Rheumatoid Arthritis Disease Severity Indices in Administrative Databases: A Systematic Review

ÉVELYNE VINET, BINDEE KURIYA, JESSICA WIDDIFIELD, and SASHA BERNATSKY

ABSTRACT. ***Objective.*** We aimed to systematically review rheumatoid arthritis (RA) disease severity indices for use in administrative healthcare databases. We also provide an overview of alternative methods to control for RA disease severity in administrative database research.

Methods. We conducted a systematic review of studies that developed/validated an index for RA disease severity using variables in administrative databases, and compared the convergent validity/reliability of the index with a standard measure of RA severity.

Results. After reviewing 539 articles, 2 studies were included. The claims-based index for RA severity (CIRAS) was developed in one study. Components of the CIRAS included tests for inflammatory markers, number of chemistry panels/platelet counts ordered, rheumatoid factor test, number of rehabilitation and rheumatology visits, and Felty's syndrome. The CIRAS correlated moderately well with a previously validated RA medical records-based index of severity. The second study assessed whether current and lifetime treatment with disease-modifying antirheumatic drugs and/or biologics accurately predicted RA severity, as measured by the patient-reported Patient Activity Scale (PAS). Treatment variables did not fully distinguish patients in the highest and lowest quartiles of PAS scores (67.2% correctly classified).

Conclusion. Two claims-based indices of RA severity were identified but have some limitations for routine use. A concerted effort from experts in the field is needed to define, develop, and validate a widely applicable measure of RA disease severity for administrative database research. (First Release Sept 15 2011; J Rheumatol 2011;38:2318–25; doi:10.3899/jrheum.110587)

Key Indexing Terms:
RHEUMATOID ARTHRITIS   PHARMACOEPIDEMIOLOGY   SEVERITY OF ILLNESS INDEX
INSURANCE CLAIM REPORTING   DATABASE MANAGEMENT SYSTEM

Rheumatoid arthritis (RA) is a chronic autoimmune condition, primarily characterized by pain, swelling, and deformity of the joints[1]. RA severity may vary over time, and its heterogeneous clinical manifestations require complex management. Healthcare databases are increasingly used to examine population-based outcomes in RA. Patients with RA have an increased risk of adverse events including infections, malignancies, and cardiovascular diseases[4,5,6]. Many studies have linked this higher risk of comorbidities, at least in part, to disease-modifying antirheumatic drugs (DMARD) and biologic

agents used to treat RA. For example, tumor necrosis factor-α (TNF-α) inhibitors are associated with infections like tuberculosis and invasive fungal disease[7,8,9,10]. TNF-α inhibitors and some traditional DMARD have also been related to an excess risk of lymphoma, although this remains controversial[11,12,13].

An important limitation of these pharmacoepidemiological studies is the inadequate control for confounding by disease severity. In practice, disease severity is difficult to define but, as a construct, may reflect disease activity, damage, or both[14]. In most circumstances, the severity of RA determines the intensity of treatment. In observational studies, patients with RA who are selected to receive DMARD or biologic therapies are systematically different from those who are not treated with these agents[15,16]. Any increased harm found among the treated patients may be mistakenly attributed to the treatment, when in fact it may be the result of more active disease[17]. Indeed, confounding by disease severity may occur when a drug is preferentially prescribed to patients with a worse baseline prognosis[18]. If adjustment for RA severity is incomplete, an overestimation of the relationship between drug exposure and outcome may result[19].

Administrative or healthcare claims databases (e.g., physician billing and hospitalization databases) were created notably to administer payments to healthcare professionals.

From the Division of Rheumatology, University of Toronto, Toronto, Ontario; and McGill University Health Centre, Montreal, Quebec, Canada.

É. Vinet, MD, McGill University Health Centre; B. Kuriya, MD, MS, Division of Rheumatology, University of Toronto; J. Widdifield, BSc, University of Toronto; S. Bernatsky, MD, PhD, McGill University Health Centre.

Dr. Vinet and Dr. Kuriya contributed equally to this work.

Address correspondence to Dr. É. Vinet, Montreal General Hospital, McGill University Health Centre, 1650 Cedar Avenue, Room A6 162.2, Montreal, Quebec H3G 1A4, Canada. E-mail: evelyne.vinet@mail.mcgill.ca

These databases prospectively collect patient information such as demographics, hospitalizations, and physician visits, and often can be linked with prescriptions data, cancer, and birth-malformation registries[20]. Several administrative databases are currently available for observational studies in North America, including Canadian provincial healthcare databases (e.g., Québec and Ontario), different American health maintenance organizations (e.g., Kaiser Permanente and US Veterans Affairs), and Medicaid databases (e.g., Tennessee and New Jersey)[20].

