



OPEN ACCESS

## EXTENDED REPORT

## Patient characteristics influence the choice of biological drug in RA, and will make non-TNFi biologics appear more harmful than TNFi biologics

Thomas Frisell,<sup>1</sup> Eva Baecklund,<sup>2</sup> Karin Bengtsson,<sup>3</sup> Daniela Di Giuseppe,<sup>1</sup> Helena Forsblad-d'Elia,<sup>4</sup> Johan Askling,<sup>1,5</sup> on behalf of the ARTIS Study group

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-2017-212395>).

<sup>1</sup>Clinical Epidemiology Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

<sup>2</sup>Rheumatology Unit, Department of Medical Sciences, Uppsala University, Uppsala, Sweden

<sup>3</sup>Department of Rheumatology and Inflammation Research, Institute of Medicine, Sahlgrenska Academy at University of Gothenburg, Göteborg, Sweden

<sup>4</sup>Department of Public Health and Clinical Medicine, Rheumatology, Umeå University, Umeå, Sweden

<sup>5</sup>Department of Rheumatology, Karolinska University Hospital, Stockholm, Sweden

## Correspondence to

Dr Thomas Frisell, Clinical Epidemiology Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm 17176, Sweden; [thomas.frisell@ki.se](mailto:thomas.frisell@ki.se)

Received 15 September 2017

Revised 23 November 2017

Accepted 27 November 2017

Published Online First

13 December 2017

## ABSTRACT

**Objectives** With the wide range of biological disease-modifying anti-rheumatic drugs (bDMARDs) available for treating rheumatoid arthritis (RA), and limited evidence to guide the choice for individual patients, we wished to evaluate whether patient characteristics influence the choice of bDMARD in clinical practice, and to quantify the extent to which this would bias direct comparisons of treatment outcome.

**Methods** Register-based study of all Swedish patients with RA initiating necrosis factor inhibitor (TNFi), rituximab, abatacept or tocilizumab in 2011–2015 as their first bDMARD (n=6481), or after switch from TNFi as first bDMARD (n=2829). Group differences in demographics, clinical characteristics and medical history were assessed in multivariable regression models. Predicted differences in safety and treatment outcomes were calculated as a function of patient characteristics, through regression modelling based on observed outcomes among patients with RA starting bDMARDs 2006–2010.

**Results** Patients starting non-TNFi were older than those starting TNFi, had lower socioeconomic status, higher disease activity and higher burden of diseases including malignancy, serious infections and diabetes. Differences were most pronounced at first bDMARD initiation. These factors were linked to treatment outcome independent of therapy, yielding worse apparent safety and effectiveness for non-TNFi biologics, most extreme for rituximab. Standardising to the age/sex distribution of the TNFi group reduced differences considerably.

**Conclusions** There was significant channelling of older and less healthy patients with RA to non-TNFi bDMARDs, in particular as first bDMARD. Whether this channelling represents a maximised benefit/risk ratio is unclear. Unless differences in age, medical history and disease activity are accounted for, they will substantially confound non-randomised comparative studies of available bDMARDs' safety and effectiveness.

## BACKGROUND

Following a rapid development over the past two decades, a wide range of biological disease-modifying anti-rheumatic drugs (bDMARDs) are currently available for the treatment of rheumatoid arthritis (RA). In many countries, Sweden included, RA treatment guidelines have expanded the recommended options for first bDMARD in recent years, from necrosis factor inhibitor (TNFi)

drugs exclusively to include abatacept, tocilizumab and rituximab,<sup>1–4</sup> ranking the drugs as comparable in overall safety and efficacy. For historical reasons, TNFi drugs remain the most common choice as first bDMARD, but many patients will switch from their initial bDMARD,<sup>5</sup> and similar to the first, the choice of the next bDMARD (eg, switching to another TNFi or to another mode of action) is seldom strictly regulated. In clinical practice, perceived or established differences between bDMARD options lead to a non-random allocation of treatment. Although many clinicians may be aware of the *existence* of such channelling, its *magnitude* (ie, how different treatment outcomes it will give rise to) is seldom quantified, yet essential for a correct evaluation of the drugs' relative effectiveness and safety.<sup>6,7</sup>

In all situations with an element of preference-guided choice of therapy, it is important to monitor which patient gets which therapy for at least two reasons. First, if physicians' show a preference for a specific drug for certain patients, it should be a research priority to tell whether this is motivated by an increased tolerability or efficacy in this group, or merely a misconception about the drug's (side) effects, potentially leading to inequities in care.<sup>8</sup> Second, non-random choice of therapy will hamper studies of RA therapies by introducing *confounding by indication*, which occurs when factors associated with the choice of therapy are also predictors of the studied outcome, and is generally considered the major limitation of non-randomised comparisons of therapies.<sup>9</sup> The case of bDMARDs in RA illustrates both of these needs.

Thus, the objective of this paper is twofold. First, to describe baseline patient characteristics at initiation of different bDMARDs at two clinically distinct and common time points: (1) at first bDMARD initiation, (2) at switch to a second bDMARD after having used a TNFi as first bDMARD. Second, to estimate the potential of the observed channelling to confound comparative safety and effectiveness studies in RA.

## METHODS

Data on clinical characteristics, demographics and medical history among all patients with RA in Sweden who initiated a first or second bDMARD therapy during 2011–2015 were identified by linking the Swedish Rheumatology Quality register (SRQ) to nationwide Swedish healthcare registers.

Check for updates

**To cite:** Frisell T, Baecklund E, Bengtsson K, et al. *Ann Rheum Dis* 2018;**77**:650–657.

## Data sources

The database used for this study has been described previously.<sup>10–11</sup> Briefly, the SRQ is a clinical register with longitudinal data on disease activity and treatment at each rheumatology visit,<sup>12</sup> with a national coverage for bDMARD treatment in RA of 95%.<sup>13</sup> The National Patient Register provided all diagnoses set in inpatient and non-primary outpatient visits; validation against medical files has found a high positive predictive value (85%–95%) for diagnoses in inpatient care.<sup>14–15</sup> The Prescribed Drug Register provided all dispensed prescriptions in Sweden since July 2005; the register has virtually complete coverage.<sup>16</sup> The Swedish Cancer Register contains clinical data on all cancers since 1958; estimated coverage is greater than 95%.<sup>17</sup> Registers on communicable diseases provided dates for verified tuberculosis, hepatitis B and C. Socioeconomic data were available through census registers.

