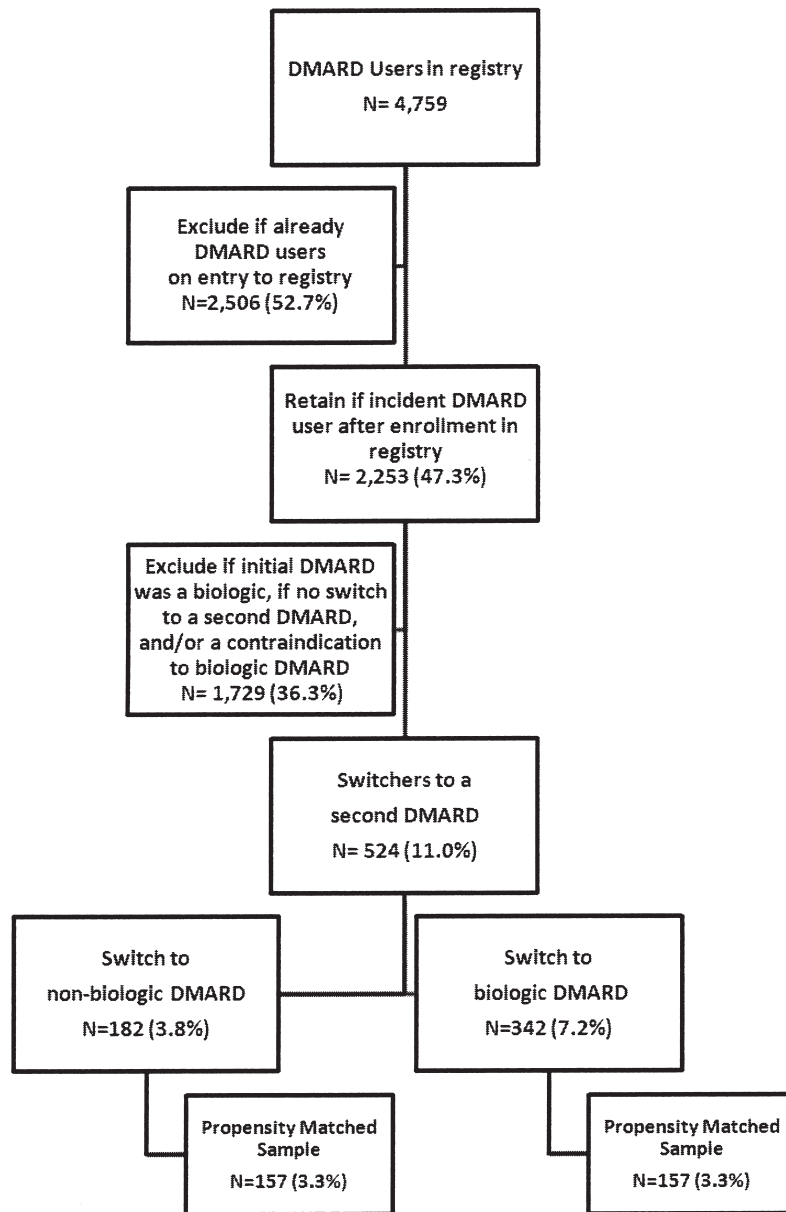
Figure 1. Derivation of samples for study of comparative effectiveness. DMARD: disease-modifying antirheumatic drugs.

were chosen to evaluate whether longer-term effects of treatment were evident.

We applied 2 analytic approaches to evaluate the relative effectiveness of the treatments: traditional multivariable adjustment applied to the full, unmatched cohort and to a matched cohort through use of propensity scores. The same variables were used in each analytic approach.

Multivariable regression on unmatched cohort. The first approach was to retain all patients meeting the inclusion criteria and to use multivariable regression modeling to statistically adjust for any potential differences in patient or site characteristics. Missing data were handled with imputation using the sample mean for continuous variables and the mode response for categorical variables. To take into account the clustering effect of patients nested within physician practices, we employed 3-level mixed linear models to examine the change in CDAI after patients switched from the initial treatment with a nonbiologic DMARD. Treatment was modeled as a

fixed effect, and time was modeled as both fixed and random effects. Random intercepts for study sites and individual patients were also specified. Thus, the models' 3 levels comprised site, time, and patient. We also included several clinical and demographic variables measured at baseline or at the time of switch from the initial nonbiologic DMARD (Table 1). The effectiveness (i.e., change in CDAI) of the nonbiologic DMARD group is represented in the measurement estimate for time, in months. Differential effectiveness of bDMARD relative to nonbiologic DMARD is represented as the interaction term between time and treatment.

Multivariable regression on propensity score-matched cohort. The second approach was to pair patients with similar characteristics from the 2 treatment groups before making comparisons. We derived propensity scores for matching patients in the 2 treatment groups based on the conditional probability of receiving a bDMARD[13]. The model used for derivation of propensity scores (PS) was initially developed in a separate

Table 1. Select baseline characteristics of study cohorts. For overall registry sample, baseline is at initial use of DMARD. For "Switchers," baseline is at time of switch.

