



OPEN ACCESS

EXTENDED REPORT

Organ damage in patients treated with belimumab versus standard of care: a propensity score-matched comparative analysis

Murray B Urowitz,¹ Robert L Ohsfeldt,^{2,3} Ronald C Wielage,³ Kari A Kelton,³ Yumi Asukai,⁴ Sulabha Ramachandran⁵**Handling editor** Josef S Smolen

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/annrheumdis-2018-214043>).

¹Centre for Prognosis Studies in Rheumatic Diseases, Toronto Western Hospital, Toronto, Ontario, Canada

²School of Public Health, Texas A&M University, College Station, Texas, USA

³Medical Decision Modeling Inc, Indianapolis, Indiana, USA

⁴Value Evidence & Outcomes, GSK, Brentford, UK

⁵Value Evidence & Outcomes, GSK, Philadelphia, Pennsylvania, USA

Correspondence to

Prof Murray B Urowitz, Centre for Prognosis Studies in the Rheumatic Diseases, University of Toronto Senior Scientist, Krembil Research Institute, Toronto Western Hospital, Toronto, ON M5T 2S8, Canada; m.urowitz@utoronto.ca

Data from this study have previously been presented at the American College of Rheumatology Annual Scientific meeting (ACR 2017).

Received 3 July 2018

Revised 1 November 2018

Accepted 12 November 2018



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Urowitz MB, Ohsfeldt RL, Wielage RC, *et al.* *Ann Rheum Dis* Epub ahead of print: [please include Day Month Year]. doi:10.1136/annrheumdis-2018-214043

ABSTRACT

Objectives The study (206347) compared organ damage progression in patients with systemic lupus erythematosus (SLE) who received belimumab in the BLISS long-term extension (LTE) study with propensity score (PS)-matched patients treated with standard of care (SoC) from the Toronto Lupus Cohort (TLC).

Methods A systematic literature review identified 17 known predictors of organ damage to calculate a PS for each patient. Patients from the BLISS LTE and the TLC were PS matched posthoc 1:1 based on their PS (\pm calliper). The primary endpoint was difference in change in Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI) score from baseline to 5 years.

Results For the 5-year analysis, of 567 (BLISS LTE n=195; TLC n=372) patients, 99 from each cohort were 1:1 PS matched. Change in SDI score at Year 5 was significantly lower for patients treated with belimumab compared with SoC (-0.434 ; 95% CI -0.667 to -0.201 ; $p<0.001$). For the time to organ damage progression analysis (≥ 1 year follow-up), the sample included 965 (BLISS LTE n=259; TLC n=706) patients, of whom 179 from each cohort were PS-matched. Patients receiving belimumab were 61% less likely to progress to a higher SDI score over any given year compared with patients treated with SoC (HR 0.391; 95% CI 0.253 to 0.605; $p<0.001$). Among the SDI score increases, the proportion of increases ≥ 2 was greater in the SoC group compared with the belimumab group.

Conclusions PS-matched patients receiving belimumab had significantly less organ damage progression compared with patients receiving SoC.

INTRODUCTION

In the treatment of systemic lupus erythematosus (SLE), the clinical efficacy of belimumab plus standard SLE therapy (standard of care, SoC) compared with placebo plus SoC has been demonstrated in four phase III clinical trials.¹⁻⁴ Long-term extension (LTE) studies (BEL112233, NCT00724867 US patients and BEL112234, NCT00712933 outside-US patients) of the phase III studies BLISS-52 (BEL110752, NCT00424476) and BLISS-76 (BEL110751, NCT00410384) demonstrated that long-term exposure to belimumab was safe and well tolerated and that clinical efficacy was maintained.^{1,3,5} However, as the LTE studies did not have comparator arms, comparison of belimumab plus SoC (referenced as belimumab throughout)

Key messages**What is already known about this subject?**

► In the treatment of systemic lupus erythematosus (SLE), the clinical efficacy of belimumab plus standard SLE therapy (standard of care, SoC) compared with placebo plus SoC has been demonstrated in four phase III clinical trials; in addition to disease lowering benefits at phase III, two long-term extension (LTE) studies demonstrated that long-term exposure to belimumab was safe and well tolerated, with low rates of organ damage accrual over time.

What does this study add?

► As the LTE studies did not have comparator arms, comparison of belimumab plus SoC with SoC alone was not possible; the present study extends the work of the LTE studies by enabling a comparison of belimumab plus SoC versus SoC alone, using a matched comparator group.

► This study uses propensity score matching (PSM) to match patients in one of the belimumab LTE studies with patients from an external SLE cohort (the Toronto Lupus Cohort) treated with SoC alone, thus enabling a robust and credible comparison of belimumab's impact on organ damage over a 5-year period, compared with SoC treatment alone.

How might this impact on clinical practice or future developments?

► The results of this study demonstrate that belimumab plus SoC treatment results in lower rates of organ damage accrual compared with SoC alone. The benefits of treatment with belimumab have been demonstrated over a 5-year period, which may assist healthcare providers in determining the appropriate long-term treatment for patients with SLE.

► This study demonstrates the utility and value of PSM in providing a comparator for a treatment observed within an LTE. The results suggest that this methodology could be applied more frequently to reinforce the conclusions that can be drawn from these studies.

with SoC alone was not possible. Consequently, the question of long-term relative efficacy required further investigation.

Propensity score matching (PSM) is a statistical technique that allows matching of patients within two treatment groups based on their propensity score (PS).^{6,7} A patient's PS is based on their relationship to known clinical demographics and disease characteristics that would likely lead to the patient receiving a particular treatment. Therefore, once patients have been matched based on their PS, any observed difference in outcomes is assumed to be a direct result of the treatment. PSM is often used in observational studies to compare treatments, clinical techniques or subgroups^{8,9} or to adjust for confounding in two groups of patients with different observed outcomes, to examine the effect of the treatment.⁸

This study (206347) used PSM to match patients with SLE treated with belimumab in the US BLISS LTE study (BEL112233) with patients from an external SLE cohort treated with SoC to enable a long-term comparative analysis of belimumab versus SoC.

