Comorbidities Before and After the Diagnosis of Rheumatoid Arthritis: A Matched Longitudinal Study

Mark R. Tatangelo, 1 George Tomlinson, 1 Edward Keystone, 2 J. Michael Paterson, 3 Nick Bansback, 4 and Claire Bombardier 1

Objective. To determine the contribution of rheumatoid arthritis (RA) to conditions and medical events. A secondary objective is to quantify this association before and after the introduction of biologic medications.

Methods. All data were collected as health administrative data in Ontario, Canada. Patients with RA (n = 136,678) matched 1:1 to a pool of possible controls without RA from 1995 to 2016. The study was a retrospective longitudinal observational administrative data-based cohort study with cases (RA) and controls (two non-RA comparator groups). The main exposure was new-onset RA identified by a validated diagnosis algorithm. The secondary exposure was the calendar year, which provided a natural experiment to compare years in which biologics were unavailable (pre-2001) to increasing utilization over time.

Results. Patients experienced increases in conditions and medical events up to 5 years before RA disease incidence—4.9 conditions per patient-year compared with 4.6 conditions per patient-year in matched controls. Comorbidities increased to 8.7 conditions per patient-year in the year of RA incidence but were lower in the years after diagnosis—6.9 conditions per patient-year at 5 years postdiagnosis.

Conclusion. This study reframes the clinical manifestations of RA with detailed data on the marginal contribution of RA to conditions and medical events. These results show that a large portion of disease burden is due to the indirect effects of RA.

INTRODUCTION

There is increasing evidence that the effects of rheumatoid arthritis (RA) extend beyond the joints to other organ systems through systemic inflammation (1–8). RA-related inflammation has been linked to a range of conditions, from depression to cardiovascular and pulmonary diseases (3). A comorbid condition acquired by a patient with RA could be a result of RA or occur irrespective of RA diagnosis (9).

Medications for the treatment of RA have been studied in clinical trials and phase 4 observational studies, showing their efficacy, safety, and effectiveness for reducing the burden due to RA.
Observational research often relies on matched designs in which patients with RA are matched to controls on age and sex alone then followed forward for outcomes with differences in occurrence between the groups attributed to RA. Matching studies may misrepresent the attributable comorbidities of RA because these patients are likely to have different prediagnosis medical histories compared with age- and sex-matched non-RA controls. The advantage of adding additional matching covariates to observational studies in addition to age and sex ensures that the selected controls are more similar in medical history to RA cases. More recently developed statistical techniques can be used to separate the direct effects of RA from underlying age- and sex-related risk and from the indirect effects of RA, such as conditions and events arising from the acquisition of RA that are not direct manifestations of the disease. These statistical techniques allow for testing the association of RA with conditions and medical events, and also time-varying exposures, such as, for instance, the introduction of biologics (16,17).

METHODS

Study design. The occurrence of each condition or medical event can be classified broadly into three parts (18) (Figure 1): (a) Direct comorbidities of RA (RA cases – age-/sex-/medical history–matched non-RA controls), (b) indirect comorbidities of RA (age-/sex-/medical history–matched non-RA controls – age-/sex-matched non-RA controls), and (c) underlying comorbidities that are a secular control for comorbidities of an “average patient” of the same age and sex (age-/sex-matched non-RA controls).

The study was a matched longitudinal cohort study using individual-level administrative data. Data were housed at the ICES (formerly the Institute for Clinical Evaluative Sciences (19)), a
nonprofit research institute funded by the Ontario Ministry of Health, which holds administrative data for all health care billings from a publicly funded single payor, meaning all patients under study had identical health insurance coverage.

The study was approved by the University Health Network Research Ethics Board. All analyses were conducted in R 3.4.1 (20) statistical software with data abstracted using SAS Enterprise 7.1 (21). High-performance computing was conducted in Redhat Linux 7 (22) with IBM LSF (23) and SLURM workload management software (24).

Population. Cases were health-insured residents of Ontario, Canada, for at least 1 year who received a diagnosis of RA from age 15 or older from 1995 to 2016 inclusive (n = 136678). Health insurance in Ontario covers physician visits, acute care hospital use for all patients, and all prescriptions for patients over the age of 65. RA incidence date was determined using a validated algorithm (25). Eligible controls in a given year were insured residents of Ontario but were not diagnosed with RA.