One advantage of these databases is that multiple drug exposures can be evaluated simultaneously. In addition, they offer population-based data and are thus more representative of the general population than drug trials. Also, the large sample sizes allow the assessment of relatively rare outcomes, such as malignancies[21,22]. Because these databases are created for reimbursement purposes, their major disadvantage is that they lack information on important clinical variables, such as the number of tender and swollen joints, which would traditionally be used to assess disease severity in RA[20].

Administrative databases may, however, contain correlates of disease severity that could, theoretically, reduce bias due to confounding by disease severity. Surrogates, including the number of physician visits, acute-care hospitalizations, joint replacement surgeries, or dispensing of different DMARD or biologic agents, are typically indicative of more severe RA[23,24]. Thus, partial adjustment for disease severity may be possible in administrative databases.

The objective of this investigation was to systematically review the literature for studies that aimed to develop and/or validate a health claims-based index for RA disease severity. Studies that developed a tool using variables available in administrative databases and that compared its validity and/or reliability with a clinical measure of RA severity were included for review. We also discuss alternative methods to control for RA disease severity in administrative database research.

MATERIALS AND METHODS

This study was developed according to the Consort Group's Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines, and followed a protocol that prespecified study selection, eligibility criteria, quality assessment, and data abstraction[25].

Search strategy. We conducted a systematic review of articles using the following electronic databases: PubMed (1950-July 2010), Embase (1980-July 2010), and Web of Science (1991-July 2010). No language restrictions were applied. We used Medical Subject Heading and free text terms adapted for each database (see Appendix 1 for complete list of search terms). We included search terms for (1) rheumatoid arthritis; (2) disease activity, disease severity, or disability; and (3) administrative database. All terms within each set were combined using the Boolean operator "OR" and then the 3 sets were combined using "AND". We hand-searched reference lists and review articles for relevant articles not identified by the electronic searches. Any pertinent secondary references, including meeting abstracts, were also reviewed.

Study selection. Two independent reviewers (BK and EV) screened the titles and abstracts of all studies for eligibility. We included cohort studies of adult patients with RA (aged ≥ 18 years) that assessed disease severity using information from an administrative data source (or that would be routinely available in administrative databases). We also assessed the convergent validity or reliability (qualitatively or quantitatively) of the health claim-based severity index against at least 1 clinical measure of disease severity (see Appendix 2 for complete list). Any disagreement between the 2 reviewers was resolved by consensus, or if necessary, by a third party.

Data extraction. Two observers (BK and EV) independently extracted data on study type, patient population and baseline characteristics, components of the health claim-based index used, type of clinical comparator of disease severity used, and any reported measures of the strength of association/correlation between the index and clinical comparator.

Quality assessment. All included studies were independently assessed for methodological features most relevant to the control of bias in observational studies by using the quality assessment tool developed by Hayden, *et al*[26]. Before data extraction and quality assessment were finalized, any remaining disagreements were resolved by consensus.

RESULTS

Studies included. The literature search identified a total of 536 studies that matched the predefined search terms. Three were additionally identified by hand-searching reference lists. Of the potentially relevant studies retrieved, the majority were excluded because the data were not from and/or could not be derived from an administrative data source (Figure 1). A total of 2 studies were included in the final analysis and were judged to be of sufficient quality.

Study characteristics. *Study 1*. Ting, *et al*[24] aimed to develop a healthcare claims index of RA severity, using a previously developed RA medical records-based index of severity (RARBIS). The RARBIS was based on ratings of potential indicators of RA severity commonly found in medical charts by a Delphi panel of 6 rheumatologists[27]. Indicators ranked as having strong or very strong association with RA severity were included in the RARBIS, and subjectively weighted by the panel based on the perceived association with disease severity. The RARBIS includes items pertaining to the following categories: radiological and laboratory results, surgeries, extraarticular manifestations, clinical and functional status, and medications. The RARBIS has been shown to correlate moderately well with RA treatment intensity and the 28-joint Disease Activity Score (DAS28) in a cohort of US Veterans Administration (VA) patients[28].

The population studied by Ting, *et al* comprised patients from the New England VA Health System who had at least 2 recorded visits with a diagnosis of RA (International Classification of Disease, 9th ed, code 714.0) and at least 2 hospital outpatient visits between July 1999 and June 2001, with sufficient evidence of RA from their medical record. One hundred twenty patients were included and their characteristics are summarized in Table 1.