The primary objective was to compare organ damage progression (mean change in Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI) score) from baseline to Year 5 in patients treated with belimumab or SoC, using PS-matched data from the BLISS LTE study and the external cohort. Secondary objectives included comparing the time to organ damage progression and the magnitude of damage accrual.

METHODS

Identifying the comparator cohort

A systematic literature review was performed to identify cohorts, registries or other databases formed to support studies in SLE, with the objective of identifying an SLE comparison cohort with population characteristics similar to those of the BLISS LTE study population and an adequate sample with ≥5 years of follow-up. In total, 393 publications were identified referring to 92 cohorts, from which 21 cohorts/databases were selected for more in-depth evaluation, using the criteria of at least 400 patients and 3 peer-reviewed

publications. Data for each of the 21 cohorts were extracted to evaluate cohort size, ethnicity, age, duration of SLE, severity of disease activity, extent of organ damage progression, duration of follow-up, loss to follow-up, scope of data collection and data availability. The review identified the Toronto Lupus Cohort (TLC)^{10,11} as the preferred source of SoC data for this study, based on the size of the cohort, the extent of organ damage in the patients and the severity of SLE disease activity within the cohort. The TLC collected patient data at each visit and at 3–4-month intervals, and the scales used within the TLC for recording disease severity and organ damage progression were similar to those used within the BLISS LTE study.

Study design

This was a posthoc longitudinal PS-matched study comparing patients from the BLISS LTE study to clinically and demographically similar patients in the TLC. Baseline for the BLISS LTE study was the date of first exposure to belimumab (1 mg/kg or 10 mg/kg). For patients in the TLC, baseline was the first date they obtained an SLE Disease Activity Index-2000 (SLEDAI-2K) score ≥6, as this was an inclusion criterion for the BLISS LTE parent study (figure 1). Patients in the TLC did not receive belimumab because it was not available at the time.

Patients

The eligibility criteria used in the US BLISS LTE study were applied to patients in the TLC before PSM. Patients were ≥18 years of age, had a diagnosis of SLE using ≥4 of 11 American College of Rheumatology criteria (710.0)^{12,13} and a Safety of Estrogens in Lupus Erythematosus National Assessment–SLEDAI (SELENA-SLEDAI)/SLEDAI-2K score ≥6 at baseline and were autoantibody positive.⁵ Patients were excluded if they had active severe lupus nephritis or central nervous system lupus or if they had received B-cell target therapy at any time, if their baseline data preceded 1990

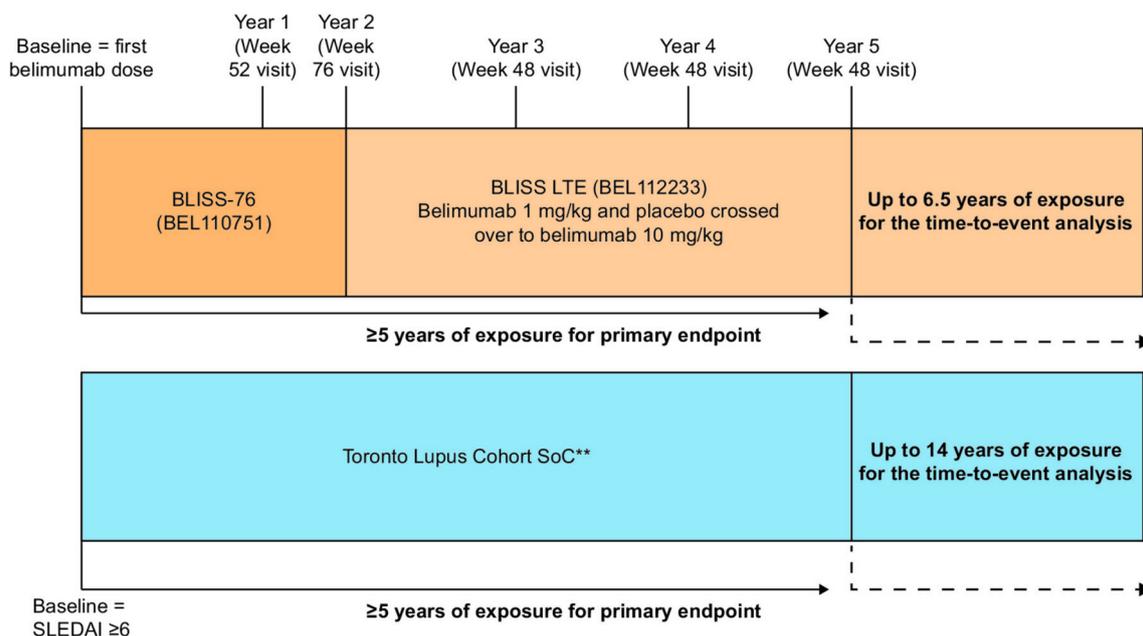


Figure 1 Study design and visit* assessments. *For the BLISS LTE study, the final visit in the parent study of the BLISS LTE study was recorded at 76 weeks. Thereafter, SDI was recorded every 48 weeks ('annual' visits). In the TLC, annual visits were defined as the visit closest to each 48-week interval from baseline that deviated by no more than 24 weeks from that interval. **Patients within the TLC had no exposure to belimumab as it was not available at the time. LTE, long-term extension; SDI, Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SLE, systemic lupus erythematosus; SLEDAI, SLE Disease Activity Index; SoC, standard of care; TLC, Toronto Lupus Cohort.

(to enhance the comparability of the period of treatment across groups) or if they had no visit within 24 weeks of their scheduled annual visit. The time to organ damage progression analysis included all patients with >1 year of follow-up and excluded TLC patients with ≥ 15 years of follow-up.