Primary exposure. The exposure was the RA diagnosis date defined as the calendar year that a patient met the validated administrative algorithm definition.

Secondary exposure. To measure the association of biologics and the risk of developing comorbidities, the calendar year of RA diagnosis was used as a proxy for population exposure to biologics using a natural experiment as a bias reduction technique (26,27). Biologics were introduced to the Canadian market in 2001, with increased utilization over time.

Outcomes. Outcomes were counts of per-patient conditions measured yearly, operationalized by The Johns Hopkins Adjusted Clinical Groups Case Mix System: Version 10, Expanded Diagnosis Clusters (EDC) (28), a tool that allows for the identification of people with specific diseases or symptoms. The EDC methodology assigns event-level classifiers to each diagnosis code found in the Ontario Health Insurance Plans physician services claims, the Canadian Institute for Health Information Discharge Abstract Database, which captures inpatient hospitalizations, and The National Ambulatory Care Reporting System (29), which captures day surgery, outpatient/community-based clinics, and emergency department visits. Discrete encounters from these three data sources were assigned to one of 264 EDCs and were further organized into 27 categories called Major Expanded Diagnosis Clusters (MEDCs) and grouped by severity (high, eg, myocardial infarction; medium; and low, eg, allergic rhinitis). As broad groupings of diagnosis codes, MEDCs help to control differences in coding behavior between practitioners (eAppendices 1 and 2). Our study used EDCs as the primary outcome measure while reporting individual MEDCs to provide measurement of specific comorbid conditions.

Outcomes for each group were reported as a percentage of total comorbid conditions, counts by exposure group, and percentage of differences between cases and each control group. Summary measures were reported between cases and controls at 5, 2, and 1 year before and after RA diagnosis and during the year of RA diagnosis by the exposure group (Figure 2).

For the secondary outcome of the association of biologics on immediate and future comorbid conditions, counts per patient per year of EDC conditions and medical events were compared in years 1995, 2000, 2005, 2010, and 2015. We infer that biologics reduce downstream comorbid conditions if the difference in the occurrence of comorbid conditions between patients with RA and controls decreases in later time periods. Biologics medications received by patients were measured using the Ontario Drug Benefits Program (public formulary), which contains all prescriptions for patients over the age of 65, receiving government disability benefits, or on social assistance.

Statistical analysis. Two matched control groups without RA were created by sampling with replacement to provide each case with an eligible control at the year of index (1:1 case to control match in each group).

The first non-RA control group was matched only to the birth year and sex of cases. The second non-RA control group was also matched on year of birth and sex but similar medical history in each year before RA diagnosis was added. Medical history before diagnosis was operationalized by selecting the control with the smallest pairwise Mahalanobis distance to each case based on the similarity of the cumulatively available medical history (30–35). This technique uses all given covariates, normalizes them on a common scale based on the shape of the variance of each covariate, and gives each case a “similarity score” to each potential control. The similarity score compares the mean of each case to the mean of the controls and upweights or downweights the importance of the variable based on the variance and the direction the mean deviates from the variance (30–35):

\[ D^2 = (x-m)^T C^{-1} (x-m) \]

where \( D^2 \) = Mahalanobis distance, \( x \) = mean of the data variable, \( m \) = mean of population on data variable, \( C^{-1} \) = inverse covariance matrix of the covariates, and \( T \) = indicates vectors should be transposed.

The more similar a case is to a control, the lower the similarity score. Therefore, two patients would be identical on all covariates if their pairwise “similarity score” was 0 (totally similar medical

*One hospital admission or three physicians’ claims bearing an RA diagnosis (with at least one provided by a musculoskeletal specialist) within 2 years are included in the database, first triggering event in algorithm considered index point.
COMORBIDITIES OF RHEUMATOID ARTHRITIS

Figure 2. Annual per-patient Expanded Major Expanded Diagnosis Clusters (MEDC) comorbidities before and after rheumatoid arthritis diagnosis. Y-Axis: MEDC Comorbidities Per Patient. X-Axis: Time in Relation to RA Incidence (Years). Grouping Variable: Exposure Group. [Color figure can be viewed at wileyonlinelibrary.com]

history). Twenty-seven MEDC covariates (eAppendixes 1 and 2) were included in the Mahalanobis distance measured in each year before RA diagnosis to ensure similar medical histories between cases and controls (35).