## Covariates

We considered an inclusive list of baseline characteristics to capture factors that we a priori considered may influence choice of therapy, safety or treatment outcome. Variables included sociodemographic background (highest education, country of birth), RA-specific clinical characteristics (rheumatoid factor (RF), disease duration, Health Assessment Questionnaire - Disability Index (HAQ-DI), Disease Activity Score - 28 joints (DAS28) with components, visual analogue scale (VAS) pain), concomitant treatment with non-steroidal anti-inflammatory drugs (NSAIDs), glucocorticosteroids, conventional synthetic DMARDs and medical history. Disease activity and current therapy was extracted from the visit in the SRQ with valid data on each variable closest to treatment start (within -90 to +14 days, chosen to increase data availability, while avoiding values influenced by the treatment effect). Medical history was measured as having been diagnosed with either of a range of specific conditions (definitions in online supplementary table s1), within 5 years before treatment start, except for serious infections (defined as 'recent' within 1 year, and 'non-recent' within 1 to 5 years) and malignancy ('recent' within 5 years, and 'non-recent' more than 5 years earlier). Analysis of individual conditions was preferred over a combined comorbidity score since the latter would mask disease-specific associations and increase risk for residual confounding.<sup>18</sup> We used three continuous measures intended to capture patients' general health: (1) number of days hospitalised, (2) days of lost work due to sick leave or disability pension (only for those aged 25–65 years) and (3) total healthcare costs. Healthcare costs were calculated by summing costs for dispensed drugs and visits in inpatient and non-primary outpatient care, weighted by disease-related group with annual national tariffs, inflation corrected to 2012.

## Statistics

To assess differences in patient characteristics across biologics, we tabulated means and proportions of baseline covariates, with adjusted differences for each non-TNFi bDMARD compared with TNFi, modelled in multivariable linear regression models with bootstrapped CIs.<sup>19–20</sup> The main model was adjusted for sex, age and geographic region, and a supplemental model further adjusted for country of birth, education level, RF, disease duration, erythrocyte sedimentation rate (ESR), DAS28 calculated with C-reactive protein (DAS28-CRP), recent infections, recent malignancy, joint surgery, chronic lung disease and acute coronary syndrome. The choice of covariates in model 2 was based on observed differences and availability of data.

Therapy after switch from TNFi was defined as the first bDMARD therapy started within 1 year of discontinuing an initial TNFi as the first ever bDMARD. The main analysis focused on the difference between abatacept, rituximab, tocilizumab and the class of TNFi (adalimumab, certolizumab pegol, etanercept, infliximab and golimumab). Supplementary analyses were performed comparing individual TNFi drugs.

The expected impact of confounding was assessed through a series of prediction models. Logistic regressions were used to estimate associations of patient characteristics and treatment outcomes among all individuals with RA starting a bDMARD (to maximise cohort size and precision, we included up to third bDMARD) in the years 2006–2010 (immediately prior to our study period). We defined safety outcomes as the proportions experiencing the following events within 5 years of starting therapy: (1) death, (2) serious infection, (3) major acute cardiovascular event (MACE), (4) non-benign malignancy (definitions in supplementary table s2). Similarly, we defined treatment effectiveness outcomes as the proportion: (1) discontinuing therapy before 1 year and (2) remaining on therapy and having reached good European League Against Rheumatism (EULAR) response after 1 year. Separate models were created for each outcome using the full list of covariates. To allow some deviation from linearity, continuous variables were entered as second-degree polynomials; the only included interaction was between age and sex. Line of therapy was included as a binary variable (biologics naive vs not). Work loss was excluded from model building, since it is restricted to those of working age. The coefficients from the final models were used to calculate the predicted probability of each outcome, by treatment, in our main cohort. Since specific bDMARDs were not included in the prediction models, these predicted probabilities will reflect the proportions expected only from baseline characteristics, averaged across all bDMARDs. To further assess how much of the predicted difference between treatments would be removed by adjustment for age and sex rather than by other patient characteristics (eg, medical history), we standardised each treatment group to the age (in 10-year categories) and sex distribution in the largest group (TNFi as first bDMARD).

Linear regression with bootstrapped CIs was made using the boot package in R V.3.3.1. SAS V.9.4 was used for all other analyses.

## RESULTS

### Patient characteristics at start of first bDMARD

We identified 6481 patients with RA starting a first bDMARD between 1 January 2011 and 31 December 2015. Most started a TNFi (n=5307, 82%), with all available TNFi in common use, ranging from etanercept (n=1502, 28% of all TNFi) to golimumab (n=745, 14%). The most common non-TNFi was rituximab (n=655, 10% of all first bDMARD). Demographical and clinical characteristics are shown in table 1. Initiators of non-TNFi therapy were older and less well educated than those starting a TNFi, with largest difference for rituximab. Compared with those starting TNFi, rituximab initiators were also more often seropositive, had longer disease duration and slightly higher ESR. Abatacept initiators were similar to the TNFi group, but had higher ESR. Tocilizumab initiators were most extreme in terms of disease activity, with significantly higher ESR and CRP, and borderline higher tender joint counts and HAQ. Initiators of non-TNFis had lower use of concomitant methotrexate.