Choosing the PSM variables

A systematic literature review¹⁴ was used to identify publications that reported predictors of SLE organ damage and progression.^{15–18} Key predictors found in the literature (figure 2) were reviewed by a clinical expert (Professor Murray B Urowitz) and limited to those available in both the BLISS LTE study and the TLC. This generated a list of 14 predictors, which correlated to 17 operationalised variables used in the PSM analysis (figure 2).

Endpoints and assessments

The primary endpoint was the difference in change of total SDI score from baseline to 5 years between the BLISS LTE study and the TLC in patients with ≥ 5 years of follow-up. Secondary endpoints, assessed in patients with ≥ 1 year of follow-up, included time to first worsening (SDI score increase) in total SDI score, which used full longitudinal data for all patients (up to 6.5 years and 14 years within the BLISS LTE study and TLC cohorts, respectively). Magnitude of SDI score worsening was also assessed, and data from patients with ≥ 1 year of follow-up and up to 5 years of data were analysed. SDI score was taken from each visit within the BLISS LTE study and the TLC (figure 1).

Statistical analyses

PS were calculated using the logistic regression procedure in SAS V.9.4 software (SAS Institute, Cary, North Carolina, USA). The model specification included all potential predictor variables as independent variables.

The PS value for matching was defined as the estimated log-odds (ie, the $X\beta$ value) from the logistic regression, rather than the predicted probability, to enhance the range of variation in the PS distribution for matching. Patients from the BLISS LTE study were matched 1:1 to patients from the TLC based on similar PS values (within a calliper value defined as 20% of the SD for the distribution of the PS variable in the full sample). The standardised distance (bias) in the PS-matched sample across groups for all variables used to determine PS values ideally should be <5%; however, a standardised distance of <10% for all variables is considered adequate balance.¹⁹

An augmented regression model for the 5-year change in SDI score in the PS-matched samples was also estimated, adjusting for baseline corticosteroid dosage and decade of entry into the study to account for changes in the management of SLE over time.

For sensitivity analyses, the difference in change in total SDI score from baseline to 5 years across groups was evaluated using the inverse PS weighting (IPSW), method that uses the entire patient sample and the PS to weight the observations. The IPSW method aimed to confirm the robustness of the PSM method. Regression-augmented IPSW was also conducted as an additional sensitivity analysis to overcome any inadequate balance with the IPSW analysis, adding variables with bias >10% as covariates in the regression model. To assess the potential for nonlinearity in the magnitude of the 5 year change in SDI score, an ordered logistic regression model (SDI change equal to 0, 1 or 2+) was estimated using the PS-matched samples. Finally, changes from baseline in SDI organ damage system subscores were compared using Fisher's exact tests.

All inferential statistics were two-tail tests performed with an alpha of $p=0.05$. The difference in organ damage progression from baseline was evaluated using linear regression with change of total SDI score from baseline as the dependent variable and with a variable indicating treatment group (belimumab or SoC). Unbalanced matching variables were added as covariates. The difference in time to organ damage progression was analysed using parametric survival models with a binary indicator for treatment with belimumab as a covariate. Unbalanced matching variables were added as covariates. If statistically significant, the decade of entry into the study was also added as a covariate in both the difference in organ damage progression endpoint and time to organ damage progression analysis. The descriptive statistics for the magnitude of year-to-year organ damage progression were estimated at the end of Years 1–5 for both the belimumab and the SoC groups. Patient year-to-year increases in total SDI scores were classified as either +1 or ≥ 2 . A two-proportion z-test was used to determine whether there was a statistically significant difference between treatment cohorts in the proportion of total SDI increases that were ≥ 2 .

RESULTS

There were 259 patients from the BLISS LTE study and 706 patients from the TLC who met the inclusion criteria and were included in the PSM analysis.

PS-matched population for patients with 5 years of follow-up data (primary endpoint)

The range of PS distribution for patients with 5 years of follow-up was -9.927 to 4.701 . The range of common support (overlap in the PS distributions) for patients in the BLISS LTE study and the TLC was -3.648 to 2.893 (online supplementary figure S1A). With the calliper value of 0.53 (20% of the SD for the PS distribution), the range of support was -4.178 to 3.423 . Matching was not possible for 11 patients from the BLISS LTE study and 95 patients from the TLC with PS values outside of the range of support (including the calliper).

Using PSM, 99 patients from the BLISS LTE study and 99 patients from the TLC were 1:1 PS-matched from a larger pool of 567 patients (BLISS LTE $n=195$; TLC $n=372$). This sample was well balanced, with percentage bias <5% for 9 of 17 variables and <10% for all variables (mean bias=4.6%) (table 1).

PS-matched population for patients with ≥ 1 year of follow-up data

The range of PS distribution for patients from the BLISS LTE study and the TLC with ≥ 1 year of follow-up was -8.475 to 3.645 , and the range of overlap in the PS distributions was -3.928 to 2.171 (online supplementary figure S1B). With the calliper value of 0.40, the range of support was -4.328 to 2.571 . Matching was not possible for 13 BLISS LTE study patients and 246 TLC patients with PS values outside of the range of support (including the calliper).

Using PSM, 179 patients from the BLISS LTE study and 179 patients from the TLC were 1:1 PS-matched from a larger pool of 965 patients (BLISS LTE $n=259$; TLC $n=706$). This sample was well balanced, with percentage bias <5% for all but one variable and <10% for all variables (mean bias=2.2%) (table 2).