The authors developed the claims-based index of RA severity (CIRAS) by using linear regression models. Modeling the RARBIS as a function of different administrative data variables, they used diverse model selection procedures to obtain the best model, as defined by the highest model R^2. The CIRAS was composed of the administrative variables present in the best model, which included the fol-
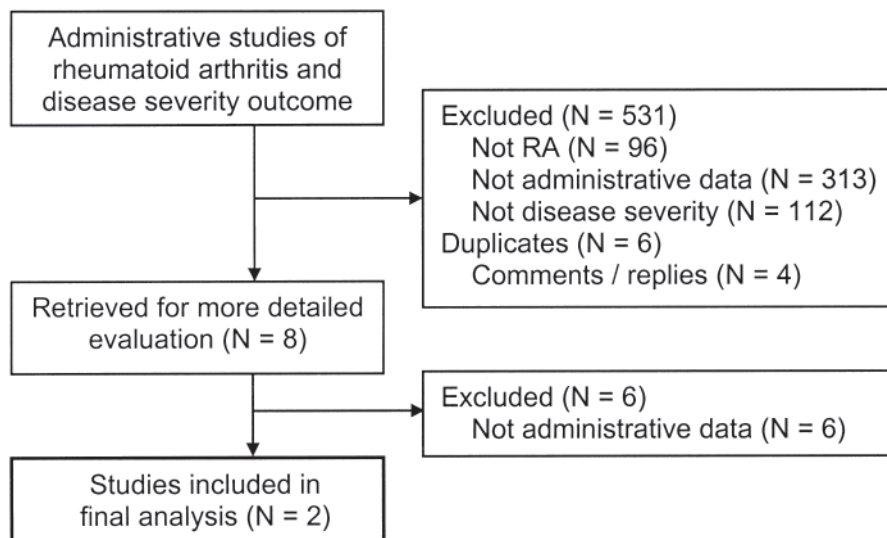
Figure 1. The selection process of studies evaluated for inclusion in the systematic review.

Table 1. Characteristics of the patients in the 2 studies included for review.

Characteristic	Ting, 2008[24], n = 120	Wolfe, 2006[30], n = 7541
Age, yrs (mean ± SD)	70.6 ± 11.1	60.8 ± 13.4
Women	11 (9)	5882 (78)
Disease Duration, yrs (mean ± SD)		14.6
ACR functional class		
I	93 (78)	
II	8 (7)	
III	6 (5)	
IV	4 (3)	
Rheumatology visits	3 (3)	—
Hospitalizations	7 (6)	—
Morning stiffness, hours		
< 1	70 (58)	
1 to 4	25 (21)	
> 4	8 (7)	
Flares		
0	65 (54)	
1	22 (18)	
1 to 4	11 (9)	
5+	3 (3)	
Swollen joints	64 (53)	—
Erosive disease	61 (51)	—
Joint space narrowing	74 (62)	—
Rheumatoid nodules	41 (34)	—
Vasculitis	7 (6)	—
Pulmonary nodule	11 (9)	—
Presence of C1-C2 subluxation	2 (2)	—
Use of glucocorticoids		2308 (30.6)
Intraarticular injection	11 (9)	
Intramuscular injection	1 (1)	
Use of DMARD	—	5595 (74.2)
Use of biologic	—	2496 (33.1)
Use of DMARD and/or biologic	—	6161 (81.7)

All numbers are n (%) unless otherwise indicated. ACR: American College of Rheumatology; DMARD: disease-modifying antirheumatic drug.

lowing variables assessed over the preceding year: age, sex, Felty's syndrome, number of rheumatology visits, rehabilitation visits (physical and occupational therapy), rheumatoid factor (RF) testing, and the number of platelet counts, inflammatory markers and chemistry panels ordered. Each of these items was weighted by its regression coefficient. For each patient in the study, both the CIRAS and RARBIS (with and without the medication subscale) were calculated using data from the administrative database and the medical charts over the past year. Then, the investigators examined the correlation between the claims-based RA severity variables (CIRAS) and the RARBIS using the Spearman correlation coefficient. The mean score for the RARBIS with the medication subscale was 4.4 (range 0–11) and without medications was 3.0 (range 0–8), and the mean CIRAS score was 4.38 (range 1.18–8.11). The Spearman correlation coefficients between the composite CIRAS score and the RARBIS with the medication subscale was 0.56, and without the medication subscale was 0.51 (both with $p < 0.0001$), indicating moderate correlation with the RARBIS[29]. The individual variables were weakly correlated with the RARBIS. Table 2 shows the adjusted correlation coefficients for each proposed variable. The authors' suggested scoring of the CIRAS is found in Table 3.

Since the CIRAS was moderately correlated with the RARBIS, the authors concluded that this index could potentially be used for adjustment of RA severity in studies using claims data, but urged future studies to examine its validity in other samples.

Study 2. Wolfe, *et al*[30] assessed whether current treatment with DMARD and/or biologics (i.e., infliximab, etanercept, adalimumab, or anakinra) and lifetime number of DMARD and/or biologics accurately predicted severity among patients with RA. To assess severity, they used the Patient Activity

Table 2. Characteristics of the included studies.

