



Published in final edited form as:

Arthritis Care Res (Hoboken). 2012 October ; 64(10): 1490–1496. doi:10.1002/acr.21736.

Identification of Rheumatoid Arthritis Patients Using an Administrative Database: A Veterans Affairs Study

Bernard Ng, MBBS MMed^{1,2},

Staff rheumatologist and health services researcher at the Michael E. DeBakey VA Medical Center Health Services Research and Development Center of Excellence, Houston, Texas

Fawad Aslam, MBBS²,

Resident, Department of Medicine, Baylor College of Medicine, Houston, Texas

Nancy J. Petersen, PhD^{1,2},

Biostatistician; Michael E. DeBakey VA Medical Center Health Services Research and Development Center of Excellence, Houston, Texas

Hong-Jen Yu, MS¹, and

Statistical Analyst; Michael E. DeBakey VA Medical Center Health Services Research and Development Center of Excellence, Houston, Texas

Maria E. Suarez-Almazor, MD PhD³

Chief, Rheumatology Section, Division of General Internal Medicine, University of Texas M.D. Anderson Cancer Center, Houston, Texas

Abstract

The accuracy of the diagnosis is vital when administrative databases are used for pharmacoepidemiologic and outcome studies. Data pertaining to the utility of databases for rheumatoid arthritis (RA) is sparse and variable. We assessed the utility of various diagnostic algorithms to identify RA patients within the Veterans Health Administration (VHA) databases.

Using the ICD code for RA at two visits, at least six months apart, we identified 1779 patients between 10/1/1998 and 9/30/2009 in our local Veteran Affairs Medical Center administrative database. Disease Modifying Anti Rheumatic Drugs (DMARD) use was ascertained from the pharmacy database. Cases were analyzed based on DMARD therapy and RA codes at clinic visits. 543 patients' medical records selected by stratification and random selection on the basis of their visits were reviewed to ascertain the clinicians' diagnoses and clinical criteria documentation. Positive predictive values (PPV) were calculated for various database case identification algorithms using diagnosis of RA by medical record review as the gold standard.

Corresponding/Reprint Author: Dr. Bernard Ng, Assistant Professor, Section of Allergy, Immunology and Rheumatology, MCL, Michael E. DeBakey Veteran Affairs Medical Center, 2002 Holcombe Blvd Mail Stop 152, Houston, Texas, 77030. Ph: 713-794-8651; fax: 713-748-7359; Bernard.Ng@va.gov.

¹Michael E. DeBakey VA Medical Center Health Services Research and Development Center of Excellence, Houston, Texas.

²Baylor College of Medicine, Houston, Texas.

³University of Texas MD Anderson Cancer Center, Houston, Texas.

Conflicts of interest and support: Dr. Bernard Ng holds a South Central VA Health Care Network Research Pilot Grant Award. Dr. Suarez-Almazor holds a K24 award from the National Institute for Arthritis, Musculoskeletal and Skin Disorders (NIAMS; K24 AR53593). The rest of the authors have no conflicts of interest to declare.

Disclaimer: The views expressed are those of the authors and do not necessarily represent those of the Department of Veterans Affairs/Baylor College of Medicine.

PPV for identification of RA with two RA codes six months apart was 30.9%. Addition of DMARD therapy increased the PPV to 60.4%. The PPV further increased to 91.4% when having RA code at the last VAMC rheumatology clinic visit criterion was added.

An algorithm using only two administrative RA codes six months apart had a low PPV for correctly identifying patients with RA in the VHA database. Including DMARD therapy and requiring a RA code at the last visit with a rheumatologist increased the performance of the data extraction algorithm.

MeSH Keywords

Rheumatoid Arthritis; Health Services Research; Computerized records; Databases; ICD codes

Introduction

Computerized administrative databases are frequently used for epidemiological research. The Veterans Health Administration (VHA) databases, amongst the largest in the country, are one such example. International Classification of Disease, 9th Revision, Clinical Modification (herein referred to as ICD) codes are often used to identify subjects for research purposes. While the strategy promises to be cost-effective and less labor intensive than extracting information from individual patient records, the validity and reliability of using administrative codes as the sole source of patient identification have been debated. A few studies have been undertaken for rheumatic diseases including rheumatoid arthritis (RA), gout and spondyloarthropathies.¹⁻³ While the coding accuracy has been good for the latter, the same cannot be said for RA and gout. For RA in particular, the results have been mixed. A study by Singh et al originating from the Minneapolis Veteran Affairs Medical Center (VAMC) included patients seen at their rheumatology clinics, and used two diagnoses of RA at least 6 weeks apart and verified by two rheumatologists as the gold standard.³ They randomly selected and screened 184 patients who had visited their rheumatology clinic (71 diagnosed with RA with the gold standard criterion) and reported that a single time entry of an ICD code for RA (714) in the administrative database had a sensitivity of 100% but a specificity of only 55% and a positive predictive value (PPV) of about 66% for identifying RA patients. The addition of a positive rheumatoid factor (RF) and/or disease modifying anti-rheumatic drugs (DMARD) therapy improved specificity to 83-97% and PPV to 81-97% while decreasing the sensitivity to 76-88%. They authors however, did not include patients with a RA code who were not seen by a rheumatologist, and could have missed patients from rural areas, or those who may elect to see a rheumatologist outside the VHA, but can potentially be captured through visits to other providers. Many RA database studies rely on data not confined only to rheumatology clinic visits, thus is important for us to verify if the results from Singh et al's paper can be extrapolated to non-rheumatology clinic visit scenarios. Their suggestion was to use the administrative database as a screening tool to identify patients, followed by a chart review to increase accuracy.