Difference in organ damage progression from baseline to 5 years

The change in SDI score from baseline to Year 5 for PS-matched patients was significantly lower for patients treated with belimumab compared with SoC (-0.434 ; 95% CI -0.667 to -0.201 ;

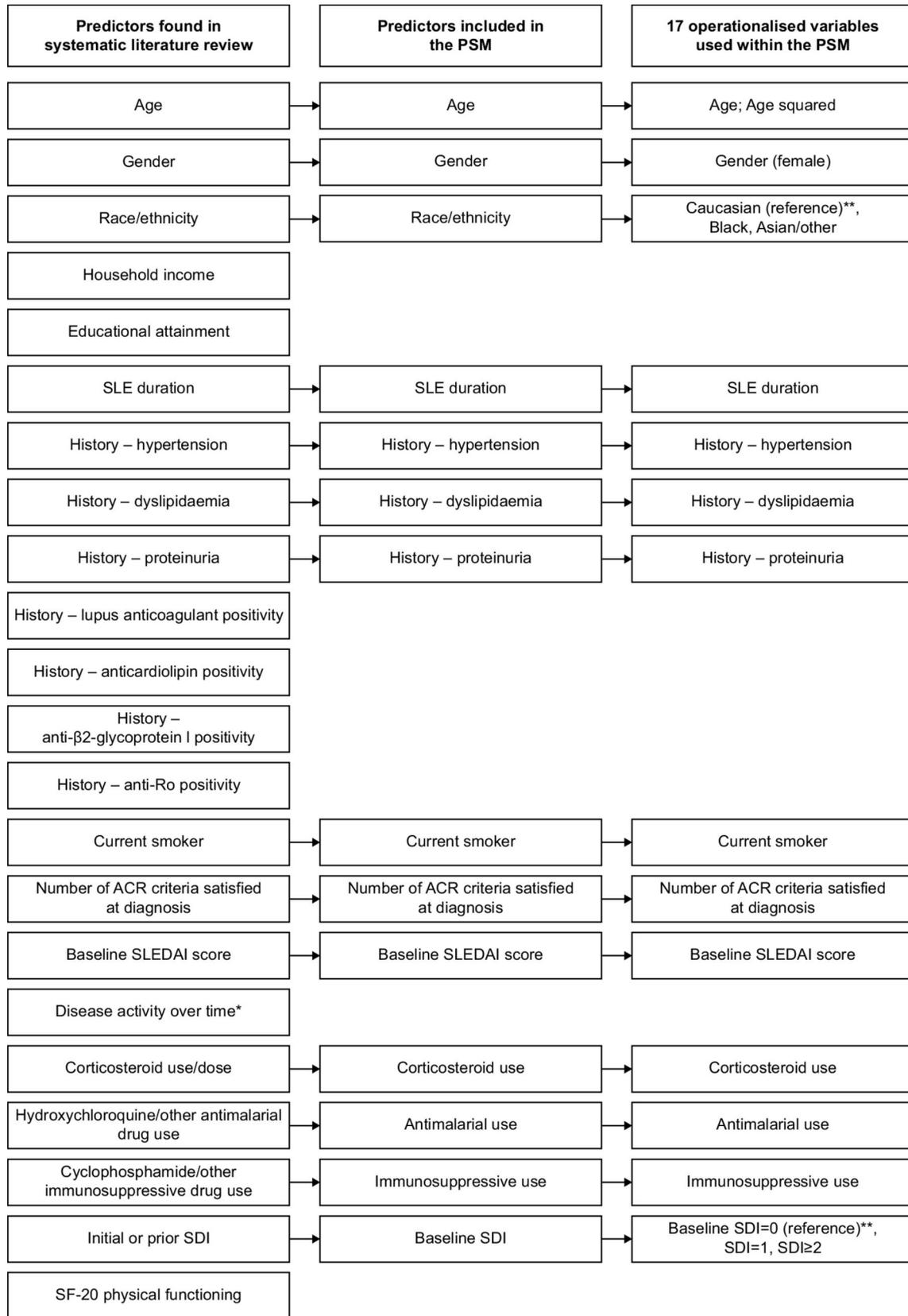


Figure 2 Predictors, variables and operationalised variables determined from a systematic literature review and used within the PSM analysis. *The variable 'disease activity over time' could not be used within the PSM as it was not a baseline variable. **The references for the 'race/ethnicity' variable and the 'baseline SDI' variable were 'Caucasian' and 'Baseline SDI=0', respectively. ACR, American College of Rheumatology; PSM, propensity score-matched; SDI, Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SF-20, 20-item short form survey; SLE, systemic lupus erythematosus; SLEDAI, SLE Disease Activity Index.

Table 1 Variables at baseline, pre-PSM and post-PSM for patients with 5 years of follow-up

Variable	Bias pre-PSM			Bias post-PSM		
	Belimumab (n=195)	SoC (n=372)	Bias (%)*	Belimumab (n=99)	SoC (n=99)	Bias (%)*
Mean age, years	42.8	37.3	45.5	40.0	39.0	8.4
Mean age squared, years	1947.4	1560.8	38.1	1733.0	1661.7	7.2
Female, %	92.8	89.5	11.6	92.9	91.9	3.8
Black, %	23.1	15.3	19.7	21.2	23.2	-4.8
Asian†/other race, %	9.2	23.4	-39.0	14.1	12.1	6.0
Mean SLE duration, years	7.9	5.8	30.0	7.4	7.6	-2.6
Smoker, %	3.6	23.7	-61.1	7.1	7.1	0.0
Hypertension‡, %	67.7	37.6	63.0	54.5	53.5	2.0
Dyslipidaemia, %	22.6	58.1	-77.5	28.3	31.3	-6.6
Proteinuria, %	12.3	31.7	-48.1	20.2	18.2	5.1
Number of ACR classification criteria satisfied	5.9	5.7	19.8	6.0	5.9	6.5
Baseline SLEDAI	7.8	10.1	-48.4	8.5	8.5	-2.2
Corticosteroid use, %	63.6	60.8	5.8	64.6	66.7	-4.2
Antimalarial use, %	73.8	51.9	46.6	69.7	68.7	2.2
Immunosuppressive use, %	53.8	31.5	46.4	45.5	44.4	2.0
SDI score=1, %	27.2	14.8	30.7	24.2	27.3	-6.9
SDI score ≥2, %	28.7	10.8	46.2	15.2	18.2	-8.1

*The extent of balance in clinical characteristics between groups was assessed using a standardised distance (bias) for each of the variables in the PS model. The standardised distance in the PS-matched sample across groups for all variables used to determine PS values ideally should be <5%; however, a standardised distance of <10% for all variables is considered adequate balance.