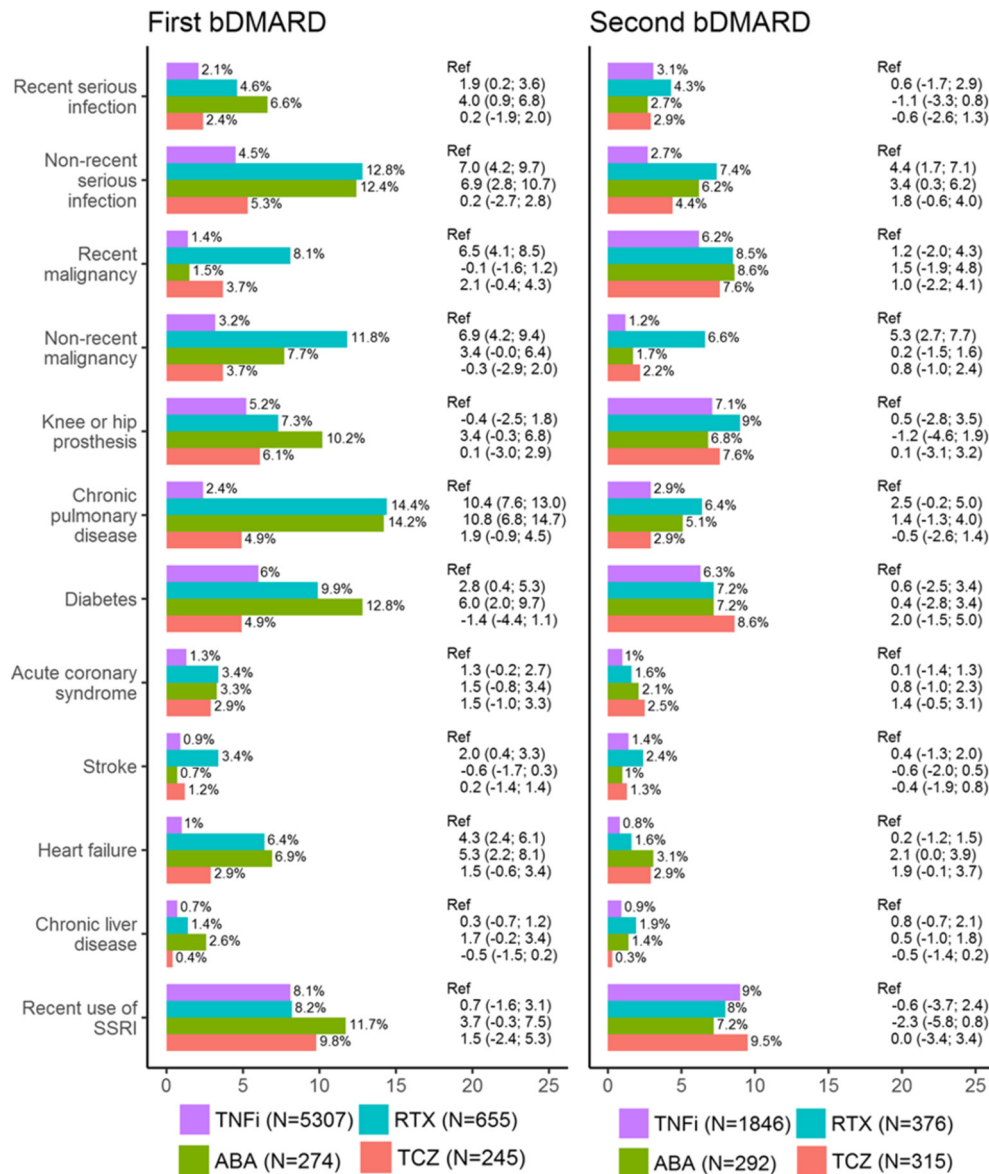
There were substantial differences in baseline medical history, with those starting rituximab or abatacept more often having