	DMARD Users in Registry		"Switchers" After Initial Nonbiologic DMARD treatment							
			Unmatched Cohort		Propensity Score-matched Cohort		Subgroup (unmatched cohort): ≤ 3 Swollen Joints		Subgroup (unmatched cohort): ≥ 4 Swollen Joints	
	DMARD Only**	Biologic User***	Second DMARD[†]	Biologic[††]	Second DMARD[†]	Biologic[††]	Second DMARD[†]	Biologic[††]	Second DMARD[†]	Biologic[††]
	2515	2244	182	342	157	157	82	114	99	225
Demographics										
Age, yrs, mean (SD)	58.5 (13.4)	55.5 (13.2)	58.8 (13.1)	56.5 (12.6)	57.7 (13.3)	57.1 (11.4)	58.1 (13.6)	55.1 (12.0)	59.3 (12.7)	57.2 (13.0)
Disease duration, yrs (SD)	7.5 (9.0)	9.5 (9.8)	6.7 (7.9)	8.6 (8.7)	7.2 (8.3)	6.9 (7.0)	7.0 (7.2)	8.1 (8.1)	6.4 (8.6)	8.9 (9.0)
Female, %	74.7	77.1	74.2	79.2	77.7	79.6	74.4	77.2	73.7	80.4
White, %	88.7	90.0	91.8	87.7	91.1	89.8	87.8	86.8	96.0	88.0
Clinical characteristics										
CDAI, mean (SD)	17.0 (13.5)	19.3 (14.4)	16.8 (11.5)	21.7 (13.6)	17.2 (11.5)	16.6 (10.5)	8.3 (6.7)	10.2 (7.2)	24.0 (9.5)	27.4 (12.4)
MHAQ, mean (SD)	0.4 (0.5)	0.4 (0.5)	0.4 (0.5)	0.5 (0.5)	0.4 (0.5)	0.4 (0.4)	0.3 (0.4)	0.4 (0.4)	0.5 (0.5)	0.6 (0.5)
Swollen joint count, mean (SD)	6.2 (6.7)	6.4 (6.5)	6.6 (7.0)	7.7 (6.8)	6.9 (7.2)	6.7 (6.9)	0.9 (1.1)	1.2 (1.1)	11.3 (6.3)	11.0 (6.1)
Tender joint count, mean (SD)	4.7 (6.0)	6.0 (6.8)	4.2 (5.3)	6.8 (6.8)	4.2 (5.4)	4.2 (5.2)	2.3 (3.8)	3.5 (4.3)	5.7 (5.9)	8.4 (7.3)
Physician global disease activity, 0–100, mean (SD)	27.4 (22.0)	32.9 (22.7)	23.8 (19.2)	32.7 (21.3)	24.6 (19.6)	25.2 (18.2)	17.4 (17.8)	24.3 (19.0)	29.3 (18.6)	36.9 (21.2)
DMARD use										
Initial, n (%)										
Leflunomide			15 (8.2)	86 (25.1)	14 (8.9)	34 (21.7)	5 (6.1)	27 (23.7)	10 (10.1)	57 (25.3)
Sulfasalazine			26 (14.3)	40 (11.7)	25 (15.9)	23 (14.6)	15 (18.3)	12 (10.5)	11 (11.1)	28 (12.4)
Methotrexate			89 (48.9)	170 (49.7)	74 (47.1)	76 (48.4)	34 (41.5)	60 (52.6)	54 (54.5)	10 (48.9)
Minocycline			5 (0.5)	1 (0.3)	4 (2.5)	5 (3.2)	2 (2.4)	2 (1.8)	3 (3.0)	6 (2.7)
Cyclosporine			4 (2.2)	3 (0.9)	4 (2.5)	2 (1.3)	2 (2.4)	1 (0.9)	2 (2.0)	2 (0.9)
Hydroxychloroquine			53 (29.1)	67 (19.6)	44 (28.0)	32 (20.4)	29 (35.4)	27 (23.7)	24 (24.2)	40 (17.8)
Subsequent, n (%)										
Leflunomide			78 (42.9)		63 (40.1)		33 (4.02)		44 (44.4)	
Sulfasalazine			27 (14.8)		25 (15.9)		9 (11.0)		18 (18.2)	
Methotrexate			42 (23.1)		38 (24.2)		18 (22.0)		24 (24.2)	
Minocycline			11 (6.0)		9 (5.7)		9 (11.0)		2 (2.0)	
Cyclosporine			5 (2.8)		4 (2.6)		2 (2.4)		3 (3.0)	
Hydroxychloroquine			25 (13.7)		23 (14.6)		14 (17.1)		11 (11.1)	
Subsequent biologic, n (%)										
Etanercept				114 (33.3)		48 (30.6)		37 (32.5)		76 (33.8)
Adalimumab				135 (39.5)		63 (40.1)		54 (47.4)		79 (35.1)
Infliximab				98 (28.7)		48 (30.6)		25 (21.9)		73 (32.4)

