

Prevalence of rheumatoid arthritis in the United States adult population in healthcare claims databases, 2004–2014

Theresa M. Hunter¹ · Natalie N. Boytsov¹ · Xiang Zhang¹ · Krista Schroeder¹ · Kaleb Michaud² · Andre B. Araujo¹

Received: 20 December 2016 / Accepted: 19 April 2017 / Published online: 28 April 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract This study aimed to determine the prevalence of rheumatoid arthritis in the United States (US) adult insured population from 2004 to 2014. This was an observational, retrospective, cross-sectional study based on US administrative health insurance claims databases (Truven Health MarketScan[®] Research database and IMS PharMetrics Plus database). Trends in RA prevalence focusing on the 10-year period covering January 1, 2004–December 31, 2014 were analyzed using a validated algorithm for the identification of RA. Prevalence rates in the databases were determined and age- and gender-adjusted rates were projected to the US population in 2014. Analysis of data from the two databases indicated that the RA prevalence rate in commercially insured adult US population ranged from 0.41 to 0.54% from 2004 to 2014. The prevalence varied substantially by gender and age in each year and increased gradually across the years for most subgroups. In 2014, out of 31,316,902 adult patients with continuous enrollment in the Truven Health MarketScan[®] Research database, 157,634 (0.50%) patients met our criteria for RA. Similarly, out of 35,083,356 adult patients in the IMS PharMetrics Plus database, 139,300 (0.50%) patients met our criteria for RA. In 2014, the overall age-adjusted prevalence of RA ranged

from 0.53 to 0.55% (0.29–0.31% for males and 0.73–0.78% for females). The prevalence of RA in the US appeared to increase during the period from 2004 to 2014, affecting a conservative estimate of 1.28–1.36 million adults in 2014.

Keywords Rheumatoid arthritis · Prevalence · Claims databases

Introduction

Over the past two decades, there has been much variation between studies that report the prevalence of rheumatoid arthritis (RA) [1]. While there have been several RA prevalence studies conducted in Europe [2–6], limited prevalence studies have been conducted in the United States (US). The most frequently referenced study on RA prevalence in the US assessed data from 1955 to 1985 and found that there was a prevalence of 1073 per 100,000 population in 1985 [7]. This study only analyzed data from Olmsted County, Minnesota, generalizable to the white population and is now 30 years old [7]. Recent studies have attempted to assess the prevalence of RA in the US, yet their generalizability to the overall US adult population is uncertain [8–10].

In addition to being outdated, there are several methodological variations among previous RA prevalence studies. The variation in the algorithms used for patient identification is a key limitation found in RA prevalence studies that utilize administrative claims databases [11–13]. When using the rheumatologist's diagnosis as the gold standard, the overall accuracy of algorithms used to identify RA cases in administrative claim-based studies differ, causing wide variations in RA prevalence rates (0.15–0.61%) [13]. Therefore, to understand US RA prevalence, additional

Electronic supplementary material The online version of this article (doi:10.1007/s00296-017-3726-1) contains supplementary material, which is available to authorized users.

✉ Theresa M. Hunter
hunter_theresa_marie@lilly.com

¹ Global Patient Outcomes and Real World Evidence, Eli Lilly and Company, Lilly Technology Center South, 1555 South Harding Street, Indianapolis, IN 46221, USA

² Division of Rheumatology, University of Nebraska Medical Center, Omaha, NE, USA

studies using validated RA case identification algorithms are needed.

The purpose of this study was to assess the current prevalence of RA among commercially insured adults in the US. To do this, data from administrative insurance claims databases over the period 2004–2014 were analyzed using a validated algorithm for the identification of RA. We sought to determine the prevalence of RA among the insured US adult population and its variations according to gender, age, and geographical regions. Our findings can be used to inform the scientific and medical community on the prevalence of RA among commercially insured adults in the US and are needed to understand the economic burden of RA on the US healthcare system.