**Table 1** Patient characteristics at start of bDMARD therapy, among all bionative Swedish patients with RA, 2011–2015

|  | TNFI          | Rituximab     | Abatacept     | Mean difference versus TNFI (95% CI) | Abatacept     | Mean difference versus TNFI (95% CI) | Tocilizumab   | Mean difference versus TNFI (95% CI) |
|--|---------------|---------------|---------------|--------------------------------------|---------------|--------------------------------------|---------------|--------------------------------------|
| Patients, n                                    | 5307          | 655           | 274           |                                      | 274           |                                      | 245           |                                      |
| <b>Demographics</b>                            |               |               |               |                                      |               |                                      |               |                                      |
| Age, mean (SD)                                 | 55.2 (14.1)   | 65.1 (12.5)   | 61.4 (13.3)   | 10.0 (9.0 to 11.1)                   | 61.4 (13.3)   | 6.2 (4.7 to 7.8)                     | 58.5 (13.7)   | 3.3 (1.7 to 5.1)                     |
| Female, n (%)                                  | 3929 (74.0)   | 471 (71.9)    | 202 (73.7)    | 1.3 (-2.4 to 5.3)                    | 202 (73.7)    | 1.7 (-3.4 to 7.2)                    | 190 (77.6)    | 4.5 (-0.7 to 10.1)                   |
| <b>Highest education, n (%)</b>                |               |               |               |                                      |               |                                      |               |                                      |
| 9 years or less                                | 908 (21.2)    | 165 (33.0)    | 60 (29.0)     | 5.7 (1.6 to 10.1)                    | 60 (29.0)     | 4.2 (-1.5 to 10.3)                   | 44 (25.1)     | 2.0 (-4.8 to 8.5)                    |
| 10 to 12 years                                 | 2091 (48.9)   | 219 (43.8)    | 102 (49.3)    | -2.6 (-7.4 to 1.9)                   | 102 (49.3)    | 1.2 (-5.5 to 8.2)                    | 87 (49.7)     | 2.0 (-5.5 to 9.9)                    |
| >12 years                                      | 1279 (29.9)   | 116 (23.2)    | 45 (21.7)     | -3.1 (-7.1 to 0.9)                   | 45 (21.7)     | -5.4 (-11.1 to 0.3)                  | 44 (25.1)     | -4.0 (-10.6 to 2.7)                  |
| <b>Country of birth, n (%)</b>                 |               |               |               |                                      |               |                                      |               |                                      |
| Sweden   | 485 (9.1)     | 82 (12.5)     | 21 (7.7)      | -6.9 (-10.1 to -3.3)                 | 21 (7.7)      | -1.6 (-5.8 to 2.7)                   | 27 (11.0)     | -1.6 (-6.3 to 3.1)                   |
| Scandinavia                                    | 267 (5.0)     | 54 (8.2)      | 19 (6.9)      | 1.6 (-0.6 to 3.9)                    | 19 (6.9)      | 1.1 (-2.2 to 4.1)                    | 13 (5.3)      | -0.3 (-3.3 to 2.4)                   |
| Other  | 4555 (85.8)   | 519 (79.2)    | 234 (85.4)    | 5.3 (2.5 to 7.8)                     | 234 (85.4)    | 0.5 (-2.9 to 3.5)                    | 205 (83.7)    | 1.9 (-2.3 to 5.7)                    |
| <b>Medical history</b>                         |               |               |               |                                      |               |                                      |               |                                      |
| Days hospitalised last 5 years, mean (SD)      | 5.5 (16.5)    | 15.2 (30.5)   | 14.9 (29.0)   | 8.0 (5.4 to 10.2)                    | 14.9 (29.0)   | 8.4 (4.7 to 11.5)                    | 7.5 (16.6)    | 1.1 (-1.2 to 3.1)                    |
| Days of work loss last 5 years, mean (SD)      | 363.1 (588.2) | 587.2 (703.2) | 560.2 (715.8) | 183.0 (103.1 to 264.7)               | 560.2 (715.8) | 151.1 (45.6 to 261.9)                | 407.9 (651.3) | 20.4 (-77.7 to 116.2)                |
| Healthcare costs last 5 years, TSEK, mean (SD) | 123.3 (133.4) | 228.0 (234.2) | 213.0 (216.0) | 85.4 (66.6 to 102.6)                 | 213.0 (216.0) | 77.6 (52.5 to 103.3)                 | 136.0 (143.1) | 2.5 (-16.7 to 20.7)                  |
| <b>RA clinical characteristics</b>             |               |               |               |                                      |               |                                      |               |                                      |
| Rheumatoid factor positive (%)                 | 3886 (75.2)   | 566 (88.2)    | 206 (76.6)    | 10.4 (7.6 to 13.3)                   | 206 (76.6)    | -0.5 (-6.0 to 4.9)                   | 182 (75.2)    | -0.7 (-6.4 to 4.9)                   |
| Disease duration, years, mean (SD)             | 9.1 (9.8)     | 12.9 (12.7)   | 9.6 (11.6)    | 2.1 (1.0 to 3.1)                     | 9.6 (11.6)    | -0.7 (-2.0 to 0.6)                   | 7.3 (10.2)    | -2.3 (-3.6 to -1.1)                  |
| HAQ, mean (SD)                                 | 1.02 (0.61)   | 1.16 (0.66)   | 1.12 (0.61)   | 0.06 (-0.00 to 0.13)                 | 1.12 (0.61)   | 0.07 (-0.02 to 0.16)                 | 1.14 (0.55)   | 0.10 (0.02 to 0.17)                  |
| DAS28, mean (SD)                               | 4.70 (1.28)   | 4.83 (1.28)   | 4.95 (1.23)   | -0.04 (-0.17 to 0.09)                | 4.95 (1.23)   | 0.16 (-0.01 to 0.34)                 | 5.16 (1.13)   | 0.39 (0.23 to 0.55)                  |
| Tender joints, 0–28, mean (SD)                 | 6.9 (5.6)     | 6.7 (5.6)     | 7.0 (5.5)     | -0.1 (-0.6 to 0.5)                   | 7.0 (5.5)     | 0.2 (-0.6 to 0.9)                    | 7.9 (6.0)     | 0.9 (-0.0 to 1.7)                    |
| Swollen joints, 0–28, mean (SD)                | 6.3 (4.8)     | 6.1 (4.6)     | 6.7 (5.0)     | -0.4 (-0.9 to 0.0)                   | 6.7 (5.0)     | 0.3 (-0.4 to 1.0)                    | 6.8 (5.1)     | 0.5 (-0.2 to 1.2)                    |
| Global health, mean (SD)                       | 54.4 (24.8)   | 53.6 (24.8)   | 56.4 (23.5)   | -1.3 (-3.6 to 1.3)                   | 56.4 (23.5)   | 1.5 (-1.5 to 4.5)                    | 57.3 (21.8)   | 2.7 (-0.4 to 5.7)                    |
| ESR, mean (SD)                                 | 24.2 (20.4)   | 30.6 (22.5)   | 30.4 (22.2)   | 2.2 (0.1 to 4.4)                     | 30.4 (22.2)   | 4.0 (1.1 to 6.9)                     | 33.4 (26.3)   | 7.7 (4.1 to 11.2)                    |
| CRP, mean (SD)                                 | 15.9 (21.9)   | 18.8 (22.4)   | 20.6 (25.9)   | 0.7 (-1.4 to 2.8)                    | 20.6 (25.9)   | 3.4 (-0.1 to 6.7)                    | 23.3 (28.0)   | 7.4 (3.6 to 10.9)                    |
| VAS pain, mean (SD)                            | 54.8 (24.9)   | 53.5 (24.5)   | 56.9 (24.1)   | -1.7 (-4.1 to 0.8)                   | 56.9 (24.1)   | 1.8 (-1.5 to 5.3)                    | 57.3 (23.8)   | 2.2 (-1.1 to 5.5)                    |
| Conc. use of MTX, n (%)                        | 3812 (71.8)   | 352 (53.7)    | 161 (58.8)    | -17.8 (-21.8 to -13.6)               | 161 (58.8)    | -12.6 (-18.7 to -6.7)                | 135 (55.1)    | -15.5 (-21.7 to -9.4)                |
| Conc. use of non-MTX csDMARD, n (%)            | 1035 (19.5)   | 166 (25.3)    | 54 (19.7)     | 6.9 (3.2 to 10.6)                    | 54 (19.7)     | 0.4 (-5.1 to 5.2)                    | 28 (11.4)     | -6.9 (-11.2 to -2.9)                 |
| Conc. use of oral steroids, n (%)              | 2692 (50.7)   | 375 (57.3)    | 139 (50.7)    | 4.0 (-0.0 to 8.0)                    | 139 (50.7)    | -1.7 (-7.8 to 4.2)                   | 116 (47.3)    | -4.3 (-10.8 to 2.3)                  |
| Conc. use of NSAIDs, n (%)                     | 1573 (29.6)   | 128 (19.5)    | 60 (21.9)     | -8.4 (-11.7 to -5.0)                 | 60 (21.9)     | -6.1 (-11.2 to -1.2)                 | 56 (22.9)     | -4.3 (-9.9 to 0.9)                   |

Mean differences are compared with TNFI in multivariable linear regression adjusted for age, sex and geographical region with bootstrapped CIs.

bDMARD, biological disease-modifying anti-rheumatic drug; csDMARD, conventional synthetic disease modifying anti-rheumatic drug; RA, rheumatoid arthritis; TNFI, tumour necrosis factor inhibitor; TSEK, Thousand SEK; MTX, methotrexate.



**Figure 1** History of disease at treatment start of bDMARD therapy among all patients with rheumatoid arthritis in the SRQ, 2011–2015. Differences in proportion (with 95% CIs) are with reference to TNFi, and adjusted for age, sex and geographical region. bDMARD, biological disease-modifying anti-rheumatic drug; SRQ, Swedish Rheumatology Quality register; SSRI, selective serotonin reuptake inhibitor; TNFi, tumour necrosis factor inhibitor.

a history of the assessed diseases, and having consumed more healthcare resources before treatment start (table 1 and figure 1). Adjusting for age, sex and geographical region decreased these differences, but most of the associations remained. The same was not seen for tocilizumab, where baseline medical history was more similar to the TNFi group. Of particular note, rituximab had a higher proportion with recent (within 5 years) or non-recent (more than 5 years before) malignancy at treatment start (8.1% with recent malignancy, compared with 1.4% on TNFi) (figure 1). Due to low numbers, it was not possible to assess a difference in history of tuberculosis (n=9 patients had tuberculosis before starting first bDMARD), hepatitis B (n<5), hepatitis C (n=6) or multiple sclerosis/demyelinating events (n=14). For brevity, these conditions and other inflammatory conditions are presented in online supplementary table s3-s5.