** Cohort who used only nonbiologic DMARD during study period. *** Cohort who ever used a biologic DMARD during study period. [†] Group who switched from incident nonbiologic DMARD to second nonbiologic DMARD. [††] Group who switched from incident nonbiologic DMARD to treatment including a biologic DMARD. DMARD: disease-modifying antirheumatic drugs; CDAI: Clinical Disease Activity Index; MHAQ: modified Health Assessment Questionnaire.

study by Curtis, *et al*[14]. Variables included in the logistic regression model to estimate propensity scores are listed in Table 2. Patients were excluded if they had 4 or more missing covariates needed to estimate propensity scores. For the patients missing data for 3 or fewer covariates, missing values were imputed using the mean value for continuous variables and the mode response for categorical variables. Then, we matched patients between comparator groups using a "greedy match" algorithm applied to the propensity scores. Greedy matching moves sequentially through each of the observations that are to be matched and selects the match having the most similar propensity score. After matching, we applied a multivariable mixed linear model using the same covariates (Table 2) as in the multi-variable regression approach applied to unmatched samples described above. In effect, these 2 steps served to "double-adjust" with PS matching,

creating 2 similar groups with regard to baseline characteristics, and the subsequent adjustment with multivariable regression to increase statistical precision, akin to applying multivariable regression in a randomized trial[15].

Subgroup analyses. We also performed the analyses described above for 2 subgroups of patients: those with at least 4 active (swollen) joints at the point of switching from the initial nonbiologic DMARD and those with 3 or fewer active (swollen) joints. These subgroups were chosen because having 4 active joints on the 28-joint count corresponds to an inclusion criterion for clinical trials[16,17]. By contrast, patients with 3 or fewer active joints had correspondingly low disease activity by CDAI scores, with little room for improvement on this measure with a change in therapy. As a sensitivity analysis, the multivariable regression analysis was also performed on the full sample after censoring patients in the second nonbiologic DMARD

Table 2. Multivariable regression model and propensity score-matched adjusted model results comparing effectiveness of nonbiologic versus biologic DMARD treatment in reducing CDAI in the full cohort.

Time Period	Multivariable Regression Model			Propensity Score-matched Model		
	Estimate (SE)	95% CI	p	Estimate (SE)	95% CI	p
5-month results						
Difference in baseline CDAI of biologic DMARD user compared to nonbiologic DMARD user	0.03 (0.05)	–0.07, 0.13	0.55	0.04 (0.07)	–0.10, 0.18	0.57
Monthly change in CDAI after switch to nonbiologic DMARD	–0.22 (0.39)	–0.99, 0.55	0.58	–0.12 (0.37)	–0.84, 0.60	0.74
Incremental monthly change in CDAI after switch to biologic DMARD*	–2.68 (0.49)	–3.64, –1.71	< 0.0001	–1.24 (0.54)	–2.29, –0.19	0.02
9-month results						
Difference in baseline CDAI of biologic DMARD user compared to nonbiologic DMARD user	–0.42 (0.62)	–1.69, 0.85	0.50	–1.12 (0.71)	–2.58, 0.33	0.12
Monthly change in CDAI after switch to nonbiologic DMARD	–0.62 (0.19)	–0.99, –0.26	< 0.001	–0.63 (0.18)	–0.98, –0.29	< 0.001
Incremental monthly change in CDAI after switch to biologic DMARD*	–0.53 (0.23)	–0.99, –0.08	0.02	0.21 (0.25)	–0.27, 0.70	0.39
24-month results						
Difference in baseline CDAI of biologic DMARD user compared to nonbiologic DMARD user	–0.39 (0.66)	–1.75, 0.97	0.56	–0.86 (0.74)	–2.38, 0.66	0.26
Monthly change in CDAI after switch to nonbiologic DMARD	–0.27 (0.06)	–0.39, –0.15	< 0.0001	–0.30 (0.06)	–0.41, –0.19	< 0.001
Incremental monthly change in CDAI after switch to biologic DMARD*	–0.16 (0.07)	–0.30, –0.02	0.03	–0.02 (0.08)	–0.18, 0.14	0.81

* Point estimate represents the incremental change in CDAI each month in the biologic DMARD group compared to the nonbiologic DMARD group. This is the modeled interaction term between time (month) and treatment group. Additional covariates applied in propensity-score matching and multivariable models: age, duration of disease, sex, tender joint count, swollen joint count, subcutaneous nodules, joint deformity, disease activity, HAQ-DI, comorbidity history, patient visual analog scale (VAS), patient pain VAS, morning stiffness, body mass index, American College of Rheumatology functional class, tobacco use, alcohol use, and exercise. Tender joint count, swollen joint count, disease activity, and patient VAS had significant p values (alpha level, 0.05) at all timepoints. DMARD: disease-modifying antirheumatic drugs; CDAI: Clinical Disease Activity Index; HAQ-DI: Health Assessment Questionnaire-Disability Index.

group who later went on to receive treatment with a bDMARD (i.e., patients whose first switch was from a nonbiologic to a different nonbiologic DMARD, but who later switched to a bDMARD).