Study	Index Studied	Disease Severity Comparator	Claims-based Variables	Adjusted Measure of Association	
				Correlation to RARBIS with Medication Subscale	Correlation to RARBIS without Medication Subscale
Ting, 2008[24]	CIRAS	RARBIS	Age and sex	0.08	0.05
			Rheumatology visits	0.01	—
			Rehabilitation visits	0.01	0.04
			Felty's syndrome	0.01	0.03
			No. inflammatory markers ordered	0.14	0.08
			Rheumatoid factor test ordered	0.02	0.04
			No. platelet counts ordered	0.03	0.01
			No. chemistry panels ordered	0.01	0.02
			Composite CIRAS score	0.56 ($p < 0.0001$)	0.51 ($p < 0.0001$)
Wolfe, 2006[30]	Composite score for RA severity index	PAS	Age	AUC = 0.64	AUC = 0.70
			Sex	For detection of PAS above or below median value for cohort	For detection of 1st vs 4th PAS quartile for cohort
			Disease duration		
			Current DMARD and/or biologics		
			Lifetime number of DMARD and/or biologics		

AUC: Area under the curve; CIRAS: claims-based index for RA severity; PAS: Patient Activity Score; RARBIS: RA medical records-based index of severity; RA: rheumatoid arthritis; DMARD: disease-modifying antirheumatic drug.

Scale (PAS), which is composed of the Health Assessment Questionnaire (HAQ), a visual analog scale (VAS) for pain, and a VAS for global severity. This patient-completed scale has been shown to be strongly correlated with disease activity and severity, and is reported on a 0–10 scale by multiplying the HAQ by 3.33, and then dividing the sum of the VAS pain, VAS global, and HAQ by 3[31].

The investigators studied patients with RA enrolled in the National Data Bank for Rheumatic Diseases (NDB) longitudinal study of RA outcomes. Participants in the NDB cohort are recruited from rheumatologists' practices in the US and are prospectively followed with semiannual questionnaires, recording information such as demographic and treatment variables. For the present study, the authors included patients with RA who had completed at least 1 semiannual questionnaire between January 2002 and December 2004. They excluded patients who were participants in a safety registry. For each patient, a single visit was randomly selected in the specified time interval. Although the study population did not originate from an administrative database, this study was included in our review because it used treatment information available in administrative databases as a surrogate of disease severity, and specifically assessed the validity of this surrogate.

Characteristics of the 7541 patients included in the study are presented in Table 2. Similar to the study by Ting, *et al*[24], information on baseline variables typically associated with more severe disease was not reported. The majority of patients (81.7%) were currently using a DMARD and/or a biologic, 74.2% were using a DMARD, 33.1% a biologic, and 18.3% were taking neither DMARD nor biologics. The median PAS

score was 3.4. The greatest median PAS scores were found in patients not taking a DMARD or biologic [PAS 3.7, interquartile range (IQR) 1.7-5.7], and in those who were currently using a DMARD and biologic (PAS 3.7, IQR 2.0–5.4).

Wolfe, *et al* assessed the ability of treatment variables to predict PAS by calculating the area under the receiver-operating characteristic (ROC) curve and the percentage correctly classified for 2 logistic regression models. Both models used the same independent variables to predict PAS group: current treatment with each individual DMARD and/or biologic, the lifetime number of DMARD and/or biologics ever used, age, sex, and RA disease duration. In the first model, the authors measured the predictive ability of treatment variables to correctly classify patients with PAS scores above or below the median PAS value for the entire cohort. In the second model, they aimed to adequately predict patients with a PAS score in the fourth quartile compared to the first quartile. The area under the ROC curve was 0.64 for the first model, and the percentage correctly predicted was 60.5%. In the second model, the area under the ROC curve was 0.70 and the percentage properly classified was 67.2% (Table 1).

In their conclusion, the authors determined that the sole use of DMARD/biologics and demographic variables in administrative data does not distinguish disease severity groups with adequate sensitivity and/or specificity.

DISCUSSION
Through this systematic review, we found 2 studies that aimed to develop or validate an RA disease severity index for use in administrative database research.

Table 3. Suggested scoring method for claims-based index of rheumatoid arthritis severity (CIRAS) by Ting, *et al*[24].

Claims-based Variables	Score
Age, continuous	–0.066
Sex	–0.092
0: male	
1: female	
No. inflammatory markers ordered*	0.60
0: no	
1: yes	
Rehabilitation visits*	0.69
0: no	
1: yes	
Rheumatoid factor test*	2.1
0: no	
1: yes	
Felty's syndrome*	2.3
0: no	
1: yes	
No. platelet counts ordered*	0.42
0: platelet count = 0	
1: platelet count = 1	
2: platelet count = 2	
3: platelet count = 3	
4: platelet count ≥ 4	
No. chemistry panels ordered*	–0.14
0: chemistry panels = 0	
1: chemistry panels = 1	
2: chemistry panels = 2	
3: chemistry panels = 3	
4: chemistry panels = 4	
5: chemistry panels ≥ 5	
Rheumatology visits*	0.52
0: no. rheumatology visits = 0	
1: no. rheumatology visits = 1, 2, 3, or 4	
2: no. rheumatology visits > 4	
Intercept	6.5

* Time period for which data should be recorded for these variables is 1 year. Claims-based variables shown were selected from automated modeling procedures with the rheumatoid arthritis records-based index of severity (RARBIS) as the dependent variable and potential claims-based variables as independent variables. Scores represent parameter estimates for these explanatory variables and can be used as weights when computing the CIRAS. To obtain an overall CIRAS score, multiply the value of each claims-based variable with its corresponding score and then sum all the scores and the value for the intercept.