In sensitivity analyses, further adjustment for demographics, disease activity and medical history did not materially alter the observed pattern of differences in baseline characteristics (online supplementary table s3). Comparisons of individual TNFis

revealed few noteworthy differences (online supplementary table s4), but those starting infliximab were slightly older and had slightly higher disease activity compared with the others, while those starting etanercept more often were female, and had accrued higher healthcare costs.

#### Patient characteristics at switch from first TNFi

We identified 2829 patients with RA who initiated a second ever bDMARD within 1 year of discontinuing a first TNFi. (For reference, during the same period, 1144 patients discontinued a first TNFi without starting a bDMARD within 1 year). It was common to start a second TNFi (n=1846, 65%), regardless of recorded reason for discontinuing the first TNFi. The switch cohort was more homogenous than the first bDMARD cohort, with overall smaller differences across therapies (table 2 and figure 1). Patients starting rituximab and abatacept were older than those starting a TNFi, and had a higher proportion with recent serious infections. Those starting rituximab had a

**Table 2** Demographics and clinical characteristics at start of second bDMARD therapy, among Swedish patients with RA switching from TNFi 2011–2015

|  | TNFi          | Rituximab     | Mean difference versus TNFi (95% CI) | Abatacept     | Mean difference versus TNFi (95% CI) | Tozilumab     | Mean difference versus TNFi (95% CI) |
|--|---------------|---------------|--------------------------------------|---------------|--------------------------------------|---------------|--------------------------------------|
| Patients, n                                    | 1846          | 376           |                                      | 292           |                                      | 315           |                                      |
| <b>Demographics</b>                            |               |               |                                      |               |                                      |               |                                      |
| Age, mean (SD)                                 | 55.1 (14.4)   | 60.3 (12.3)   | 5.1 (3.6 to 6.5)                     | 58.7 (13.3)   | 3.7 (2.0 to 5.4)                     | 56.6 (13.8)   | 1.6 (–0.0 to 3.2)                    |
| Female, n(%)                                   | 1408 (76.3)   | 270 (71.8)    | –2.6 (–7.6 to 2.5)                   | 233 (79.8)    | 5.0 (–0.0 to 9.9)                    | 241 (76.5)    | 1.0 (–3.8 to 5.8)                    |
| <b>Highest education, n (%)</b>                |               |               |                                      |               |                                      |               |                                      |
| 9 years or less                                | 318 (21.7)    | 82 (26.5)     | 1.2 (–4.0 to 6.3)                    | 57 (26.1)     | 2.3 (–4.1 to 8.1)                    | 70 (27.5)     | 4.1 (–1.8 to 10.0)                   |
| 10 to 12 years                                 | 705 (48.1)    | 156 (50.3)    | 2.9 (–3.5 to 8.9)                    | 111 (50.9)    | 4.1 (–3.1 to 11.3)                   | 110 (43.1)    | –4.1 (–10.6 to 2.3)                  |
| >12 years                                      | 443 (30.2)    | 72 (23.2)     | –4.1 (–9.2 to 1.2)                   | 50 (22.9)     | –6.4 (–12.0 to –0.4)                 | 75 (29.4)     | 0.1 (–6.0 to 6.2)                    |
| <b>Country of birth, n (%)</b>                 |               |               |                                      |               |                                      |               |                                      |
| Sweden   | 145 (7.9)     | 28 (7.4)      | –1.8 (–5.6 to 2.0)                   | 31 (10.6)     | –4.0 (–8.7 to 0.6)                   | 35 (11.1)     | –2.3 (–6.5 to 2.1)                   |
| Scandinavia                                    | 90 (4.9)      | 24 (6.4)      | 0.5 (–2.2 to 3.2)                    | 15 (5.1)      | 0.2 (–2.5 to 2.8)                    | 11 (3.5)      | –1.2 (–3.6 to 1.1)                   |
| Other  | 1611 (87.3)   | 324 (86.2)    | 1.3 (–1.8 to 4.1)                    | 246 (84.2)    | 3.8 (0.1 to 7.4)                     | 269 (85.4)    | 3.5 (–0.5 to 7.1)                    |
| <b>Medical history</b>                         |               |               |                                      |               |                                      |               |                                      |
| Days hospitalised last 5 years, mean (SD)      | 6.9 (14.7)    | 11.0 (20.7)   | 3.1 (0.7 to 5.3)                     | 8.8 (18.3)    | 1.5 (–0.8 to 3.5)                    | 9.9 (19.4)    | 2.3 (0.0 to 4.4)                     |
| Days of work loss last 5 years, mean (SD)      | 478.6 (623.9) | 615.1 (723.0) | 183.0 (103.1 to 264.7)               | 564.2 (661.3) | 20.4 (–76.8 to 118.1)                | 486.0 (667.6) | –46.3 (–144.5 to 43.7)               |
| Healthcare costs last 5 years, TSEK, mean (SD) | 319.2 (256.8) | 370.6 (283.4) | 30.8 (–1.6 to 61.3)                  | 354.8 (269.0) | 11.3 (–20.6 to 42.9)                 | 339.1 (274.5) | 12.0 (–20.7 to 43.8)                 |
| <b>RA-related characteristics</b>              |               |               |                                      |               |                                      |               |                                      |
| Rheumatoid factor positive, n (%)              | 1333 (75.1)   | 334 (90.8)    | 14.2 (10.7 to 17.9)                  | 221 (77.0)    | 0.6 (–4.8 to 6.1)                    | 237 (78.0)    | 1.9 (–3.3 to 7.2)                    |
| Disease duration, years, mean (SD)             | 12.0 (10.6)   | 13.9 (11.4)   | 0.5 (–0.8 to 1.7)                    | 12.8 (11.4)   | –0.3 (–1.6 to 1.0)                   | 10.8 (10.8)   | –1.7 (–3.0 to –0.5)                  |
| HAQ, mean (SD)                                 | 1.05 (0.63)   | 1.20 (0.63)   | 0.11 (0.03 to 0.18)                  | 1.21 (0.67)   | 0.12 (0.03 to 0.22)                  | 1.17 (0.64)   | 0.09 (0.01 to 0.18)                  |
| DAS28, mean (SD)                               | 4.37 (1.37)   | 4.88 (1.22)   | 0.44 (0.28 to 0.61)                  | 4.82 (1.30)   | 0.38 (0.19 to 0.57)                  | 5.03 (1.38)   | 0.64 (0.45 to 0.82)                  |
| Tender/joints, 0–28, mean (SD)                 | 5.6 (5.2)     | 6.3 (5.5)     | 0.8 (0.1 to 1.4)                     | 6.5 (5.4)     | 0.7 (–0.0 to 1.5)                    | 7.5 (6.7)     | 1.8 (1.0 to 2.6)                     |
| Swollen joints, 0–28, mean (SD)                | 4.7 (4.5)     | 5.6 (4.9)     | 0.9 (0.3 to 1.5)                     | 5.1 (4.1)     | 0.3 (–0.3 to 0.8)                    | 6.2 (5.3)     | 1.5 (0.8 to 2.1)                     |
| Global health, mean (SD)                       | 53.6 (24.9)   | 56.7 (24.4)   | 2.9 (–0.2 to 5.9)                    | 58.9 (24.2)   | 4.9 (1.5 to 8.4)                     | 59.2 (25.1)   | 5.0 (1.6 to 8.3)                     |
| ESR, mean (SD)                                 | 24.5 (21.4)   | 33.8 (25.1)   | 7.2 (4.1 to 10.2)                    | 29.9 (22.3)   | 4.0 (0.9 to 6.9)                     | 33.9 (25.5)   | 9.2 (6.0 to 12.4)                    |
| CRP, mean (SD)                                 | 14.7 (22.1)   | 21.4 (25.8)   | 5.7 (2.6 to 8.7)                     | 18.5 (25.0)   | 3.5 (–0.0 to 6.8)                    | 24.5 (29.1)   | 9.4 (6.1 to 12.9)                    |
| VAS pain, mean (SD)                            | 53.8 (25.4)   | 56.1 (25.3)   | 2.1 (–1.2 to 5.2)                    | 59.3 (24.1)   | 4.9 (1.5 to 8.3)                     | 59.8 (25.0)   | 5.3 (1.8 to 8.5)                     |
| Conc. use of MTX, n (%)                        | 1343 (72.8)   | 292 (77.7)    | 4.9 (0.4 to 9.6)                     | 220 (75.3)    | 4.1 (–1.2 to 9.8)                    | 232 (73.7)    | 1.2 (–4.1 to 6.5)                    |
| Conc. use of non-MTX csDMARD, n (%)            | 409 (22.2)    | 90 (23.9)     | 2.0 (–2.9 to 6.3)                    | 58 (19.9)     | –1.8 (–7.0 to 3.3)                   | 51 (16.2)     | –5.5 (–10.1 to –0.8)                 |
| Conc. use of oral steroids, n (%)              | 1106 (59.9)   | 252 (67.0)    | 7.4 (2.2 to 12.4)                    | 207 (70.9)    | 10.5 (4.7 to 16.5)                   | 207 (65.7)    | 7.0 (1.0 to 12.9)                    |
| Conc. use of NSAIDs, n (%)                     | 852 (46.2)    | 166 (44.1)    | –0.1 (–5.7 to 5.4)                   | 115 (39.4)    | –4.6 (–10.9 to 1.4)                  | 145 (46.0)    | 0.6 (–5.2 to 6.3)                    |
| <b>Reason for switch, n (%)</b>                |               |               |                                      |               |                                      |               |                                      |
| Adverse event                                  | 374 (20.6)    | 72 (19.5)     | –1.5 (–6.0 to 3.0)                   | 67 (23.8)     | 3.1 (–2.3 to 8.3)                    | 49 (15.8)     | –5.8 (–10.6 to –1.4)                 |
| Lack of effect                                 | 1170 (64.4)   | 235 (63.7)    | –0.1 (–5.7 to 5.4)                   | 165 (58.7)    | –6.4 (–12.9 to 0.3)                  | 235 (75.6)    | 10.9 (5.4 to 16.3)                   |
| Other  | 272 (15.0)    | 62 (16.8)     | 1.6 (–2.9 to 5.7)                    | 49 (17.4)     | 3.4 (–1.6 to 8.1)                    | 27 (8.7)      | –5.1 (–8.6 to –1.7)                  |