Methods

Study design

This study was an observational, retrospective, cross-sectional study based on two US administrative insurance claims databases. First, data from Truven Health MarketScan® Research database (Truven Health, Ann Arbor, MI, USA) were analyzed to assess trends in RA prevalence focusing on the 10-year period covering January 1, 2004–December 31, 2014. Prevalence rates were analyzed overall and stratified by age and gender. For the 2014 population, demographic characteristics were assessed, and the age-adjusted prevalence rate was measured overall and by gender. Additionally, for comparative purposes, prevalence rates assessed from the IMS PharMetrics Plus database (IMS Health, Waltham, MA, USA) were also reported from January 1, 2006 to December 31, 2014.

The setting for this study was US clinical practice, as reflected by the insurance claims in the databases. Truven Health MarketScan® Commercial Claims and Encounters and Medicare Supplemental databases contain de-identified data on over 50 million covered lives and capture the continuum of care in all settings including physician office visits, hospital stays, and pharmacies. The IMS PharMetrics Plus database is the largest claims database of integrated medical claims in the US and is comprised of adjudicated claims for more than 150 million unique enrollees across the US.

Study variables were defined in the Truven Health MarketScan® Research and IMS PharMetrics Plus databases using enrollment records and International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes. All data from the databases are Health Insurance Portability and Accountability Act (HIPAA) compliant to protect patient privacy.

Identification of RA in claims database

Several published algorithms have been utilized to define RA in claims databases. When choosing a definition for our study, we assessed the published sensitivity, specificity, and accuracy of multiple potential case definitions for RA that were validated in the US [14, 15] and Canada [13, 16–18]. The RA algorithms that were assessed differed in the number of diagnostic codes and source of diagnoses (rheumatologist versus general practice physician), and varied in regards to specificity, sensitivity, and positive and negative predictive values. The algorithm chosen for this study had a sensitivity of 92.0, a specificity of 74.3, and an accuracy of 77.8 [13], which was deemed appropriate by the research team.

For each calendar year of analysis, a base cohort was assembled that consisted of all patients ≥ 18 years on January 1st of the calendar year with continuous enrollment in medical benefits throughout the calendar year allowing for an enrollment gap of <30 days. From these base cohorts, utilizing ICD-9 codes, the sub-set of patients with RA were identified by the following diagnostic criteria:

- two non-rheumatology physician visits with a listed RA code (ICD-9: 714.0, 714.1, 714.2) occurring at least 2 months apart;
- or at least one RA code contributed by a rheumatologist;
- or at least one inpatient hospitalization for which RA was in the diagnostic codes.

Within this definition, patients were not counted as having RA if they had at least two visits, at least two months apart, subsequent to the second RA visit above (if a second visit occurred), with two identical diagnoses of other autoimmune and connective tissue diseases [psoriatic arthritis (ICD-9: 696.0), ankylosing spondylitis (ICD-9: 720.0), and other spondyloarthropathies (ICD-9: 720.1, 720.2, 720.8, 720.9), systemic lupus erythematosus (ICD-9: 710.0), scleroderma (ICD-9: 710.1), Sjögren's syndrome (ICD-9: 710.2), dermatomyositis (ICD-9: 710.3), polymyositis (ICD-9: 710.4), primary systemic vasculitis (ICD-9: 446.0, 446.2, 446.4, 446.5, 446.7, 447.6) and other connective diseases (ICD-9: 710.5, 710.8, 710.9)] [13].

A second RA case definition was tested for comparative purposes that did not exclude comorbidities. Utilizing ICD-9-CM codes, this sub-set of patients with RA was defined by the following diagnostic criteria: two physician visits for RA at least 2 months apart [14] or at least one hospitalization where RA was in the diagnostic codes [13, 18]. These case definitions were previously tested and validated [14, 18].

Estimation of prevalence

Prevalence is defined as the proportion of individuals who have the disease of interest in a specified time period (includes both new and existing cases). In our study, annual RA prevalence was estimated using the US adult population in the US health claims databases during the period of 2004–2014. For each calendar year, a base cohort was assembled and the case identification algorithm was applied separately in each year. The numerator in the prevalence estimation was the number of patients that met the RA definition described in the previous section. The denominator was the number of patients in the base cohort.