Ting, *et al*[24] built a complex RA severity index, the CIRAS, as a function of a previously developed medical records-based severity index, RARBIS. Although the CIRAS was derived from the selection of the best model predicting the RARBIS, it offered only a moderate correlation with the RARBIS. Their study has some limitations. The investigators used as their reference standard the RARBIS, previously shown to correlate only moderately with disease activity, and for which the reliability has not been thoroughly assessed. Moreover, they used their reference standard to develop the new index. In doing so, and by using the same group of patients in which the new index was developed, some correla-tion between both indices was expected[32]. Moreover, variables included in the CIRAS were selected because they were part of the best model identified through diverse selection procedures, not because of an *a priori* hypothesis that they were related to disease severity, although most variables (extra-articular involvement, RF testing) would be expected to correlate with clinical disease, and so they do provide face validity to the CIRAS. Some have questioned whether components of the RARBIS are appropriate measures of disease severity (e.g., joint surgery) or whether they principally represent aspects of irreversible damage, although joint surgery is obviously likely a correlate of prior disease activity and overall disease severity[14]. Although the CIRAS index may allow adjustment of some aspects of RA disease severity, further research is required.

Wolfe, *et al*[30] studied a large RA cohort to determine if treatment variables could predict disease severity[30]. Their model (including current use of a specific DMARD and/or biologic and lifetime number of these drugs) only partially discriminated patients with high disease severity from those with low severity, as defined by the PAS. Patients with high severity scores were found at both ends of the therapeutic spectrum: those not currently taking a DMARD or a biologic, and those currently taking both a DMARD and a biologic. Again, some limitations might be noted; most importantly, the reference standard was a patient-reported measure, which likely led to significant misclassification of disease severity. It is unclear why the authors decided to exclude patients enrolled in a safety registry. They stated that this was done to prevent selection bias and avoid including participants in a registry who typically have more severe disease. Since patients in the NDB cohort came from the general practice of US rheumatologists, this exclusion criterion may have prevented the investigators from evaluating the full range of disease severity ideal for validation of an index. Further, the study findings may not be generalizable to other countries where a larger proportion of the population is covered by a government medication plan and must satisfy strict clinical criteria (often based on disease severity) to be reimbursed for biologic agents. Thus, it is possible that, in other settings, the use of demographic and treatment variables may better distinguish RA severity.

While performing our systematic search, we found 1 study that assessed the relationship between the number of visits to a rheumatologist and the changes in functional disability (measured by the HAQ) in a community-based cohort of 127 patients with RA[33]. The study by Ward, *et al* was not included in our final set of articles because it did not specifically aim to develop a measure of disease severity and did not use an administrative data source. However, this study is of particular interest since it offers the opportunity to examine if rheumatology visit frequency is a good correlate of disease severity, as measured by the HAQ. Adjusting for potential confounders, the investigators observed a U-shaped relation-

ship between the average visit frequency and the rate of progression of functional disability (the minimum rate of progression corresponded to a visit frequency of 7 visits per year). Although that study was conducted before the biologics era — and at a time when many visits to rheumatologists were for intramuscular gold injections — it emphasizes the complex relationship between medical visits and RA severity. On one hand, infrequent visits may reflect inactive disease, but may also occur among noncompliant patients, who are more likely to have worse disease. Similarly, frequent visits may indicate high adherence to medical management or may represent poorly controlled symptoms that warrant close monitoring. Therefore, more work is needed to assess whether or how visit frequency can independently be used as a correlate of disease severity in health-claims databases. Interestingly, comparison of unadjusted and adjusted risk ratios from observational data do suggest that "proxy" measures of disease severity (like rheumatology visits or orthopedic interventions) may partly correct for biases like channeling[34].

An alternative to an RA severity index is the use of propensity scores. This is a method for producing treatment effects adjusted for nonrandom assignment, which is a problem in observational RA studies, since presumably treatment is preferentially used in more severe disease. A propensity score is based on the conditional probability of receiving a particular treatment given certain patient characteristics; patients with the same propensity score have theoretically the same probability of receiving treatment[16]. In contrast to other approaches to address confounding by indication, including restriction, stratification, matching, or multivariate adjustment, which are limited by the number of variables that can be controlled for simultaneously, the propensity score method can handle a large number of covariates[35]. This method creates a single variable (the propensity score) that represents the combined effect of each of the variables considered upon treatment initiation, which can be used for restriction, stratification, matching, or modeling[35].