Our study was approved by the Institutional Review Board (IRB) of Duke University Medical Center, and data collection within CORRONA is governed by the New England IRB. All statistical analyses were conducted using SAS 9.2.

RESULTS

Table 1 reports sample sizes and select baseline characteristics of the DMARD users in the overall registry population and for "switchers" after initial nonbiologic DMARD treatment. The switchers are presented as the unmatched and the propensity score-matched study cohorts, and patients in the unmatched cohort are presented further in subgroups of those with 3 or fewer active joints, or 4 or more active joints at the time of the switch. Table 1 also presents the initial nonbiologic DMARD used and the subsequent DMARD.

The initial comparison shows the overall sample of patients in the registry treated with any DMARD during the study period. These were divided into those who used only nonbiologic DMARD compared to those who ever used a bDMARD during the study period. There were 4759 patients in the registry who were users of any DMARD during the study period. Of those patients, 2253 had their first incident use of any DMARD, nonbiologic or biologic,

after enrollment into the registry. After restricting the cohort to patients who were incident users of nonbiologic DMARD (1729) who later switched to an alternate DMARD regimen, and eliminating those who had contraindication to bDMARD therapy, overall sample size was reduced by 89% to 524. Progressive restriction of the treatment groups by 1:1 matching based on propensity scores resulted in further loss of sample size (Figure 1).

As designed, matched treatment groups were more similar on baseline characteristics compared to the unmatched groups. Compared to the unrestricted cohort of bDMARD users, the bDMARD users in the matched cohort had lower mean CDAI and affected joint counts, reflecting a less severe disease profile of patients for analysis. When examining patients with 4 or more swollen joints, the severity profile increased markedly and sample size was reduced by 38.2%, from 524 to 324 patients. The mean duration of followup from the time of switch was 727 days. Some patients in the second nonbiologic DMARD group went on to begin a bDMARD within the period of analysis in the following proportions: at 5 months, 7.4%; at 9 months, 13.2%; and by 2 years, 25.0%.