Statistical analyses

RA prevalence was estimated for subgroups stratified by gender and age (18–34, 35–44, 45–54, 55–64, and ≥ 65) for each calendar year from 2004 to 2014. To account for the distortion caused by the age distributions in the datasets, we also calculated the age-adjusted prevalence of RA from 2004 to 2014 using direct standardization. The age- and gender-specific prevalence rates in 2014 were applied to the corresponding population estimates from the US Census Bureau to project the total number of persons in the US expected to have RA in 2014 and in 2020.

Results

Rheumatoid arthritis prevalence: 2004–2014

Annual RA prevalence rates ranged from 0.41 to 0.52% from 2004 to 2014 for adult US patients in the Truven MarketScan® Research database. The prevalence varied

substantially by gender and age in each year and increased gradually across the years for most subgroups. Specifically, prevalence among females was more than twice the prevalence among males (Fig. 1). In the Truven MarketScan® Research database, overall prevalence in females gradually increased from 0.56% in 2004 to 0.71% in 2014, whereas the overall prevalence among males remained relatively stable over the same period (0.23% in 2004 to 0.26% in 2014) (Fig. 1). At the same time, RA prevalence increased with age among both males and females, and for most age groups the rates rose consistently across the study period (Fig. 2a, b). We also calculated the age-adjusted prevalence rates from 2004 to 2014, which ranged from 0.37 to 0.55%.

The overall RA prevalence rate for the adult US population in the IMS PharMetrics Plus database was similar to the rate in Truven Health MarketScan® Research database and ranged from 0.47 to 0.54% from 2006 to 2014. Similar to the findings in the Truven Health MarketScan® Research database, the prevalence varied substantially by gender and age in each year and increased gradually across the years for most subgroups. Rheumatoid arthritis prevalence increased with age among both males and females, and for most age groups the rates rose consistently across study years (Supplemental Figure 1A, B).

Age-adjusted RA prevalence in 2014: Truven Health MarketScan® Research database

In 2014, out of a total of 31,316,902 adult patients with continuous enrollment in the Truven Health MarketScan® Research database, there were 157,634 (0.50%) patients with RA. Of these 157,634 patients, 119,692 (75.93%) were female and 37,942 (24.07%) were male. Mean age for overall RA population was 57.42 years [standard deviation (SD) 13.32]. A majority of patients were

Fig. 1 Rheumatoid arthritis prevalence trends stratified by gender (2004–2014)