For example, unadjusted results from an observational study of the effect of anti-TNF therapy on malignancy very likely suffer from confounding by indication, which is difficult to control for completely because the number of variables that potentially differ between treated and untreated patients is large. As mentioned, one approach to address confounding by indication is the use of a matched cohort design, in which unexposed subjects with a specific covariate level (or pattern of covariates) are matched to exposed subjects with the same characteristic. However, matching works well with only a few variables and becomes more difficult as the numbers of matched variables increase. Since the reasons for allocation of anti-TNF agent may be complex and based on numerous variables (e.g., not only disease severity but also potentially history, healthcare access, etc.), it is likely that attempts at exact matching on all these variables will fail to find appropriate matches despite large sample sizes obtained from administra-

tive databases (and failure to match on all relevant variables will result in residual confounding). Propensity scores avoid this by combining all the covariates into a single value that can serve as a matching variable, providing excellent balance between treated and untreated groups with respect to the covariates at hand[35].

The propensity score model can include multiple variables that are not related to disease severity. Indeed, a propensity score is tailored to the treatment-outcome relationship under study, compared to a "one-size-fits-all" index of severity. It makes use of all available variables present in the database under study that predict treatment exposure (and outcome), as opposed to an index of severity designed for widespread use, which would rely only on variables available in most administrative databases. Moreover, propensity scores developed within an administrative database can provide a relatively detailed picture of a patient's healthcare use, allowing integration of healthcare use variables such as counts of physician visits, hospital stays, and laboratory tests performed[35]. As noted, these measures of healthcare use may represent proxies for factors influencing medical decisions, including disease severity.

Although propensity score methods have the advantage of producing "quasi-randomization" in observational studies through the use of rich exposure models, the generalizability of their results is limited and they do not provide adjustment for unobserved confounders[16]. The propensity score can control only for known confounders that are measured and available in the dataset in question. Further, certain types of variables entered into the propensity score model can actually introduce more bias, so careful consideration is required during the variable selection process[16,34]. Because of these limitations, the use of propensity scores is not a "magic bullet" to completely eliminate confounding by disease severity. Hence, some authors prefer to continue to adjust for factors that are known or believed to be confounders, as opposed to using propensity score adjustment.

Instrumental variable analyses are another tool that could be used to deal with confounding by indication[36]. This method has the potential to adjust for both measured and unmeasured confounders of the exposure-outcome association under study. The concept is the following: the effect of an exposure on an outcome can be identified by making use of a relationship between the exposure and another variable, the instrumental variable (IV). However, to produce unbiased estimates, strict assumptions have to be fulfilled: (1) the IV must be strongly associated with the exposure; (2) the IV must not have a direct effect on the outcome; and (3) the IV must be independent of the exposure-outcome confounders[37]. The difficulty with RA disease severity is to find a perfect surrogate variable, satisfying the required assumptions for IV. Indeed, it has been argued that although IV can be useful in case of moderate confounding, they are less useful in the case of strong confounding, which is presumably the case for disease severity, because

strong instruments are not always available, and the required assumptions are often violated[38]. Traditionally, well-recognized IV have been genetic markers, which are obviously not available in administrative databases. Thus, to date this approach has not been used to address confounding by RA disease severity in administrative database research, and one can arguably ask if good candidates for IV actually exist in most administrative databases.

Our systematic review highlights the absence of a comprehensive claims-based index of RA severity, and the paucity of studies aiming to develop one. In administrative database research, it is crucial to adequately measure RA activity/severity to reduce bias from confounding by disease severity. To date, only a few investigators have addressed this challenge, and we applaud their efforts. Further concerted effort from experts in the field is needed to define, develop, and validate a widely applicable measure of RA disease severity for use in administrative databases research. Such a measure could be used by different investigators to perform sensitivity analyses of the influence of confounding by disease severity on their effect estimates, which would enhance the validity and usefulness of the study findings. In the future, results obtained through this approach could eventually be compared with those alternatively obtained with a propensity score model using a minimal set of included covariates. As administrative database research has many advantages, particularly in relatively rare conditions such as RA, it is likely that this area of rheumatic disease research will fuel a lot of interest in the next few years. Until existing methodological gaps are filled, authors can only adjust for disease severity using 1 or more of the available methods, and must always include in their discussion the direction and magnitude of potential effects of residual confounding.