The addition of medical history to age- and sex-matched controls ensured that cases were similar to controls on other conditions, controlling for comorbidities other than RA and the potential for interactions between other comorbid conditions. For example, without controlling simultaneously for both diabetes and hypertension, if diabetes was uncontrolled, higher levels of hypertension might be observed because of diabetes (uncontrolled variable) rather than RA.

RESULTS

Patient demographics. There were 136,678 patients with RA with a diagnosis from 1995 to 2016 (Table 1, RA Cases), and 136,678 controls matched on identical birth year, sex, and preexisting medical history (Table 1, Age/Sex/Medical History Matched Controls) with a second 136,678 patient control group matched only by birth year and sex (Table 1, Age/Sex Matched Controls). The pool of unique available controls for matching was 18.6 million residents alive and eligible for health care within the study window.

At diagnosis, RA cases were predominantly female (70%, n=95,073), had a mean age of birth of 1950 (SD = 17.7, range = 1910-2000), mean age of 56.8 (SD = 16.7, range 15-99) with 2006 being the mean year of diagnosis (SD = 6.3, range = 1995-2016).

Main effects for annual comorbidities before and after diagnosis. Our results (Figure 2) showed that patients with RA accured more total comorbid conditions before and after diagnosis compared with age-/sex-/medical history–matched controls (non-RA) and age-/sex-matched controls (non-RA). Differences in comorbidities of patients with RA (4.9/patient) and matched controls (age-/sex-/medical history–matched, 4.57/patient; age-/sex-matched, 3.35/patient) were observable 5 years before the diagnosis of RA. These differences grew over time and peaked in the year of diagnosis, in which patients with RA accrued 11,914,59 (8.7/patient) medical events or conditions compared with 712,456 (5.2/patient) for age-/sex-/medical history–matched controls (non-RA), and 514,520 (3.7/patient) for age-/sex-matched controls (non-RA). These differences mean that the patients with RA had 67.2% more
comorbidities than age-/sex-/medical history–matched controls and 131.6% more comorbidities compared with age-/sex-matched controls in the year of RA diagnosis. At 1 year postdiagnosis, the conditions and medical events experienced by patients with RA declined slightly to 975,144 (7.6/patient) compared with 675,270 (5.3/patient) for age-/sex-/medical history–matched controls (non-RA), reducing the increase in comorbidities between patients with RA and matched controls to 93.6%. At 3 years postdiagnosis, the comorbidities experienced by patients with RA decreased further to 859,648 (7.2/patient) compared with 627,980 (5.3), Δ37.6% for age-/sex-/medical history–matched controls (non-RA), and 477,623 (3.9), Δ82.4% for age-/sex-matched controls (non-RA). These differences further declined at 5 years postdiagnosis to 619,259 (6.9/patient) comorbidities at 5 years postdiagnosis compared with 488,628 (5.3), Δ29.5% for age-/sex-/medical history–matched controls (non-RA) and 385,143 (4.1), Δ66.8% for age-/sex-matched controls (non-RA).

The differences between patients with RA and matched controls were observable 5 years before RA diagnosis, rose until the RA diagnosis year, then declined in the years after RA diagnosis.

In terms of specific comorbidities incurred by patients with RA, the top five comorbidities from a volume perspective at the year of incidence were musculoskeletal (191.9 thousand), rheumatologic (151.9 thousand), cardiovascular (100.3 thousand), eye (80.2 thousand), and neurologic (77.7 thousand) conditions (Table 1). Some notable differences between patients with RA and controls in the year of diagnosis were rheumatologic (3266%), musculoskeletal (148%), hematologic (134%), neurologic (120%), toxic effects and adverse events (106%), renal (105%), general signs and symptoms (85%), and infections (55%) (Table 1. Age/Sex/Medical History Matched Controls, %).