Mean differences are compared with TNFi in multivariable linear regression adjusted for age, sex and geographical region with bootstrapped CIs. bDMARD, biological disease-modifying anti-rheumatic drug; RA, rheumatoid arthritis; TNFi, tumour necrosis factor inhibitor.

higher proportion with recent malignancy and with seropositive RA. Unlike the channelling at first bDMARD, all non-TNFi groups had higher disease activity at switch than the TNFi group. Tocilizumab was more common among those who discontinued the first TNFi due to lack of effect; abatacept was more common among those discontinuing due to adverse events.

In sensitivity analyses of specific TNFis, individual drugs were overall very similar, although several differences reached nominal significance (online supplementary table s5). Infliximab initiators had lower average education (38% had 9 years or less, vs 20% in other groups), more work loss, and less psoriasis/psoriatic arthritis (PsA). There was also a significant difference in the proportion female, ranging from 69% for infliximab to 81% for golimumab. Those starting etanercept had accrued lower healthcare costs.

### Expected differences in safety and effectiveness due to confounding by indication

Modelling using observed outcomes of patients starting bDMARDs in 2006–2010 indicated that several of the factors associated with choice of therapy were also significant predictors of safety and treatment outcomes (associations in online supplementary table s6). Age and sex were strong predictors of all outcomes except remaining on drug less than 1 year. Components of baseline disease activity were predictive of all outcomes, although with varying magnitude (HAQ was associated with risk of MACE; DAS28 and its components with achieving good EULAR response). Concomitant therapy at baseline was also a predictor of most outcomes, for example, glucocorticoids at baseline were predictive of adverse events and decreased proportion with good EULAR response. Medical history also predicted treatment outcomes, for example, a history of infection or cardiovascular disease predicted future infections and cardiovascular disease, while history of malignancy significantly predicted drug retention and (weakly) new onset of malignancy.