APPENDIX 1

Search terms used, according to electronic database.
PubMed
1. rheumatoid arthritis[tw] OR arthritis, rheumatoid[MeSH Terms] OR RA OR inflammatory arthritis[tw] OR early arthritis[tw]
2. administrative*[tw] OR database*[tw] OR database management systems[MeSH Terms] OR pharmacoepidemiology[MeSH Terms] OR insurance claim review[MeSH Terms] OR insurance claim reporting[MeSH Terms] OR health maintenance organizations[MeSH Terms] OR claim*[tw] OR billing code*[tw] OR claims-based ind*[tw] OR CIRAS[tw]
3. severity of illness index[MeSH Terms] OR disease activity[tw] OR disease severity[tw] OR disease disability[tw] OR DAS[tw] OR DAS28[tw] OR CDAI[tw] OR SDAI[tw] OR RARBIS[tw] OR HAQ[tw] OR global assessment[tw] OR VAS[tw] OR antirheumatic agents[MeSH] OR disease-modifying antirheumatic drug[tw] OR DMARD[tw]
EMBASE
1. rheumatoid arthritis/ or rheumatoid arthritis.mp. or RA.mp. or inflammatory arthritis.mp. or early arthritis.mp.
2. administrative*.mp. or data base/ or database*.mp. or database management systems.mp. or pharmacoepidemiology/ or health insurance/ or insurance claim review.mp. or insurance claim reporting.mp. or claim*.mp. or billing code*.mp. or claims-based ind*.mp. or CIRAS.mp.
3. severity of illness index.mp. or disease activity/ or disease severity/ or disability/ or disease disability.mp. or DAS.mp. or DAS28.mp. or CDAI.mp. or

SDAI.mp. or RARBIS.mp. or HAQ.mp. or global assessment.mp. or VAS.mp. or antirheumatic agent/ or disease-modifying antirheumatic drug.mp. or DMARD.mp.
Web of Science
1. rheumatoid arthritis OR RA OR inflammatory arthritis OR early arthritis
2. administrative* OR database* OR database management system OR pharmacoepidemiology OR insurance claim review OR insurance claim reporting OR medicare OR health maintenance organizations OR claim* OR billing code* OR claims-based index OR CIRAS
3. severity of illness index OR disease activity OR disease severity OR disability OR disease disability OR DAS OR DAS28 OR CDAI OR SDAI OR RARBIS OR HAQ OR global assessment OR VAS OR antirheumatic agent OR disease-modifying antirheumatic drug OR DMARD

APPENDIX 2

Clinical measures of disease severity considered for analysis.
1. Disease Activity Score
2. Disease Activity Score in 28 Joints
3. Simplified Disease Activity Index
4. Clinical Disease Activity Index
5. Health Assessment Questionnaire
6. Physician Global Assessment of Disease
7. Patient Global Assessment of Disease
8. Visual Analogue Scale Pain or Severity
9. Rheumatoid Arthritis Medical Records-Based Index of Severity
10. Disease-modifying Antirheumatic Drug Use
11. Patient Activity Scale

REFERENCES

1. Lee DM, Weinblatt ME. Rheumatoid arthritis. Lancet 2001;358:903.
2. Sihvonen S, Korpela M, Laippala P, Mustonen J, Pasternack A. Death rates and causes of death in patients with rheumatoid arthritis: a population-based study. Scand J Rheumatol 2004;33:221-7.
3. Myasoedova E, Davis JM 3rd, Crowson CS, Gabriel SE. Epidemiology of rheumatoid arthritis: rheumatoid arthritis and mortality. Curr Rheumatol Rep 2010;12:379-85.
4. Doran MF, Crowson CS, Pond GR, O'Fallon WM, Gabriel SE. Frequency of infection in patients with rheumatoid arthritis compared with controls: A population-based study. Arthritis Rheum 2002;46:2287-93.
5. Gridley G, McLaughlin JK, Ekbom A, Klareskog L, Adami HO, Hacker DG, et al. Incidence of cancer among patients with rheumatoid arthritis. J Natl Cancer Inst 1993;85:307-11.
6. Aviña-Zubieta JA, Choi HK, Sadatsafavi M, Etminan M, Esdaile JM, Lacaille D. Risk of cardiovascular mortality in patients with rheumatoid arthritis: a meta-analysis of observational studies. Arthritis Rheum 2008;59:1690-7.
7. Tubach F, Salmon D, Ravaud P, Allanore Y, Goupille P, Breban M, et al. Risk of tuberculosis is higher with anti-tumor necrosis factor monoclonal antibody therapy than with soluble tumor necrosis factor receptor therapy: The three-year prospective French research axed on tolerance of biotherapies registry. Arthritis Rheum 2009;60:1884-94.
8. Bernatsky S, Habel Y, Rahme E. Observational studies of infections in rheumatoid arthritis: a meta-analysis of tumor necrosis factor antagonists. J Rheumatol 2010;37:928-31.
9. Lee JH, Slifman NR, Gershon SK, Edwards ET, Schwieterman WD, Siegel JN, et al. Life-threatening histoplasmosis complicating immunotherapy with tumor necrosis factor alpha antagonists infliximab and etanercept. Arthritis Rheum 2002;46:2565-70.
10. Curtis JR, Patkar N, Xie A, Martin C, Allison JJ, Saag M, et al. Risk of serious bacterial infections among rheumatoid arthritis patients exposed to tumor necrosis factor antagonists. Arthritis Rheum 2007;56:1125-33.