†Asian from the BLISS LTE population refers to patients with Central Asian, East Asian, Japanese, South Asian and Southeast Asian heritage; and Asian from the TLC refers to patients of Chinese heritage.

‡Patients who met any of the following criteria at baseline were defined as having hypertension: (1) systolic blood pressure ≥140 mm Hg or (2) diastolic pressure ≥90 mm Hg or (3) use of antihypertensive medications. The BLISS database had a flag for patients with baseline hypertension, but any BLISS patients who were not flagged were defined as having hypertension based on the same criteria used for TLC patients.

ACR, American College of Rheumatology; PS, propensity score; PSM, propensity score matching; SDI, Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SLE, systemic lupus erythematosus; SLEDAI, SLE Disease Activity Index; SoC, standard of care; TLC, Toronto Lupus Cohort.

$p < 0.001$) (table 3). The IPSW sensitivity analysis, using the patient full sample, also showed belimumab treatment to be associated with a smaller increase in SDI score compared with SoC (-0.441; 95% CI -0.669, to -0.222; $p < 0.001$); however, bias within this analysis was statistically inadequate (table 3), indicating the sample was not well balanced. Regression-augmented IPSW analysis, adding variables with bias >10% as covariates, produced similar results, with a smaller SDI score increase for patients treated with belimumab compared with SoC (-0.450; 95% CI -0.676 to -0.223; $p < 0.001$) (table 3).

An additional posthoc analysis was conducted to re-estimate the regression model for the 5 year change in SDI, to adjust for baseline corticosteroid dose and decade of entry into the study. In this augmented model, the estimated coefficient of the belimumab treatment variable remained essentially unchanged (-0.448; 95% CI -0.739 to -0.157; $p = 0.003$). A breakdown of the use of corticosteroids within the BLISS LTE and the TLC, by corticosteroid dose, has been provided in online supplementary table S1. A posthoc, regression-augmented model estimating the differences between groups in daily average cumulative corticosteroid usage through to Year 5, adjusted for decade of entry, indicated that cumulative corticosteroid usage was lower each day by 2.045 units (95% CI -3.625 to -0.465; $p = 0.011$) for patients treated with belimumab compared with SoC. When immunosuppressive medication use was added as a covariate in the 5 year SDI score change model for the PS-matched samples, the estimated belimumab coefficient remained essentially unchanged (-0.449; 95% CI -0.739 to -0.159); however, this was not statistically significant, and the variation in types of immunosuppressive medication used was not clinically meaningful (online supplementary table S2).

As the 5 year SDI change measure has a significant floor effect (ie, zero change) and does not necessarily increase in a linear manner, the analysis was re-estimated using an ordered logistic regression model (for response levels 0, 1 and 2+), using the PS-matched sample. The results indicated that patients treated with belimumab plus SoC were 60% less likely than patients from the TLC treated with SoC only to have a 5 year change in total SDI score. If patients treated with belimumab did experience a change, they were 60% less likely to have seen a change of more than 1 unit.

Changes in SDI subscores at Year 5 for patients treated with belimumab and patients treated with SoC were analysed, and the only subscores with significant differences were musculoskeletal (OR 0.163; 95% CI 0.030 to 0.599; $p < 0.003$) and skin (OR 0; 95% CI 0 to 0.559; $p = 0.007$). For patients treated with belimumab, the odds of experiencing an SDI increase from baseline in musculoskeletal system subscores were significantly smaller in the first year ($p = 0.010$) and continued to be significantly smaller for the intervening years versus those treated with SoC. Similarly, for patients treated with belimumab the odds of experiencing an SDI increase from baseline in skin subscore were significantly smaller for all but the first year.

Difference in time to organ damage progression

Patients receiving belimumab were 61% less likely to progress to a higher SDI score over any given year of follow-up compared with patients treated with SoC (HR 0.391; 95% CI 0.253 to 0.605; $p < 0.001$) (figure 3). A patient receiving belimumab had a 3.5% annual probability of organ damage progression compared with

Table 2 Variables at baseline, pre-PSM and post-PSM for patients with ≥1 year of follow-up

Variable	Bias pre-PSM			Bias post-PSM		
	Belimumab (n=259)	SoC (n=706)	Bias (%)*	Belimumab (n=179)	SoC (n=179)	Bias (%)*
Mean age, years	42.6	36.9	46.0	40.4	40.7	-2.4
Mean age squared, years	1937.4	1541.0	37.6	1763.4	1792.3	-3.0
Female, %	93.4	88.8	16.3	91.6	91.6	0.0
Black, %	21.6	14.6	18.3	22.3	23.5	-2.7
Asian†/other race, %	9.3	28.0	-49.6	12.8	12.8	0.0
Mean SLE duration, years	7.7	6.2	21.5	7.5	7.7	-3.2
Smoker, %	3.9	24.2	-61.2	5.6	6.7	-4.6
Hypertension‡, %	53.3	38.0	31.1	45.8	45.8	0.0
Dyslipidaemia, %	22.8	34.7	-26.5	25.1	22.9	5.2
Proteinuria, %	13.5	33.0	-47.4	16.8	17.9	-2.9
Number of ACR classification criteria satisfied	6.0	5.7	22.0	6.0	5.9	1.9
Baseline SLEDAI	7.9	10.0	-49.0	8.4	8.5	-3.7
Corticosteroid use, %	64.9	62.5	5.0	68.2	69.3	-2.4
Antimalarial use, %	71.8	56.4	32.6	65.9	67.0	-2.4
Immunosuppressive use, %	55.2	34.4	42.7	45.8	46.4	-1.1
SDI score=1, %	27.8	14.2	33.9	24.6	25.7	-2.6
SDI score ≥2, %	27.8	10.2	46.0	16.8	16.8	0.0

*The extent of balance in clinical characteristics between groups was assessed using a standardised distance (bias) for each of the variables in the PS model. The standardised distance in the PS-matched sample across groups for all variables used to determine PS values ideally should be <5%; however, a standardised distance of <10% for all variables is considered adequate balance.