Taken together, the observed baseline differences led to substantial differences in the predicted risk of all-cause mortality, MACE and serious infections; smaller differences in risk for malignancy and for achieving EULAR good response; and virtually no expected differences in 1-year drug survival (table 3). In summary, a crude comparison of the non-TNFi drugs with the TNFi group would be particularly biased against rituximab and abatacept regarding both safety and EULAR response. The

predicted bias was much less at switch from first TNFi, reflecting the greater similarity in patient groups.

Age and sex standardisation greatly reduced predicted bias, in particular for safety outcomes (table 3). The expected risks were still inflated for all safety outcomes except malignancies, however, and this standardisation did not reduce the biased difference in EULAR response, reflecting that age was not a strong predictor of that outcome, and that the differences in sex were minor between drugs.

### DISCUSSION

In this large, nationwide study of contemporary Swedish patients with RA, we found evidence of substantial differences in baseline characteristics among patients assigned to different bDMARDs. Many predictors of treatment assignment were also predictors of adverse treatment outcomes, and in quantifying the magnitude of this, we showed that a direct comparison across therapies would not give accurate estimates of the treatments' relative effect, but would be biased in favour of TNFi.

Those not starting 'standard' TNFi therapy were older, had lower socioeconomic status and had a higher burden of other diseases. There was similar, although slighter, channelling at switch from a first ever TNFi, where a higher RA disease activity was also predictive of receiving a non-TNFi. While there were limited differences between those starting individual TNFi, channelling to and between non-TNFi bDMARD was substantial. Rituximab initiators were oldest, dominantly RF-positive and had the highest burden of other diseases (in particular malignancy), while those starting tocilizumab differed less from those starting TNFi in terms of medical history, but had significantly higher disease activity.

These differences are partly expected based on the tentative recommendations in favour of specific drug choice for some risk groups, where for example, American College of Rheumatology (ACR) guidelines have listed 'very low' evidence to support preference of rituximab over TNFi among patients with a history of malignancy, and of abatacept over TNFi among patients with serious infections, and 'moderate to very low' evidence to prefer non-TNFi among patients with congestive heart disease.<sup>1 2</sup> It seems clear that these tentative recommendations have been followed for some, but not most, patients.

By modelling the expected risk for several treatment outcomes conferred by observed patient characteristics, we

**Table 3** Potential for confounding by indication; predicted percentage with adverse events within 5 years, and treatment outcome after 1 year, based on observed baseline characteristics

| Cohort           | All-cause mortality |     | Malignancy |     | MACE  |     | Serious infection |      | Drug survival <1 year |      | Good EULAR response at 1 year |      |
|------------------|---------------------|-----|------------|-----|-------|-----|-------------------|------|-----------------------|------|-------------------------------|------|
|                  | Crude               | STD | Crude      | STD | Crude | STD | Crude             | STD  | Crude                 | STD  | Crude                         | STD  |
| First bDMARD     |                     |     |            |     |       |     |                   |      |                       |      |                               |      |
| TNFi             | 4.8                 | –   | 5.6        | –   | 5.4   | –   | 14.4              | –    | 30.3                  | –    | 31.0                          | –    |
| Rituximab        | 13.3                | 7.0 | 8.8        | 6.1 | 10.0  | 6.1 | 24.2              | 17.7 | 29.4                  | 28.9 | 25.3                          | 23.2 |
| Abatacept        | 11.9                | 8.1 | 7.0        | 5.8 | 9.1   | 6.9 | 21.3              | 18.1 | 31.2                  | 31.1 | 27.9                          | 29.2 |
| Tocilizumab      | 8.8                 | 7.1 | 6.1        | 5.4 | 7.1   | 6.1 | 17.9              | 15.9 | 30.7                  | 30.9 | 30.3                          | 31.6 |
| Switch from TNFi |                     |     |            |     |       |     |                   |      |                       |      |                               |      |
| TNFi             | 5.3                 | 5.3 | 4.8        | 4.7 | 6.1   | 6.1 | 16.9              | 16.7 | 36.2                  | 36.1 | 17.6                          | 17.6 |
| Rituximab        | 8.1                 | 6.3 | 5.7        | 4.9 | 7.6   | 6.3 | 21.2              | 19.0 | 35.1                  | 34.8 | 18.2                          | 19.1 |
| Abatacept        | 7.3                 | 6.8 | 5.3        | 4.8 | 7.0   | 6.9 | 19.5              | 18.2 | 37.9                  | 37.8 | 18.0                          | 18.3 |
| Tocilizumab      | 6.8                 | 6.4 | 5.1        | 5.0 | 6.8   | 6.4 | 18.1              | 17.6 | 37.1                  | 37.7 | 18.3                          | 18.2 |

Predicted observed percentage (crude) and age-sex standardised to TNFi as first bDMARD (STD).

bDMARD, biological disease-modifying anti-rheumatic drug; EULAR, European League Against Rheumatism; MACE, major acute cardiovascular event; TNFi, tumour necrosis factor inhibitor.

showed that even if there were no true differences in drug effect, confounding by indication will make the non-TNFi drugs appear less safe and effective than the TNFi as first bDMARD. For many of the perceived differences, a simple adjustment for age and sex reduced this confounding dramatically. Residual confounding is, however, expected to give higher rates of adverse events and less treatment response, such that comparisons should be adjusted for medical history and disease activity when possible. As expected, the predicted bias was less when studying those switching from an initial TNFi, reflecting the reduced patient heterogeneity in this specific clinical situation.

We believe that the predictive modelling approach is helpful in combining the multitude of observed baseline differences in a metric comparable across cohorts, but several limitations should be noted. The models were based on historical data, and will be incorrect if the strength of association with each risk factor has changed over calendar time. The models were also limited by the covariates we had available, and we lacked data on for instance body mass index and smoking. Unless some unknown predictors work in the opposite direction, it is likely that we underestimate the predicted bias. The prediction models were intended as a convenient way of illustrating the risk of confounding by indication, not as the best possible prediction model for these outcomes. For this reason, we used a simple model building strategy, and did not perform cross-validation to assess the models' general predictive value or construct confidence bounds on the predictions. In other limitations, it should be noted that we made a large number of statistical comparisons and present adjusted differences between groups with standard CIs; several significant differences are likely to reflect false positives. Finally, these data are by their nature relevant to the Swedish clinical setting, where the physician is free to prescribe any drug of their choosing and the state (single payer) has made recommendations (but not restrictions) based on therapy cost. The relative costs of therapy and payer restrictions may vary by country. Therefore, although the pattern of use (preferentially TNFi as first biologic) is commonplace and the Swedish national guidelines are similar to the EULAR and ACR guidelines, the generalisability to other countries may vary.