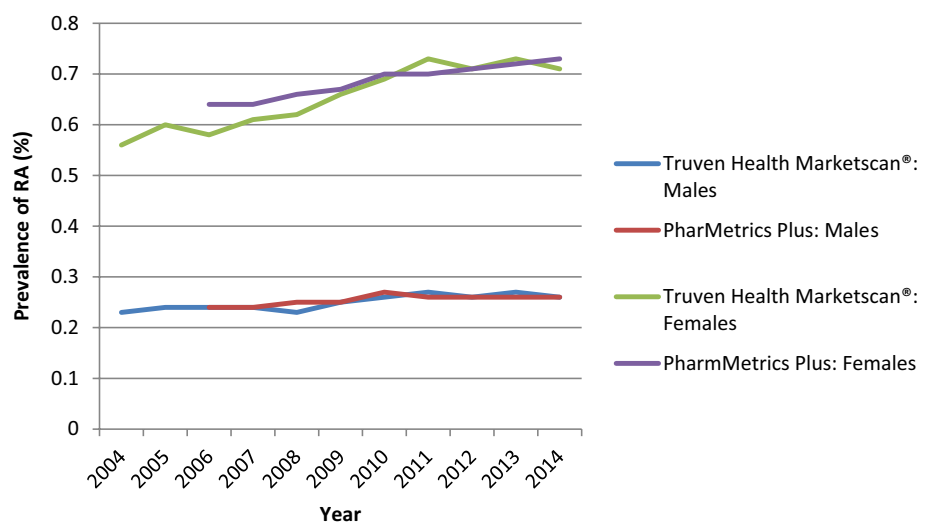
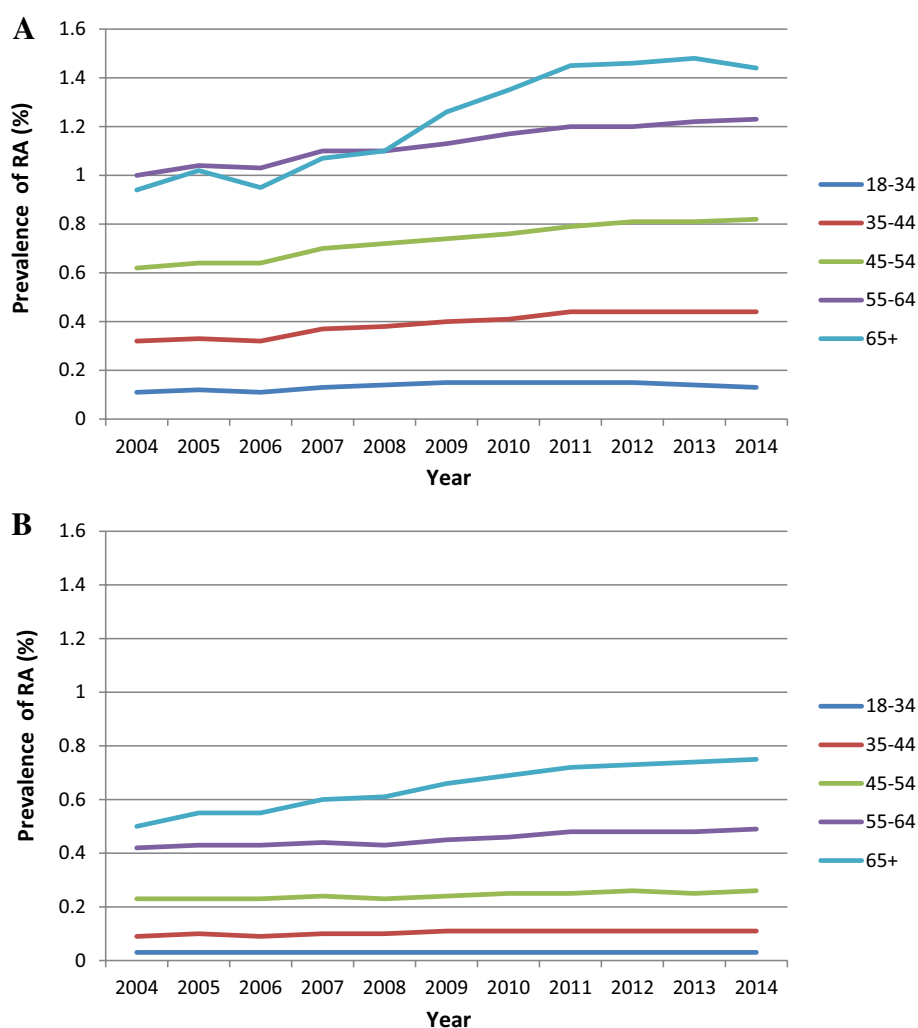


Fig. 2 a Rheumatoid arthritis prevalence among females stratified by age. **b** Rheumatoid arthritis prevalence among males stratified by age. Source: Truven Health MarketScan® Research Database



commercially insured and located in the Atlantic and Central US regions. The patients' demographic information is presented in Table 1.

The overall age-adjusted prevalence of RA among individuals who were aged 18 years or older on January 1, 2014 was 0.53%. Males had an age-adjusted prevalence of 0.29% and females had an age-adjusted prevalence of 0.73% in 2014.

Age-adjusted RA prevalence in 2014: IMS PharMetrics Plus database

In 2014, out of 35,083,356 adult patients in the IMS PharMetrics Plus database, there were 139,300 (0.50%) patients with RA. Of these patients, 103,442 (74.26%) were female, and 35,858 (25.74%) were male. The patients' demographic information is presented in Table 1. Mean age for the overall RA population was 56.70 (SD 12.4).