†Asian from the BLISS LTE population refers to patients with Central Asian, East Asian, Japanese, South Asian and Southeast Asian heritage; and Asian from the TLC refers to patients of Chinese heritage.

‡Patients who met any of the following criteria at baseline were defined as having hypertension: (1) systolic blood pressure ≥140 mm Hg or (2) diastolic pressure ≥90 mm Hg or (3) use of antihypertensive medications. The BLISS database had a flag for patients with baseline hypertension, but any BLISS patients who were not flagged were defined as having hypertension based on the same criteria used for TLC patients.

ACR, American College of Rheumatology; PS, propensity score; PSM, propensity score matching; SDI, Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SLE, systemic lupus erythematosus; SLEDAI, SLE Disease Activity Index; SoC, standard of care; TLC, Toronto Lupus Cohort.

an 8.7% annual probability of progression with SoC alone. There were no statistically significant differences in the HRs between the 1 mg/kg and 10 mg/kg doses from the parent study.

Table 3 Change in SDI from baseline to 5 years using PSM, IPSW and regression augmented IPSW

Method/variable	SoC	Belimumab	Difference
PSM sample			
n	99	99	
5 year SDI change, mean (SE)	0.717	0.283	-0.434 (0.119)
95% CI	0.500 to 0.934	0.166 to 0.400	-0.667 to -0.201
p values			p<0.001
IPSW sample			
n	372	195	
5 year SDI change, mean (SE)	0.777	0.336	-0.441 (0.116)
95% CI	0.607 to 0.947	0.184 to 0.488	-0.669 to -0.222
p values			p<0.001
Regression augmented IPSW sample			
n	372	195	
5 year SDI change, mean (SE)	0.782	0.333	-0.450 (0.116)
95% CI	0.630 to 0.935	0.167 to 0.498	-0.676 to -0.223
p values			p<0.001

CI, confidence interval; IPSW, inverse propensity score weighting; PSM, propensity score matching; SDI, Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index; SE, standard error; SoC, standard of care.

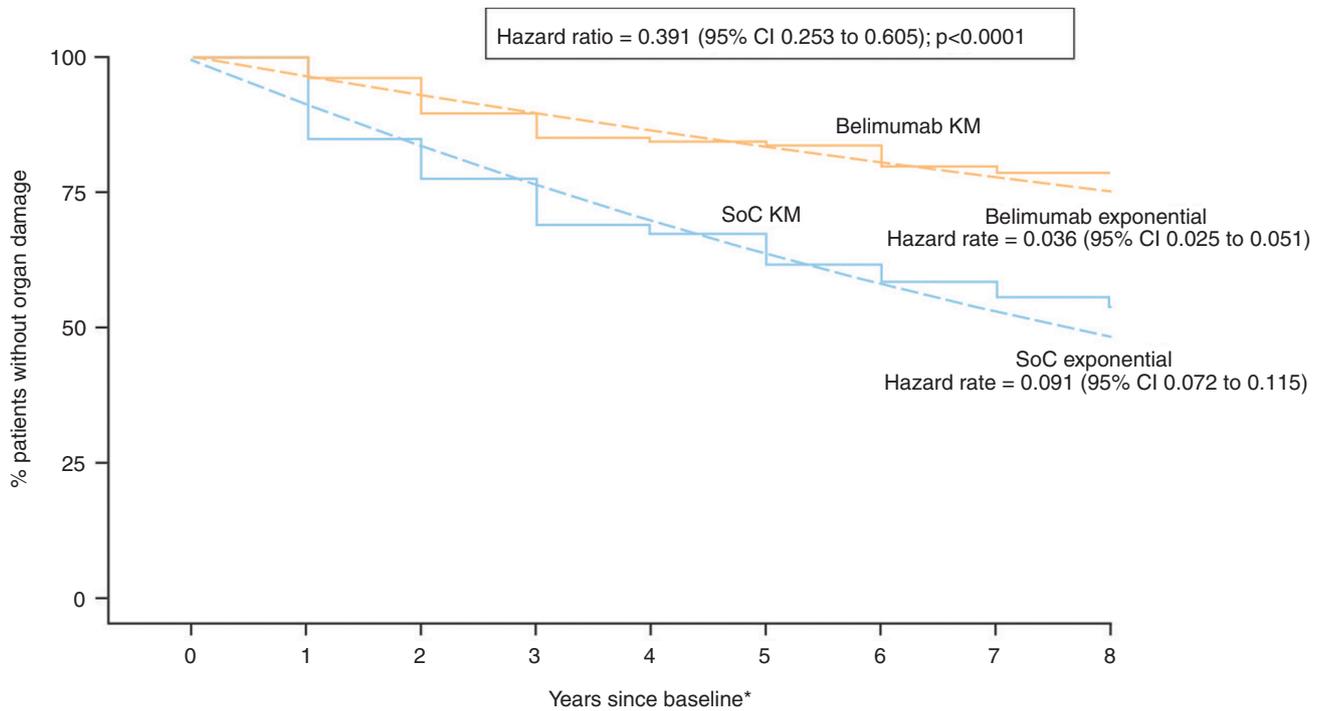
Magnitude of year-to-year organ damage progression

Of the patients treated with belimumab, there were 33 instances of an SDI score increase of ≥1 compared with 72 instances in patients treated with SoC. Of these, the proportion of SDI score increases ≥2 over any given year of follow-up was five times greater with SoC (n=22/72, 30.56%) compared with belimumab (n=2/33, 6.06%; p=0.006).