### Table 1. Demographics and comorbidities in the year of rheumatoid arthritis diagnosis

| Demographic Variables at Diagnosis Year | RA Cases | Age-/Sex-/Medical History–Matched Control (Non-RA) | Age-/Sex-Matched Control (Non-RA) |
|----------------------------------------|----------|--------------------------------------------------|----------------------------------|
| Number of patients                     | 136,678  | 136,678                                          | 136,678                          |
| Age (mean, median, SD, range)          | 56.8, 58, 16.7, (15-99) | 56.8,58,16.7, (15-99) | 56.8,58,16.7, (15-99) |
| Female sex (% count)                   | 70, 95,073 | 70, 95,073                                       | 70, 95,073                       |
| Year of birth (mean, median, SD, range)| 1950, 1949, 17.7, (1910-2000) | 1950,1949,17.7,(1910-2000) | 1950,1949,17.7,(1910-2000) |
| Year of RA diagnosis (mean, median, SD, range) | 2006, 2007, 6.3, (1995-2016) | ...                                               | ...                             |

**Abbreviations:** EDC, expanded diagnosis cluster; MEDC, major EDC; RA, rheumatoid arthritis.
Analysis of annual comorbidities before and after diagnosis grouped by calendar year of RA diagnosis. In Figure 3, we outline the annual per-patient EDC comorbidities before and after RA diagnosis, grouping patients by calendar year of diagnosis. When viewed as a total aggregate grouped by year of RA (eTable 1), conditions and medical events operationalized by EDC comorbidities per year (y-axis) in the RA group grew slightly faster at each subsequent year of diagnosis (x-axis) when compared with medical history–matched controls. However, after grouping patients by calendar year of diagnosis, a decrease was observed between conditions accrued among RA cases and matched controls who received a diagnosis in later calendar years when biologics were available (2005, 2010, 2015) compared with earlier calendar years when they were unavailable (1995, 2000). Biologics prescriptions increased in each calendar year from 71 in 2001 to 6207 in 2015 as a total number of prescriptions (eTable 3). Incident users included in the cohort had increased biologics relative to year of diagnosis from 1149 (0.008 prescriptions per patient) in the year of diagnosis to 2694 (0.03 prescriptions per patient) at 5 years postdiagnosis, representing a 275% increase in prescription percentage over the first 5 years postdiagnosis (eTables 1 and 4).

Comorbid conditions before and after RA diagnosis by severity level. Figure 4 shows the annual per-patient comorbidities in EDC for patients with RA and matched controls grouped by severity level (low, medium, high). Most of the comorbidities for both patients with RA and matched controls were low severity conditions (green) and medium severity conditions (orange) comorbidities with a smaller proportion of overall comorbidities being high severity (red). These findings indicate that most of the additional comorbid conditions related to RA are of low and medium severity. When separated by RA association, about half of the low and medium comorbidities were directly associated with RA diagnosis, and the other half was indirectly associated with RA (Figure 4). This is indicated by the values of the age-, sex-, and medical history–matched groups (dotted lines), which is about 50% lower than the age- and sex-matched groups (solid lines). High-severity comorbidities occur less frequently and are almost solely attributable...
A notable finding is that the lower-severity comorbidities of RA appear earlier (up to 12 years) before diagnosis compared with high-severity comorbidities, which appear 1 year before RA diagnosis. Another notable finding is that patients with RA experience decreasing low- and medium-severity comorbidities over time since diagnosis compared with matched controls, whereas the difference in high-severity comorbidities remains constant in the years after diagnosis between patients with RA and matched controls.

**DISCUSSION**

The clinical implications of the study findings show several interesting patterns on the accrual of comorbid conditions over disease and calendar time for patients with RA. We observed differences in total comorbid conditions up to 5 years before diagnosis, with comorbidities rising until the year of diagnosis and then decreasing up to 5 years postdiagnosis.

We found that the top five comorbid conditions by volume in the year of diagnosis were rheumatologic, musculoskeletal, cardiovascular, eye, and neurologic, whereas the top five comorbidities directly associated with RA were rheumatologic, musculoskeletal, hematologic, neurologic, and toxic/adverse events.