The overall age-adjusted prevalence of RA among individuals who were aged 18 years or older on January 1,

2014 was 0.55%. Males had an age-adjusted prevalence of 0.31% and females had an age-adjusted prevalence of 0.78% in 2014.

Rheumatoid arthritis prevalence—US estimates

Using population estimates from the US Census Bureau and the RA prevalence rates in 2014, it was estimated that 1.28 million (Truven Health MarketScan® Research) to 1.36 million (IMS PharMetrics Plus) US adults were affected by RA in 2014. If age- and gender-specific RA prevalence rates remain the same, it is projected that RA will affect 1.39 million US adults by 2020. Age- and gender-specific population estimates for RA are shown in Table 2.

Discussion

This study evaluated recent trends in prevalence of RA and helped to highlight the estimated burden of RA in the US.

Table 1 Baseline characteristics for patients with rheumatoid arthritis (2014)

Variables	Truven Health MarketScan [®] Research (N = 157,634)	IMS PharMetrics Plus (N = 139,300)
Sex		
Male	37,942 (24.07%)	35,858 (25.74%)
Female	119,692 (75.93%)	103,442 (74.26%)
Age (years)		
18–34	7749 (4.92%)	6878 (4.94%)
35–44	16,822 (10.67%)	14,364 (10.31%)
45–54	37,332 (23.68%)	33,354 (23.94%)
55–64	54,983 (34.88%)	53,670 (38.53%)
65+	40,748 (25.85%)	31,034 (22.28%)
Insurance		
Commercial	113,215 (71.82%)	136,567 (98.03%)
Medicare	44,419 (28.18%)	1137 (0.82%)
Other	N/A	1596 (1.15%)
Geographic region divisions		
New England	6585 (4.18%)	7133 (5.12%)
Mid Atlantic	28,288 (17.94%)	19,093 (13.71%)
South Atlantic	30,484 (19.34%)	24,168 (17.35%)
East North Central	28,888 (18.33%)	27,519 (19.76%)
East South Central	10,362 (6.57%)	16,012 (11.49%)
West North Central	5952 (3.78%)	13,164 (9.45%)
West South Central	17,015 (10.79%)	20,357 (14.61%)
Mountain	8880 (5.63%)	5728 (4.11%)
Pacific	17,633 (11.19%)	5071 (3.64%)
Unknown	3547 (2.25%)	1055 (0.72%)

Table 2 2014 US census projected RA population estimates stratified by gender and age (years)

	Truven Health MarketScan [®] Research	IMS PharMetrics Plus
Male		
18–34	11,429	11,231
35–44	22,175	22,242
45–54	55,705	55,357
55–64	94,677	98,811
65+	152,635	161,242
Total	336,621	365,167
Female		
18–34	47,951	51,994
35–44	89,557	92,065
45–54	180,677	180,091
55–64	255,295	261,175
65+	372,844	430,223
Total	946,324	984,084
Total	1,282,945	1,364,431

This study provided the prevalence of RA during the last decade (2004–2014) in the US commercially insured adult population using two US administrative insurance claims databases (Truven Health MarketScan[®] Research and IMS PharMetrics Plus). The findings from this study indicate an increase in the RA population in the US from 2004 to 2014. Based on these findings, it is estimated that approximately 1.3 million adults were affected by RA in 2014. The findings from this study can be used to inform the scientific and medical community on the prevalence of RA among commercially insured adults in the US and help providers, payers, and patients to better understand the economic burden of RA in the US.

In the US, there have been limited studies of RA prevalence. The studies that have been published differ considerably in their methods of identifying RA patients and result in a wide variation of prevalence estimates. This study utilized a validated definition [13] of RA and administrative claims data to provide consistent estimates of RA prevalence across two different databases.

The study determined that overall prevalence of RA in the US ranged from 0.41 to 0.54% and steadily increased from 2004 to 2014. When analyzing medical expenditure panel survey (MEPS) data, Simmons and colleagues found similar results: 0.40% in 2004, 0.44% in 2005, and 0.43% in 2006 [8]. These findings were lower than the rates reported by other studies. For example, based on the 2001–2005 National Ambulatory Medical Care Survey data, RA prevalence was 1.48% [10]. The widely referenced RA prevalence based on the Olmsted County cohort was reported at 0.72% in 2005 [19]. Limitations of these studies, such as lack of generalizability [19] and identification of RA patients by a single occurrence of RA diagnostic code [10], may have overestimated these rates.