11. Bernatsky S, Clarke AE, Suissa S. Hematologic malignant neoplasms after drug exposure in rheumatoid arthritis. Arch Intern Med 2008;168:378-81.

12. Geborek P, Bladström A, Turesson C, Gulfe A, Petersson IF, Saxne T, et al. Tumour necrosis factor blockers do not increase overall tumour risk in patients with rheumatoid arthritis, but may be associated with an increased risk of lymphomas. Ann Rheum Dis 2005;64:699-703.

13. Hellgren K, Iliadou A, Rosenquist R, Feltelius N, Backlin C, Enblad G, et al. Rheumatoid arthritis, treatment with corticosteroids, and risk of malignant lymphomas — results from a case-control study. Ann Rheum Dis 2010;69:654-9.

14. Landewe R, van der Heijde D. The validity of a rheumatoid arthritis medical records-based index of severity compared with the DAS28. Arthritis Res Ther 2006;8:107.

15. Dixon WG, Symmons DP, Lunt M, Watson KD, Hyrich KL, British Society for Rheumatology Biologics Register Control Centre Consortium, Silman AJ; British Society for Rheumatology Biologics Register. Serious infection following anti-tumor necrosis factor alpha therapy in patients with rheumatoid arthritis: Lessons from interpreting data from observational studies. Arthritis Rheum 2007;56:2896-904.

16. Hudson M, Suissa S. Avoiding common pitfalls in the analysis of observational studies of new treatments for rheumatoid arthritis. Arthritis Care Res 2010;62:805-10.

17. Walker AM. Confounding by indication. Epidemiology 1996;7:335-6.

18. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med 1991;10:577-81.

19. Landewé RB. The benefits of early treatment in rheumatoid arthritis: confounding by indication, and the issue of timing. Arthritis Rheum 2003;48:1-5.

20. Suissa S, Garbe E. Primer: administrative health databases in observational studies of drug effects — advantages and disadvantages. Nat Clin Pract Rheumatol 2007;3:725-32.

21. Roos LL Jr, Nicol JP, Cageorge SM. Using administrative data for longitudinal research: Comparisons with primary data collection. J Chron Dis 1987;40:41-9.

22. Setoguchi S, Solomon DH, Weinblatt ME, Katz JN, Avorn J, Glynn RJ, et al. Tumor necrosis factor alpha antagonist use and cancer in patients with rheumatoid arthritis. Arthritis Rheum 2006;54:2757-64.

23. Ting G, Schneeweiss S, Katz JN, Weinblatt ME, Cabral D, Scranton RE, et al. Performance of a rheumatoid arthritis records-based index of severity. J Rheumatol 2005;32:1679-87.

24. Ting G, Schneeweiss S, Scranton R, Katz JN, Weinblatt ME, Young M, et al. Development of a health care utilisation data-based index for rheumatoid arthritis severity: A preliminary study. Arthritis Res Ther 2008;10:R95.

25. Consolidated Standards of Reporting Trials. [Internet. Accessed July 22, 2011.] Available from: www.consort-statement.org

26. Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med 2006;144:427-37.

27. Cabral D, Katz JN, Weinblatt ME, Ting G, Avorn J, Solomon DH. Development and assessment of indicators of rheumatoid arthritis severity: results of a Delphi panel. Arthritis Rheum 2005;53:61-6.

28. Sato M, Schneeweiss S, Scranton R, Katz JN, Weinblatt ME, Avorn J, et al. The validity of a rheumatoid arthritis medical records-based index of severity compared with the DAS28. Arthritis Res Ther 2006;8:R57.

29. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Boston: Lippincott Williams & Wilkins; 2008.

30. Wolfe F, Michaud K, Simon T. Can severity be predicted by treatment variables in rheumatoid arthritis administrative databases? J Rheumatol 2006;33:1952-6.

31. Wolfe F, Michaud K, Pincus T. A composite disease activity scale for clinical practice, observational studies, and clinical trials: the Patient Activity Scale (PAS/PAS-II). J Rheumatol 2005;32:2410-5.

32. Pai M, McCulloch M, Enanoria W, Colford JM Jr. Systematic reviews of diagnostic test evaluations: What's behind the scenes? ACP J Club 2004;141:A11-13.

33. Ward MM. Rheumatology visit frequency and changes in functional disability and pain in patients with rheumatoid arthritis. J Rheumatol 1997;24:35-42.

34. Smitten AL, Choi HK, Hochberg MC, Suissa S, Simon TA, Testa MA, et al. The risk of hospitalized infection in patients with rheumatoid arthritis. J Rheumatol 2008;35:387-93.

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

36. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. Pharmacoepidemiol Drug Saf 2010;19:537-54.

37. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 2000;29:722-9.

38. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology 2006;17:260-7.