After stratifying patients by calendar year of diagnosis, patients diagnosed with RA in later calendar years had faster reductions in comorbid conditions after diagnosis compared with patients in earlier time periods.

When grouped by severity, most of the RA-related comorbid conditions were of low or medium severity, and these additional comorbid conditions were detectable up to 12 years before diagnosis, rose until a peak in the year of diagnosis, and decreased relative to non-RA controls after diagnosis. In contrast, severe comorbidities occurred less often, were detectable in patients with RA only 1 year before diagnosis, peaked in the year of diagnosis, and did not decrease relative to matched controls after the year of RA diagnosis. Interestingly, the severe comorbidities were directly RA attributable rather than a result of the indirect effects of RA.
Compared with other studies of comorbidity for patients with RA, we observed similar increases in cardiac, lung, and neurologic (1,36,37) comorbidities. Our approach of accruing comorbidities over time before and after diagnosis was not mimicked in the existing literature, and the use of nondisease control groups was sparse.

Some notable methodological strengths of the study were the use of all available data in the prediagnosis and postdiagnosis period. We utilized flexible match points on the year of RA diagnosis to controls who did not yet have an RA diagnosis. This approach improved the pool of available controls (approximately 1:1400 for males, 1:800 for females). The benefit of having a larger pool of eligible controls is the ability to match patients who are more similar in past medical history compared with an RA case (31,38).

Another study strength was the use of Mahalanobis distance matrices to match RA cases to controls with similar medical histories but without RA. By using a medical history match, we used statistical attribution methods to estimate the directly RA-attributable portion of comorbidity compared with the indirect and underlying comorbidities of RA. Our results on individual comorbidities indicate that some studies examining individual comorbidities may be susceptible to bias in the pre-exposure period, with differences between age-/sex-matched controls and RA cases and age-/sex-/medical history–matched controls and RA cases sometimes exceeding 25%. We found that the more a comorbidity is associated with RA, the greater the degree of estimation error is due to residual confounding.

The study was powered to detect whether the introduction of biologics had no effect on comorbid conditions or was associated with an increased number of comorbid conditions over time. The combination of innovative methodology controlling for medical history and the hypothesis that RA is less severe over time (39–41) means that we expected to observe fewer comorbid conditions in patients diagnosed in later calendar time periods with higher availability of biologics. If no significant clinical benefits or comorbidities prevented are observed in subsequent time periods under a favorable set of assumptions, the plausibility of biologics as a treatment that reduces overall comorbidity, and subsequent health service resource utilization, would be weakened. It is possible that improvements in care are changing the number and distribution of comorbid conditions. However, because this is a matched design, the improvements in care for RA would need to outpace the improvement in other areas of medicine to have a marginal effect on the difference in comorbid conditions. The reason for focusing on biologics as a notable difference over time is that the gradient in uptake of use of these agents is a notable, measurable change over time. Other outcomes like early detection, improved use of high-dose methotrexate in early disease, or even adherence to medications are unlikely to outpace improvements in early detection of other diseases.

Limitations of this study include the lack of disease-specific variables needed to assess the severity of RA over time, which could have further reduced confounding through balancing on clinical covariates. Furthermore, the length of follow-up could be too short to detect important differences in comorbidity acquisition, which we maximized to use all available follow-up. A final limitation was the inability to disaggregate comorbidities to the smallest EDC level due to hardware constraints; more covariates will yield a better match.

This study suggests that though advances have been made in administrative cohort design, researchers are only beginning to scratch the surface of parsing the causal model of RA through phenotypic data from large administrative databases. The advent of low-cost modern computing combined with large administrative databases means researchers increasingly have the ability to use more granular exposure, outcome, and covariate bias reduction methods, which are needed for a disease with this amount of complexity. Future studies could examine individual comorbidities, critical treatment periods, or drug exposures using similar methods.

The findings of our study indicate that the conditions and events arising from RA begin to occur many years before diagnosis, peak at 1 to 2 years postdiagnosis, and taper over time to a steady state, which is higher than similarly matched controls. We also found evidence for the plausibility of better treatment and newer therapies for inhibiting the formation of comorbid conditions after the diagnosis of RA. For clinicians and clinical researchers, the methods and results of our study highlight the importance of considering conditions and medical as part of the inflammatory process.