The authors of the Olmsted County cohort estimated that RA affected 1.5 million US adults in 2005 [19]. Based on our estimates of national claims databases, RA prevalence in 2005 was approximately 0.44% with an estimated 0.95 million people affected. These discrepant findings may stem from differences in gender- and age-specific prevalence rates, which were higher in Olmsted County compared to the US.

Studies have consistently documented a greater prevalence of RA in women versus men; however, the relative burden differs across studies. For instance, the Olmsted County cohort found that the RA prevalence rates in women were approximately double the prevalence rates in men [19]. Our study found that RA prevalence rates among women were closer to three times higher than the rates in men, which is consistent with the results from the MEPS study [8].

It can be inferred that many of the differences in RA prevalence estimates result from the methodological variations between the previous studies. When utilizing administrative claims databases, the variations in the methods of identifying RA patients cause differentiation among estimates and results [11–13]. This study measured prevalence rates in two different large, geographically dispersed claims databases, the Truven Health MarketScan® Research and IMS PharMetrics Plus, using a robust RA case definition assessed by high sensitivity (92.0) and specificity (74.3) [13].

It is important to note that the primary purpose of insurance claims data is administrative and not research-oriented. Therefore, there are limitations to using claims data and ICD-9 codes provided for insurance claims to determine prevalence of a disease. Due to inherent limitations of claims-based data sources, there may have been a proportion of cases identified using the chosen criteria as having RA, when in fact, they might not have RA. The potential for misclassification of non-RA patients as RA patients in claims data, especially in the case of “rule-out” diagnoses (RA diagnoses coded in laboratory work-ups when RA is suspected or needs to be “ruled-out”) should be taken into

consideration when evaluating prevalence rates. This type of misclassification could potentially lead to overestimation of prevalence. Given that laboratory test results and medical charts review were not included in this study, confirmation of RA cases was not possible. Instead, we relied on a published RA definition with a level of sensitivity and specificity, and accuracy that we deemed acceptable [13].

Additionally, given both that the Truven Health MarketScan® Research and the IMS PharMetrics Plus populations are composed of patients with commercial insurance or Medicare supplement insurance, there are specific groups (uninsured people, military personnel, Medicaid patients, and Medicare enrollees without an employer-sponsored supplement plan) that are not represented in our analyses. Approximately, one-third of the RA patients report disease-related work disability [20, 21]. Given the limitations of the databases used in this study, unemployed RA patients would not be included in our analyses, thus resulting in a conservative estimate of RA prevalence in the US. There may also be RA patients that were not accounted for in these analyses because they did not meet the RA definition that was used in this study. This would also result in a conservative estimate of the RA prevalence in the US. Additionally, all individuals over 65 in our study are insured by an employer-sponsored Medicare supplemental plan. This population represents a very specific and relatively small proportion of the over 65 populations in the US. There is also the potential for the two data sets to overlap, but the actual overlap cannot be determined, because the databases are not linked.

Conclusion

The large sample size and dispersed geographic representation of our study enhances the validity of generalizing our prevalence estimates to the general US adult population that are commercially insured, and the consistency in rates observed in the two databases strengthens our observations. We observed that the prevalence of RA in the US appeared to increase during the period 2004–2014, affecting approximately 1.3 million adults in 2014. These results may be attributed to the increasing emphasis on early diagnosis of RA, regular monitoring of disease activity, increased life expectancy, as well as a growing elderly population.

Author contributions TMH, NNB, XZ, KS, KM, and AA all made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; participated in drafting the manuscript or revising it critically for important intellectual content; approved the final version of the submitted manuscript and agreed to

be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Compliance with ethical standards

Conflict of interest The authors declare that there are no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Ethical approval The study has been performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and later amendments.

Human or animal participants This study was a retrospective analysis of administrative claims data and does not contain any human participants or animals.