Figures 2 and 3 depict the change in CDAI for the treatment groups between the visit prior to the baseline visit and the baseline visit and first visit after the change in therapy. The average duration between baseline visit and
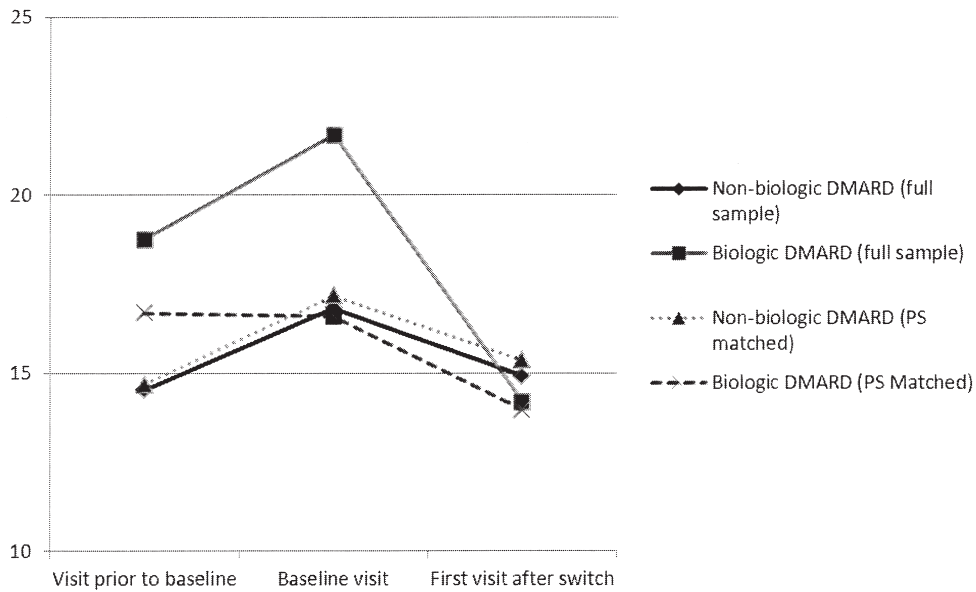
Figure 2. Mean CDAI score prior to, at visit, and after visit when treatment switch made, for the full sample. CDAI is the sample mean. Full multivariable adjustment sample size is 524. Full propensity score (PS)-matched sample size is 314. CDAI: Clinical Disease Activity Index; DMARD: disease-modifying antirheumatic drugs.
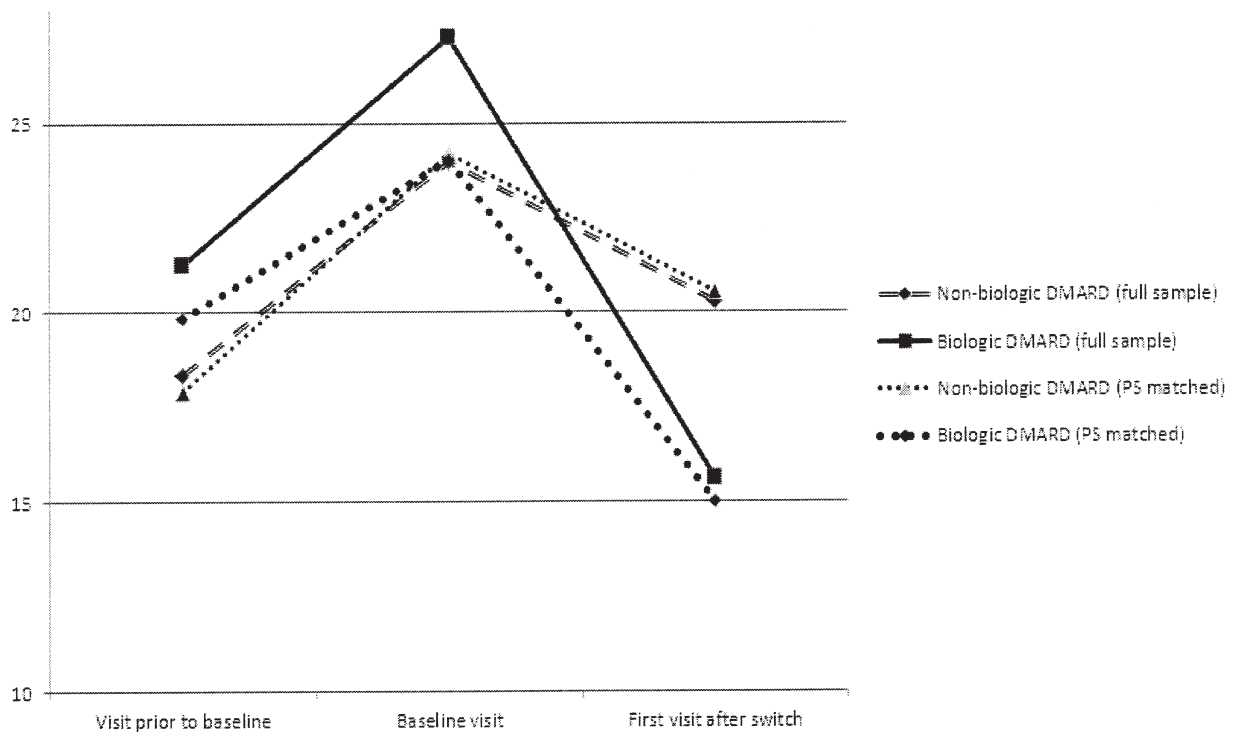

Figure 3. Mean CDAI prior to, at visit, and after visit when treatment switch made, subgroup ≥ 4 swollen joints. CDAI is the sample mean. Multivariable sample size is 324. Propensity score-matched sample size is 164. CDAI: Clinical Disease Activity Index; DMARD: disease-modifying antirheumatic drugs; PS: propensity score.

first CDAI documented in the registry after the medication switch was similar between groups, 151.7 days in the nonbiologic DMARD group and 150.1 days in the bDMARD group.

Multivariable regression on unmatched cohort. In the full, unmatched patient cohort, patients in the nonbiologic DMARD treatment group experienced improvement in CDAI over the longer timeframes, as indicated by the variable "monthly change in CDAI after switch to nonbiologic DMARD." While this was not significant over 5 months (p = 0.58), it was significant in the 9-month and 24-month analyses (p < 0.001; Table 2). In addition, a switch to bDMARD therapy was found to be more effective than a switch to a nonbiologic DMARD regimen in lowering CDAI across all time periods studied. This is represented by a significant interaction term reported in Table 2 as "Incremental monthly change in CDAI after switch to bDMARD" (p < 0.03). The point estimate in the 5-month analysis represents an additional monthly decline of 2.68 CDAI units in the biologic group compared to the nonbiologic group (p < 0.0001). In the 9-month and 24-month analyses, this differential treatment effect continued. Additional covariates consistently significant in the models were tender and swollen joint counts, global Disease Activity Score, and patient global VAS (data not shown).

Multivariable regression on the propensity score-matched cohort. In the propensity score-matched analysis there was also a favorable effect on CDAI at the 9-month (p < 0.001) and 24-month (p < 0.001) timepoints in the nonbiologic DMARD group (monthly change in CDAI after switch to nonbiologic DMARD; Table 2). However, the effect was greater among bDMARD users at 5 months compared to nonbiologic DMARD users, with a relative monthly decrease of 1.24 CDAI units (incremental monthly change in CDAI after switch to bDMARD, p = 0.02).