DISCUSSION

This PSM analysis matched patients within the BLISS LTE study 1:1 with similar patients from the TLC who had not received belimumab. This enabled the results from the two cohorts to be effectively and credibly compared in terms of organ damage progression, time to organ damage progression and magnitude of progression.

The study demonstrated that, over a 5-year period, patients treated with belimumab experienced less organ damage compared with patients treated with SoC alone. The IPSW and regression-augmented IPSW results were similar to the PS-matched results, which demonstrates the robustness of the findings across alternative PS adjustment methodologies. Patients treated with belimumab were 61% less likely to progress to a higher SDI score over any given year of follow-up compared with patients treated with SoC, indicating that belimumab appears to slow the rate of organ damage progression compared with SoC. In addition, a higher proportion of patients treated with SoC experienced an SDI score increase ≥2 compared with patients treated with belimumab. These results are in accordance with previous studies. An interim analysis of the two BLISS LTE studies demonstrated that 85.1% of patients treated with belimumab at Years 5–6 (n=403) showed no change from baseline



Total number of patients at risk at each time point

Belimumab	179	166	137	121	112	88	65	31	0
SoC	179	135	111	83	71	57	44	30	28

Figure 3 Difference in time to organ damage progression in patients with ≥ 1 year of follow-up. *Years are 48 weeks in length. KM, Kaplan-Meier; SLE, systemic lupus erythematosus; SoC, standard of care.

in SDI score, 11.4% experienced an SDI score increase of +1 and 3.2% experienced an SDI score increase of ≥ 2 .²⁰ These results, demonstrating slow rates of organ damage progression in patients treated with belimumab compare favourably with organ damage progression rates reported by other SLE cohorts, such as the SLICC inception cohort,²¹ the Tromsø cohort¹⁶ and the Lupus in Minorities Nature vs Nurture (LUMINA) cohort;¹⁵ however, it should be noted that the BLISS studies excluded patients with severe lupus nephritis and central nervous system disease.^{13 5}

The quality of the systematic literature review preceding the PSM analysis was essential in determining the appropriate, clinically reviewed variables to use within the PSM analysis to effectively and accurately match patients within the BLISS LTE study and the TLC. The operationalised variables used within the PSM analysis captured both clinical demographics and disease characteristics that enabled patients to be adequately matched, thus reducing the potential for confounders in the assessment of comparative treatment effectiveness. The PSM process, using a 20% calliper, produced well-balanced samples of patients across the BLISS LTE study and the TLC as shown by the low mean bias for all variables following PSM.

PSM has been used previously in patients with SLE. Ruiz-Irastorza *et al*²² conducted an observational prospective cohort study to determine the effect of antimalarials on thrombosis and survival in patients with SLE. Patients were divided by whether they had previously been treated with antimalarials and were PS-matched based on 24 variables related to disease severity and known prognostic significance.²² PSM enabled a comparison of patients treated with antimalarials with patients without a history of antimalarial medication, while adjusting for confounders

that would otherwise have biased the results (as it would have been unethical to randomise patients, treatment was allocated according to clinical judgement). PSM was also used to compare live patients (controls) in the LUMINA cohort with deceased patients (cases) to determine the protective effect of hydroxychloroquine on survival in patients with SLE.⁸ Patients were sorted into PS quintiles based on their likelihood of being treated with hydroxychloroquine. It was expected that patients with milder disease had a higher probability of being treated with hydroxychloroquine than those with more severe disease; however, as each PS quintile included both treated and untreated patients, the study essentially achieved de facto pseudorandomisation. Here, PSM was effectively used to adjust for patients' characteristics at baseline and to act as a single variable in multivariable analyses.^{8 23}

LTE studies with no comparator are routinely used to explore the long-term efficacy and safety of a treatment beyond the limited duration of controlled double-blinded trials; this study demonstrates the utility and value of PSM in providing a comparator for a treatment observed within an LTE. The results suggest that this methodology could be applied more frequently to reinforce the conclusions that can be drawn from these studies.

Limitations

PSM can only match patients based on known variables; other differences may exist between matched populations that cannot be observed, which may cause some degree of residual confounding.^{6 23} A strength of this study is that the patients in the TLC who were otherwise indicated for belimumab treatment did not receive belimumab solely because it was not available at the time. In addition,

although only a moderate number of patients were matched within this analysis, the IPSW and regression-augmented IPSW sensitivity analyses, which used the whole available population samples, produced similar results to the main PS-matched analysis, thus confirming the credibility of the original PS-matched analysis.

One variable that could not be matched across groups was the year of entry into the study. To enhance comparability, patients in the TLC database with baseline dates preceding 1990 were excluded from the study sample. Nonetheless, the PS-matched analysis could be confounded by remaining changes in SoC over time between the two study periods. Although no revolutionary changes in SLE treatment occurred during the study period, there may have been evolutionary changes in the management of SLE over this period. For example, the proportion of patients using corticosteroids was well balanced; but among patients using corticosteroids, those treated with belimumab were more likely to use a lower dosage compared with patients treated with SoC. Similarly, among patients using immunosuppressive medications, there were some differences in the specific immunosuppressive medication used. Nevertheless, the PS-matched model was re-estimated to adjust for baseline corticosteroid dose, immunosuppressive use and decade of study entry and the change in SDI score from baseline to Year 5 for PS-matched patients was similar to the primary PS-matched analysis (reported in table 1), which did not adjust for these factors. Thus, the evolution in the management of SLE over this time period does not appear to have had an impact of a magnitude sufficient to confound the results of the PSM analysis.

The endpoints presented here made use only of the USA LTE dataset. Therefore, the generalisability of these results to other countries may be in question and may require further studies.