Informed consent The permission from a research ethics committee was not required and formal informed consent was not obtained.

References

- Alamanos Y, Drosos AA (2005) Epidemiology of adult rheumatoid arthritis. *Autoimmun Rev* 4:130–136
- Riise T, Jacobsen BK, Gran JT (2001) Incidence and prevalence of rheumatoid arthritis in the county of Troms, northern Norway. *J Rheumatol* 27:1386–1389
- Simonsson M, Bergman S, Jacobsson LT et al (1999) The prevalence of rheumatoid arthritis in Sweden. *Scand J Rheumatol* 28:340–343
- Symmons D, Turner G, Webb R et al (2002) The prevalence of rheumatoid arthritis in the UK: new estimates for a new century. *Rheumatology (Oxford)* 41:793–800
- Power D, Codd M, Ivers L et al (1999) Prevalence of rheumatoid arthritis in Dublin, Ireland: a population based survey. *Ir J Med Sci* 168:197–200
- Aho K, Kaipiainen-Seppänen O, Heliövaara M et al (1998) Epidemiology of rheumatoid arthritis in Finland. *Semin Arthritis Rheum* 27:325–334
- Gabriel SE, Crowson CS, O’Fallon WM (1999) The epidemiology of rheumatoid arthritis in Rochester, Minnesota, 1955–1985. *Arthritis Rheum* 42:415–420
- Simons WR, Rosenblatt LC, Trivedi DN (2012) The economic consequences of rheumatoid arthritis: analysis of Medical Expenditure Panel Survey 2004, 2005, and 2006 Data. *J Occup Environ Med* 54:48–55
- Helmick CG, Felson DT, Lawrence RC et al (2008) Estimates of the prevalence of arthritis and other rheumatic condition in the US. *Arthritis Rheum* 58:15–25
- Sacks JJ, Luo YH, Helmick CG (2010) Prevalence of specific types of arthritis and other rheumatic conditions in the ambulatory health care system in the US, 2001–2005. *Arthritis Care Res* 62:460–464
- Widdifield J, Labrecque J, Lix L et al (2013) Systematic review and critical appraisal of validation studies to identify rheumatic diseases in health administrative databases. *Arthritis Care Res* 65:1490–1503
- Kim SY, Servi A, Polinski JM et al (2011) Validation of rheumatoid arthritis diagnoses in health care utilization data. *Arthritis Res Ther* 13:R32
- Hanly JG, Thompson K, Skedgel C (2015) The use of administrative health care databases to identify patients with rheumatoid arthritis. *Rheumatol Res Rev* 7:69–75
- MacLean C, Louie R, Leake B et al (2000) Quality of care for patients with rheumatoid arthritis. *JAMA* 284:984–992
- Kim SY, Servi A, Polinski JM et al (2011) Validation of rheumatoid arthritis diagnoses in health care utilization data. *Arthritis Res Ther* 13(1):R32
- Lacaille D, Anis AH, Guh DP et al (2005) Gaps in care for rheumatoid arthritis: a population study. *Arthritis Rheum* 53(2):241–248
- Shipton D, Glazier RH, Guan J et al (2004) Effects of specialty services on disease-modifying antirheumatic drug use in the treatment of rheumatoid arthritis in an insured elderly population. *Med Care* 42(9):907–913
- Widdifield J, Bombardier C, Bernatsky S et al (2014) An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. *BMC Musculoskelet Disord* 15:216
- Myasoedova E, Crowson CS, Kremers HM et al (2010) Is the incidence of rheumatoid arthritis rising? Results from Olmsted County, Minnesota: 1955–2007. *Arthritis Rheum* 62:1576–1582
- Sokka T, Kautianinen H, Pincus T et al (2010) Work disability remains a major problem in rheumatoid arthritis in the 2000s: data from 32 countries in the QUEST-RA study. *Arthritis Res Ther* 12:R42
- Allaire S, Wolfe F, Niu J et al (2008) Contemporary prevalence and incidence of work disability associated with rheumatoid arthritis in the US. *Arthritis Rheum* 59:474–480