Subgroup analyses. In patients with higher disease activity, represented by patients with at least 4 swollen joints, the pattern of results in the multivariable analysis replicated the findings in the larger cohort (Appendix 1). At 5 months, there was no significant effect on CDAI in the second nonbiologic DMARD treatment group. By 9 months and 24 months, there was documented improvement in CDAI in the second nonbiologic DMARD treatment group. There was also indication of greater incremental benefit with bDMARD at 5 months (–3.31 CDAI units; p < 0.0001) and 9 months (–0.57 CDAI units; p = 0.06). However, unlike in the main analysis, this differential effect was no longer statistically significant across 24 months of followup (p = 0.11). The result of the PS-matched comparisons of patients with at least 4 swollen joints was consistent with the multivariable analyses (Appendix 1). There was no statistically significant effect on CDAI of either treatment over time in multivariable analysis or propensity score-matched analysis restricted to the subgroup of patients with 3 or fewer swollen joints (data not shown). In the sensitivity analysis that censored patients who switched from the second nonbiologic DMARD to a bDMARD during the 24 months of analysis, the findings were consistent with the regression analysis on the unmatched cohort (data not shown).

DISCUSSION

We designed our study to address a question of clinical interest in rheumatology: in routine clinical practice, when a patient with RA changes treatment from an initial nonbiologic DMARD, will there be a differential effect on disease activity between use of nonbiologic versus bDMARD? By focusing on patients who switched treatment regimens after initial use of a nonbiologic DMARD, we identified patients who were at a common clinical juncture, providing a uniform starting point for the analysis. Although focusing on this cross-section in time resulted in a substantial reduction in sample size, we did not limit the sample on the basis of demographic or other clinical characteristics, to preserve the rich heterogeneity offered by an observational cohort. Thus we posit that our approach allowed for a satisfactory comparison of treatment alternatives within an observational registry in which decisions to prescribe bDMARD were not standardized[18]. This approach allows our findings to be generalized to patients with RA at a particular treatment decision point.

We studied the comparative effectiveness of nonbiologic DMARD versus bDMARD in patients with RA, using 2 standard analytic approaches for rigor — multivariable regression and propensity score-matched models, while anticipating that findings would be similar. Results of multivariable regression analyses on unmatched samples showed bDMARD to be more effective in lowering CDAI than nonbiologic DMARD at all the timepoints studied. While this finding was indeed replicated in the 5-month analysis in the propensity score-matched cohort, effects at 9 and 24 months were not statistically significant.

Both analytic approaches demonstrated superior effectiveness of bDMARD at 5 months, at which point there was no significant improvement yet seen in the nonbiologic DMARD group. This is consistent with the mechanism of action of the represented bDMARD, TNF antagonists, which may confer greater potential for early improvement compared to slower-acting nonbiologic DMARD. This finding was observed in the overall sample as well as the subgroup of patients with more active arthritis (defined as at least 4 active joints), as would be anticipated from clinical trials results[19]. In a subgroup of patients with less active disease, assessed as 3 or fewer swollen joints, no effect on the CDAI over time was observed with either treatment. Of note, however, no change in CDAI indicates a lack of worsening, and may also reflect a "floor effect," with inability to measure improvement[17]. Figures 2 and 3 suggest

that the short-term effect of the treatments may reflect regression to the mean. However, the 9-month and 24-month analyses showed that both treatment groups benefited over time.

The goal of comparative effectiveness research with observational data is to calculate the real-world effectiveness of treatments used in routine clinical practice. The strategy to compare the effectiveness of therapeutic choices by analytic adjustment, such as the multivariable regression analysis, allows inclusion of larger, more heterogeneous study samples. One disadvantage of analytic adjustment is the need to apply multiple modeling assumptions that may not be transparent to the audience of the study. In addition, use of these methods may obscure the patient population to whom the results can be generalized. The advantages of matching (as we applied with propensity scores) are that it identifies the patient profiles to which the results can be generalized and can reduce potential confounding[20,21]. The use of propensity score matching when combined with multivariable regression modeling in observational studies has been demonstrated to yield results closer to randomized trial results than use of either technique on its own[15]. The disadvantages of matching can include significant reductions in sample size and patient heterogeneity, arguably the main advantages to using observational data. In our study, propensity score-based matching further reduced sample size and resulted in lower mean disease severity among the bDMARD users in the matched cohorts than in the general cohort. Patients in whom change in active disease may have been more likely were culled from the analysis pool by the matching process. Thus, it is reasonable to infer that the lack of differential effect between biologic versus nonbiologic treatments in the matched analyses over the longer time horizons was due to studying subsets of patients who are different from the typical severity profile of patients receiving a biologic. However, it should also be noted that the subgroup analyses had less statistical power to detect any differences, had they been present, although the effect size we observed was relatively small, irrespective of its statistical significance. More generally, both methodological approaches can provide useful information on comparative effectiveness[22], but their limitations and assumptions should be fully transparent to those who may use this information for decision-making.