CONCLUSION

This PS-matched analysis produced adequate, well-balanced patient samples using predictors of disease and clinical characteristics as variables within the analysis. PSM enabled effective and credible comparison of patients from the BLISS LTE study to patients from the TLC. Patients receiving belimumab had significantly less SLE-related organ damage progression over 5 years compared with patients in the TLC receiving SoC only. Similarly, the study found significantly slower organ damage progression, and a smaller magnitude of progression, in patients treated with belimumab compared with SoC. Future studies may use PSM methodology to reinforce conclusions that can be drawn from long-term, open-label studies.

Acknowledgements The authors would like to acknowledge John J Dever and Mehdi Zakerifar for their contributions to this study. Medical writing support was provided by Emma Hargreaves, Fishawack Indicia Ltd, UK, and was funded by GSK.

Contributors MBU, RLO, RCW, KAK, YA and SR contributed to the study conception/design and data analysis/interpretation. SR contributed to the acquisition of data.

Funding This study (206347) was funded by GSK.

Competing interests MBU has received research grants from GSK. YA and SR are employees of GSK and hold shares in the company. RCW and KAK are employees of Medical Decision Modeling and RLO is a non-employee consultant for Medical Decision Modeling.

Patient consent Not required.

Ethics approval The original parent study (BEL110751) and the BLISS LTE study (BEL112233) were conducted in accordance with the ethical principles of the Declaration of Helsinki and the International Council for Harmonisation on Good Clinical Practice and with applicable country-specific regulatory requirements. Ethics approval was not required for this study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement GSK makes available anonymised individual participant data and associated documents from interventional clinical studies which evaluate medicines, on approval of proposals submitted to www.clinicalstudydatarequest.com. To access data for other types of GSK sponsored research, for study documents without patient-level data and for clinical studies not listed, please submit an enquiry via the website.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Navarra SV, Guzmán RM, Gallacher AE, *et al*. Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. *Lancet* 2011;377:721–31.
- Zhang F, Bae SC, Bass D, *et al*. A pivotal phase III, randomised, placebo-controlled study of belimumab in patients with systemic lupus erythematosus located in China, Japan and South Korea. *Ann Rheum Dis* 2018;77:355–63.
- Furie R, Petri M, Zamani O, *et al*. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. *Arthritis Rheum* 2011;63:3918–30.
- Stohl W, Schwarting A, Okada M, *et al*. Efficacy and safety of subcutaneous belimumab in systemic lupus erythematosus: a fifty-two-week randomized, double-blind, placebo-controlled study. *Arthritis Rheumatol* 2017;69:1016–27.
- Furie RA, Wallace DJ, Aranow C, *et al*. Long-term safety and efficacy of belimumab in patients with systemic lupus erythematosus: a continuation of a seventy-six-week phase III parent study in the United States. *Arthritis Rheumatol* 2018;70:868–77.
- D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998;17:2265–81.
- Luo Z, Gardiner JC, Bradley CJ. Applying propensity score methods in medical research: pitfalls and prospects. *Med Care Res Rev* 2010;67:528–54.
- Alarcón GS, McGwin G, Bertoli AM, *et al*. Effect of hydroxychloroquine on the survival of patients with systemic lupus erythematosus: data from LUMINA, a multiethnic US cohort (LUMINA L). *Ann Rheum Dis* 2007;66:1168–72.
- Li T, Carls GS, Panopalis P, *et al*. Long-term medical costs and resource utilization in systemic lupus erythematosus and lupus nephritis: a five-year analysis of a large medicaid population. *Arthritis Rheum* 2009;61:755–63.
- Gladman DD, Urowitz MB, Rahman P, *et al*. Accrual of organ damage over time in patients with systemic lupus erythematosus. *J Rheumatol* 2003;30:1955–9.
- Urowitz MB, Gladman DD, Tom BD, *et al*. Changing patterns in mortality and disease outcomes for patients with systemic lupus erythematosus. *J Rheumatol* 2008;35:2152–8.
- Hochberg MC. Updating the American college of rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997;40:1725.
- Tan EM, Cohen AS, Fries JF, *et al*. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982;25:1271–7.
- Sutton EJ, Davidson JE, Bruce IN. The systemic lupus international collaborating clinics (SLICC) damage index: a systematic literature review. *Semin Arthritis Rheum* 2013;43:352–61.
- Alarcón GS, Roseman JM, McGwin G, *et al*. Systemic lupus erythematosus in three ethnic groups. XX. Damage as a predictor of further damage. *Rheumatology* 2004;43:202–5.
- Becker-Merok A, Nossent HC. Damage accumulation in systemic lupus erythematosus and its relation to disease activity and mortality. *J Rheumatol* 2006;33:1570–7.
- Petri M, Purvey S, Fang H, *et al*. Predictors of organ damage in systemic lupus erythematosus: the hopkins lupus cohort. *Arthritis Rheum* 2012;64:4021–8.
- Stoll T, Sutcliffe N, Mach J, *et al*. Analysis of the relationship between disease activity and damage in patients with systemic lupus erythematosus—a 5-yr prospective study. *Rheumatology* 2004;43:1039–44.
- Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011;46:399–424.
- Bruce IN, Urowitz M, van Vollenhoven R, *et al*. Long-term organ damage accrual and safety in patients with SLE treated with belimumab plus standard of care. *Lupus* 2016;25:699–709.
- Bruce IN, O'Keefe AG, Farewell V, *et al*. Factors associated with damage accrual in patients with systemic lupus erythematosus: results from the Systemic Lupus International Collaborating Clinics (SLICC) Inception Cohort. *Ann Rheum Dis* 2015;74:1706–13.
- Ruiz-Irastorza G, Egurbide MV, Pijoan JI, *et al*. Effect of antimalarials on thrombosis and survival in patients with systemic lupus erythematosus. *Lupus* 2006;15:577–83.
- Landewé RB. The benefits of early treatment in rheumatoid arthritis: confounding by indication, and the issue of timing. *Arthritis Rheum* 2003;48:1–5.