This study has several strengths. Through the Swedish nationwide registers we were able to describe patient medical history and other characteristics using prospectively collected data, avoiding the risk for recall bias, and with a completeness that would otherwise have been difficult. We could also include the entire Swedish population, avoiding the risk of selection bias. Our main limitation is the lack of data on the physician's and patient's reasoning about the choice of treatment, which may among other factors have been influenced by the route and frequency of administration.

In conclusion, we found significant channelling of older and less healthy patients with RA to non-TNFi bDMARDs, both as first bDMARD and at switch from a first TNFi. Future studies should examine whether this channelling is medically justified or, paradoxically, act to reduce the overall effectiveness and safety of bDMARD therapy. We also demonstrated the extent to which this channelling will compromise the safety and effectiveness profile of individual bDMARDs. Unless differences in age, medical history, and RA disease activity are taken into account in studies of the relative safety and effectiveness of bDMARDs, most results will be severely confounded.

**Acknowledgements** The ARTIS Study Group conducts scientific analyses using data from the Swedish Biologics Register ARTIS, run by the Swedish Society for Rheumatology. The following were members of the ARTIS Study Group during the completion of this study: JA, Sofia Ernestam, Lars Klareskog, Ralph Nisell and Ronald van Vollenhoven (Karolinska Institutet, Stockholm, Sweden), Nils Feltelius (Medical Products Agency, Uppsala, Sweden), EB (Uppsala University, Uppsala, Sweden), Alf Kastbom (Linköping University, Linköping, Sweden), Lennart Jacobsson (Sahlgrenska Academy, Göteborg, Sweden), Carl Turesson and Elisabet Lindqvist (Lund University, Malmö/Lund, Sweden), HF-D'E and Solbritt Rantapää-Dahlqvist (Umeå University, Umeå, Sweden).

**Contributors** TF had full access to all data and takes responsibility for the integrity of the data and the accuracy of the data analysis. TF performed all analyses and drafted the first version of the manuscript. All authors were involved in drafting the study protocol before analyses, and reviewed the final manuscript for important intellectual contents.

**Funding** This study was supported by the Swedish Foundation for Strategic Research, Swedish Research Council, Swedish Cancer Society, Swedish Heart Lung Foundation, ALF funding collaboration between Karolinska Institute and Stockholm County, and the IMI funded EU project Be The Cure.

**Competing interests** ARTIS has entered into agreements with Abbvie, BMS, MSD, Lilly, Pfizer, Roche, Samsung Bioepis and UCB. Companies with products mentioned in this manuscript were given a courtesy review of the manuscript before publication, but were not involved in planning the study, performing the analysis or interpreting the results.

**Ethics approval** The Regional Ethical Review Board in Stockholm granted ethical approval (DNR: 2016/1986-32).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** Access to the national register data used for this study is granted on a restrictive basis, and may not be shared without additional specific permissions from the Swedish register-holding authorities.

**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 and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

## REFERENCES

- 1 Singh JA, Furst DE, Bharat A, *et al.* 2012 Update of the 2008 American college of rheumatology recommendations for the use of disease-modifying antirheumatic drugs and biologic agents in the treatment of rheumatoid arthritis. *Arthritis Care Res* 2012;64:625–39.
- 2 Singh JA, Saag KG, Bridges SL, *et al.* 2015 American college of rheumatology guideline for the treatment of rheumatoid arthritis. *Arthritis Rheumatol* 2016;2016:68:1–26.
- 3 Smolen JS, Landewé R, Bijlsma J, *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* 2017;76:960–77.
- 4 Smolen JS, Landewé R, Breedveld FC, *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.
- 5 Neovius M, Arkema EV, Olsson H, *et al.* Drug survival on TNF inhibitors in patients with rheumatoid arthritis: comparison of adalimumab, etanercept and infliximab. *Ann Rheum Dis* 2015;74:354–60.
- 6 Wolfe F, Flowers N, Burke TA, *et al.* Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channelling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. *J Rheumatol* 2002;29:1015–22.
- 7 Petri H, Urquhart J. Channelling bias in the interpretation of drug effects. *Stat Med* 1991;10:577–81.
- 8 Putrik P, Ramiro S, Chorus AM, *et al.* Socioeconomic inequities in perceived health among patients with musculoskeletal disorders compared with other chronic disorders: results from a cross-sectional Dutch study. *RMD Open* 2015;1:e000045.
- 9 Kyriacou DN, Lewis RJ. Confounding by indication in clinical research. *JAMA* 2016;316:1818–9.
- 10 Askling J, Fored CM, Geborek P, *et al.* Swedish registers to examine drug safety and clinical issues in RA. *Ann Rheum Dis* 2006;65:707–12.
- 11 Frisell T, Holmqvist M, Källberg H, *et al.* Familial risks and heritability of rheumatoid arthritis: role of rheumatoid factor/anti-citrullinated protein antibody status, number

- and type of affected relatives, sex, and age. *Arthritis Rheum* 2013;65:2773–82.
- 12 Eriksson JK, Askling J, Arkema EV. The Swedish rheumatology quality register: optimisation of rheumatic disease assessments using register-enriched data. *Clin Exp Rheumatol* 2014;32:S147–149.
  - 13 Wadström H, Eriksson JK, Neovius M, *et al.* How good is the coverage and how accurate are exposure data in the Swedish biologics register (ARTIS)? *Scand J Rheumatol* 2015;44:22–8.
  - 14 Socialstyrelsen. *Inpatient diseases in Sweden 1987-2008 [in Swedish]*. Stockholm: Socialstyrelsen, 2008.
  - 15 Ludvigsson JF, Andersson E, Ekblom A, *et al.* External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450.
  - 16 Wettermark B, Hammar N, Fored CM, *et al.* The new Swedish prescribed drug register—opportunities for pharmacoepidemiological research and experience from the first six months. *Pharmacoepidemiol Drug Saf* 2007;16:726–35.
  - 17 Barlow L, Westergren K, Holmberg L, *et al.* The completeness of the Swedish cancer register: a sample survey for year 1998. *Acta Oncol* 2009;48:27–33.
  - 18 Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. *Int J Epidemiol* 2000;29:891–8.
  - 19 Canty A, Ripley BD. *boot: bootstrap R (S-Plus) functions*. R package version 1.3-20 edn: McMaster University Hamilton, Ontario, Canada, 2017.
  - 20 Davison AC, Hinkley DV. *Bootstrap methods and their applications*. Cambridge: Cambridge University Press, 1997.