In addition to statistical adjustments to address potential confounding, other factors that can influence observational studies should be considered. First, the effects that we ascribed to "monthly change in CDAI after switch to nonbiologic DMARD" could reflect regression to the mean. Also, we did not have a measure of adherence. We applied intent-to-treat principles with the motivation to collect a longer duration of followup than the relatively short horizons typically used in randomized controlled trials. However, with intent-to-treat analyses, interpretations of

treatment effects are more problematic because of greater censoring and crossover over longer periods of followup. A sensitivity analysis revealed that crossover during the time of the study did not appear to alter our study conclusions. However, with the passage of time, because patients in either group may have added or switched to other medications, the degree of misclassification of treatment group becomes greater, and we are less able to isolate the longterm effectiveness of the drugs of initial interest. These issues may have biased estimates of differential treatment effects toward the null; we found point estimates decreasing across analyses with longer time periods. In addition, we grouped different medications into nonbiologic DMARD and bDMARD to maximize sample size. However, meta-analyses of TNF antagonists support that such grouping is reasonable given similar effectiveness of available preparations[23].

Although the goal of comparative effectiveness research is not to replicate clinical trial results by restricting samples to a homogeneous group of patients who would meet trial entry requirements, use of study methods that would result in balanced comparison groups in the observational sample is an important aim. Obtaining results consistent with clinical trials in a heterogeneous observational sample conveys credibility to the methodology used[18]. To maximize the internal validity of our comparisons, we relied on strategies used in clinical trials, such as limiting the study sample for comparison, applying intent-to-treat principles, examining relatively short followup periods, and (in subgroup analyses) examining patients with more severe disease. Although our comparative effectiveness analyses were consistent with results of randomized controlled trials, our study provides unique insight into the relative effectiveness of bDMARD in patients who have not been included in clinical trials because of factors such as comorbidity or not satisfying inclusion criteria, but who are routinely treated with bDMARD. In a subgroup of patients with low disease activity, we did not find evidence of differential treatment effect of bDMARD relative to nonbiologic DMARD, but this finding was expected. Although the patients may not have complete disease control, there may be little room for measurable improvement on scales currently in use in RA[17].

Insight into differential effectiveness of therapeutics such as bDMARD in the general RA population is important for conducting comparative effectiveness research. Our study demonstrates that results of comparative effectiveness studies may vary depending on choice of analytic method, and comparative effectiveness research studies should report whether findings are consistent when different methodological approaches are applied. Observational databases and registries representing routine clinical practice provide a unique opportunity to study outcomes across a patient population with a diverse range of

APPENDIX 1. Multivariable regression model and propensity score-matched adjusted model subgroup analysis results comparing effectiveness of nonbiologic versus biologic DMARD treatment in reducing CDAI in patients with 4 or more swollen joints.

Time Period	Multivariable Regression Model			Propensity Score-matched Model		
	Estimate (SE)	95% CI	p	Estimate (SE)	95% CI	p
5-month results						
Difference in baseline CDAI of biologic DMARD user compared to nonbiologic DMARD user	0.06 (0.08)	−0.11, 0.23	0.48	0.08 (0.14)	−0.22, 0.39	0.56
Monthly change in CDAI after switch to nonbiologic DMARD	−0.76 (0.55)	−1.84, 0.32	0.17	−0.83 (0.53)	−1.87, −0.21	0.12
Incremental monthly change in CDAI after switch to biologic DMARD*	−3.31 (0.66)	−4.61, −2.00	< 0.0001	−2.15 (0.79)	−3.65, −0.65	< 0.01
9-month results						
Difference in baseline CDAI of biologic DMARD user compared to nonbiologic DMARD user	−1.04 (0.91)	−2.97, 0.90	0.27	−1.78 (1.09)	−4.13, 0.57	0.13
Monthly change in CDAI after switch to nonbiologic DMARD	−1.27 (0.25)	−1.75, −0.79	< 0.0001	−1.30 (0.24)	−1.78, −0.83	< 0.0001
Incremental monthly change in CDAI after switch to biologic DMARD*	−0.57 (0.30)	−1.15, 0.01	0.06	−0.03 (0.35)	−0.72, 0.66	0.93
24-month results						
Difference in baseline CDAI of biologic DMARD user compared to nonbiologic DMARD user	−1.74 (0.93)	−3.71, 0.23	0.08	−2.16 (1.15)	−4.65, 0.33	0.08
Monthly change in CDAI after switch to nonbiologic DMARD	−0.41 (0.09)	−0.59, −0.23	< 0.0001	−0.45 (0.09)	−0.63, −0.28	< 0.0001
Incremental monthly change in CDAI after switch to biologic DMARD*	−0.17 (0.11)	−0.37, 0.04	0.11	−0.05 (0.13)	−0.30, 0.19	0.68

* Point estimate represents the incremental change in CDAI each month in the biologic DMARD group compared to the nonbiologic DMARD group. This is the modeled interaction term between time (month) and treatment group. Additional covariates applied in propensity-score matching and multivariable models: age, duration of disease, sex, tender joint count, swollen joint count, subcutaneous nodules, joint deformity, disease activity, HAQ-DI, comorbidity history, patient visual analog scale (VAS), patient pain VAS, morning stiffness, body mass index, American College of Rheumatology functional class, tobacco use, alcohol use, and exercise. Tender joint count, swollen joint count, disease activity, and patient VAS had significant p values (alpha level, 0.05) at all timepoints. DMARD: disease-modifying antirheumatic drugs; CDAI: Clinical Disease Activity Index; HAQ-DI: Health Assessment Questionnaire-Disability Index.

demographic and clinical characteristics who receive different treatments on the basis of variations in physician practice. Maintenance of such registries is resource-intensive, but may ultimately result in overall cost savings if we are able to produce generalizable knowledge to guide appropriate targeting of therapeutics across the spectrum of disease severity.