ACKNOWLEDGMENTS

The authors acknowledge Dr. Allan S. Detsky for writing mentorship.

AUTHOR CONTRIBUTIONS

Dr. Tatangelo drafted the manuscript. Drs. Tatangelo, Tomlinson, Keystone, Bansback, Bombardier and Mr. Paterson critically revised the manuscript for important intellectual content. Drs. Tatangelo, Tomlinson, Bansback, Bombardier and Mr. Paterson obtained funding. Dr. Bombadier was study supervisor.

Study concept and design. Tatangelo, Tomlinson, Keystone, Paterson, Bansback, Bombardier.

Acquisition of data. Tatangelo, Paterson, Tomlinson.

Analysis and interpretation of data. Tatangelo, Tomlinson, Keystone, Paterson, Bansback, Bombardier.

REFERENCES

1. Dougados M, Soubrier M, Antunez A, Balint P, Balsa A, Buch MH, et al. Prevalence of comorbidities in rheumatoid arthritis and evaluation of their monitoring: results of an international, cross-sectional study (COMORA). Ann Rheum Dis 2014;73:62–8.

2. Ranganath VK, Maranian P, Elashoff DA, Woodworth T, Khanna D, Hahn T, et al. Comorbidities are associated with poorer outcomes in community patients with rheumatoid arthritis. Rheumatology (Oxford) 2013;52:1809–17.

3. Ang DC, Choi H, Kroenke K, Wolfe F. Comorbid depression is an independent risk factor for mortality in patients with rheumatoid arthritis. J Rheumatol 2005;32:1013–9.
4. Michaud K, Wolfe F. Comorbidities in rheumatoid arthritis. Best Pract Res Clin Rheumatol 2007;21:885–906.
5. Wright K, Crowson CS, Gabriel SE. Cardiovascular comorbidity in rheumatic diseases: a focus on heart failure. Heart Fail Clin 2014;10:339–52.
6. Crowson CS, Liao KP, Davis JM III, Solomon DH, Matteson EL, Knutson KL, et al. Rheumatoid arthritis and cardiovascular disease. Am Heart J 2013;166:622–8.e1.
7. Simon TA, Thompson A, Gandhi KK, Hochberg MC, Suissa S. Incidence of malignancy in adult patients with rheumatoid arthritis: a meta-analysis. Arthritis Res Ther 2015;17:212.
8. Dougados M. Comorbidities in rheumatoid arthritis. Curr Opin Rheumatol 2016;28:282–8.
9. Aletaha D, Smolen JS. Diagnosis and management of rheumatoid arthritis. JAMA 2018;320:1360–72.
10. Hetland ML. Modern treatment strategies in rheumatoid arthritis. Rheumatology (Oxford) 2011;50:204–13.
11. Singh JA, Christensen R, Wells GA, Suarez-Almazor ME, Buchbinder R, Lopez-Olivo MA, et al. A network meta-analysis of randomized controlled trials of biologics for rheumatoid arthritis: a Cochrane overview. CMAJ 2009;181:787–96.
12. Lu B, Greevy R, Xu X, Beck C. Optimal nonbipartite matching and its statistical applications. Am Stat 2011;65:1582–91.
13. Rubin DB. Matched sampling for causal effects. Cambridge: Springer-Verlag New York; 2009.
14. Donahue KE, Gartlehner G, Schulman ER, Jonas B, Coker-Schwimmer E, Patel SV, et al. CER-211 Drug therapy for early rheumatoid arthritis: a systematic review update [internet]. Report No.: 18-EHC015-EF. Rockville (MD): Agency for Healthcare Research and Quality; July 2018.
15. Jeong H, Baek SY, Kim SW, Eun YH, Kim IY, Kim H, et al. Comorbidities of rheumatoid arthritis: results from the Korean Incidence, prevalence, and severity of rheumatoid arthritis: a systematic literature review. Joint Bone Spine 2016;83:625–30.
16. Minichiello E, Sermoiano L, Boissier MC. Time trends in the incidence, prevalence, and severity of rheumatoid arthritis: a systematic literature review. Joint Bone Spine 2016;83:625–30.