REFERENCES
1. Sox HC, Greenfield S. Comparative effectiveness research: A report from the Institute of Medicine. Ann Intern Med 2009;151:203-5.
2. Bansback NJ, Regier DA, Ara R, Brennan A, Shojania K, Esdaile JM, et al. An overview of economic evaluations for drugs used in rheumatoid arthritis: Focus on tumour necrosis factor-alpha antagonists. Drugs 2005;65:473-96.
3. Southwood TR, Foster HE, Davidson JE, Hyrich KL, Cotter CB, Wedderburn LR, et al. Duration of etanercept treatment and reasons for discontinuation in a cohort of juvenile idiopathic arthritis patients. Rheumatology 2011;50:189-95.
4. Greenberg JD, Reed G, Decktor D, Harrold L, Furst D, Gibofsky A, et al. A comparative effectiveness study of adalimumab, etanercept and infliximab in biologically naïve and switched rheumatoid arthritis patients: Results from the US CORRONA registry. Ann Rheum Dis 2012;71:1134-42.
5. Kremer JM. The CORRONA database. Autoimmun Rev 2006; 5:46-54.
6. Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): A review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol 2005;5 Suppl 39:S100-8.
7. Aletaha D, Funovits J, Ward MM, Smolen JS, Kvien TK. Perception of improvement in patients with rheumatoid arthritis varies with disease activity levels at baseline. Arthritis Rheum 2009;61:313-20.
8. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: dimensions and practical applications. Health Qual Life Outcomes 2003;1:20.
9. Anderson J, Sayles H, Curtis JR, Wolfe F, Michaud K. Converting modified Health Assessment Questionnaire (HAQ), multidimensional HAQ and HAQII scores into original HAQ scores using models developed with a large cohort of rheumatoid arthritis patients. Arthritis Care Res 2010;62:1481-8.
10. Pincus T, Summey JA, Soraci SA Jr, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. Arthritis Rheum 1983;26:1346-53.
11. Aletaha D, Smolen J, Ward MM. Measuring function in rheumatoid arthritis: Identifying reversible and irreversible components. Arthritis Rheum 2006;54:2784-92.
12. Aletaha D, Smolen JS. The Simplified Disease Activity Index and Clinical Disease Activity Index to monitor patients in standard clinical care. Rheum Dis Clin North Am 2009;35:759-72, viii.
13. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;

79:516-24.

14. Curtis JR, Reed G, Allison JJ, Moreland LW, Saag KG. How adequately do propensity scores adjust for confounding by indication in the absence of disease-specific markers of activity and severity? [abstract]. Arthritis Rheum 2006;54 Suppl:S346.

15. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. Annu Rev Public Health 2000;21:121–45.

16. Smolen JS, Breedveld FC, Eberl G, Jones I, Leeming M, Wylie GL, et al. Validity and reliability of the twenty-eight-joint count for the assessment of rheumatoid arthritis activity. Arthritis Rheum 1995;38:38-43.

17. Greenberg JD, Kishimoto M, Strand V, Cohen SB, Olenginski TP, Harrington T, et al. Tumor necrosis factor antagonist responsiveness in a United States rheumatoid arthritis cohort. Am J Med 2008;121:532-8.

18. Psaty BM, Siscovick DS. Minimizing bias due to confounding by indication in comparative effectiveness research: The importance of restriction. JAMA 2010;304:897-8.

19. Nam JL, Winthrop KL, van Vollenhoven RF, Pavelka K, Valesini G, Hensor EM, et al. Current evidence for the management of rheumatoid arthritis with biological disease-modifying antirheumatic drugs: A systematic literature review informing the EULAR recommendations for the management of RA. Ann Rheum Dis 2010;69:976-86.

20. Seeger JD, Williams PL, Walker AM. An application of propensity score matching using claims data. Pharmacoepidemiol Drug Saf 2005;14:465-76.

21. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997;8 Part 2:757-63.

22. Sturmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. J Clin Epidemiol 2006;59:437-47.

23. Gartlehner G, Hansen RA, Jonas BL, Thieda P, Lohr KN. The comparative efficacy and safety of biologics for the treatment of rheumatoid arthritis: A systematic review and metaanalysis. J Rheumatol 2007;33:2398-408.