A recent study identified RA patients using the Medicare claims database. Medical records were then reviewed by rheumatologists to ascertain the accuracy of the claims.⁴ They found that the PPV of two and three ICD codes for RA were 55.7 and 65.5% respectively, increasing to 66.7% when two claims were made by a rheumatologist. The addition of a DMARD prescription increased the PPV to 86.2-88.9. A prior 1997 Medicare part B study showed sensitivity and PPV of 90 and 95% respectively for RA when all claims were documented for rheumatologist visits.⁵ Another Medicare study, looked at the underlying diagnosis for patients with various rheumatic diseases undergoing hip replacement surgery.⁶ Their conclusion was that chart review is necessary to improve accuracy and validity.

Given the increasing use of administrative databases, such as those in the VHA, for comparative effectiveness research, and the surrounding controversies about the appropriateness of ICD codes for case identification of patients with RA, we examined the accuracy of various algorithms to identify RA in VHA databases, including not only patients seen by rheumatologists, but also those assigned with a RA code by other providers. The final objective of our study was to determine if robust algorithms can be proposed that would allow for robust identification of patients in administrative databases alone without requiring chart validation, as proposed by Singh.³

Materials and Methods

The study was approved by the Baylor College of Medicine Institutional Review Board and Michael E. DeBakey VAMC Research and Development committee.

Patient population

Using the VHA administrative databases, we identified patients with an ICD code for RA (714) between 10/1/1998 to 9/30/2009 and narrowed our search to patients treated at the Michael E. DeBakey VAMC in Houston. All patients identified as having two or more RA diagnostic codes at least 6 months apart met our inclusion criteria (n=1779). We used the 6 months criterion, instead of 6 weeks as used by Singh et al.,³ to increase the probability of a claim by a physician. This is because in the VHA system, it is common to have non-physician visits such as physical therapy, nutrition education, pharmacy counseling occurring close to the date of a physician visit – typically within 3 months. Such non-clinician visits often use the same diagnostic code entered at the most recent physician visit. At VAMCs clinic visits can be assigned up to 10 diagnostic codes by the clinician(s) at that particular patient encounter. We further examined which patients with RA codes never had a VAMC rheumatology clinic visit and those who had at least one VAMC rheumatology clinic visit.

We obtained DMARD prescription information from the VHA pharmacy benefits management database including methotrexate, gold, hydroxychloroquine, cyclosporine, sulfasalazine, azathioprine, minocycline, penicillamine, leflunomide, anakinra, adalimumab, etanercept, and infliximab.

Diagnostic algorithms

We evaluated the positive predictive value (PPV) of various diagnostic algorithms using validation from medical records. Within each category, PPV is defined as ratio of the number of RA diagnosed by medical record review to the total number tested positive for RA by the algorithm for that category (Figure 1). The algorithms tested were based on scenarios combining the following variables: (i) presence of at least two RA ICD codes for RA (714), at least 6 months apart; (ii) use of DMARDs for at least 180 days; and (iii) ICD claim for RA by a rheumatologist

Review of medical records for diagnostic validation

To validate the diagnosis of RA, we conducted a review of medical records in a subsample of these patients (n) selected at random for each classification group of interest described below. We stratified patients into (i) those who had received a DMARD for at least 180 days, and (ii) those who had not. Within each of these groups patients were further categorized with respect to whether they had seen a rheumatologist, and if a ICD 714 diagnostic code had been assigned to the claim (Table 1). Validation strata are described below. Rationale for the sample size per group (n) selected at random is provided in the statistical analysis section.

543 of the 550 selected for review were available, and were examined by two of the authors (BN and FA). When the charts were reviewed, the clinical diagnoses made at the last available outpatient visit superseded any earlier ones to allow for clinical diagnostic processes, under the assumption that later visits are expected to have a more updated and correct diagnosis. Our gold standard was the medical record review. All our patients had at least 2 ICD RA diagnosis codes 6 months and were thought to have RA when in addition at least one of the following criteria was met: (a) documentation of at least 4 1987 American College of Rheumatology classification criteria for RA⁷ (b) positive testing for anti-cyclic citrullinated peptide (CCP) antibodies, or (c) patient self-report of being managed by a non-VA rheumatologist for RA. The latter criterion was included because there were very few medical records for patients who had only seen PCPs fulfilling criteria (a) or (b). We felt that PCPs would not document clinical criteria for RA if they knew patients were seeing a rheumatologist elsewhere. We did not include positive rheumatoid factor (RF) as a diagnostic criterion on its own because of its lower specificity for RA, compared to anti-CCP.⁸

Laboratory and radiological data pertaining to RF, erythrocyte sedimentation rate (ESR), C reactive protein (CRP), anti-CCP and presence of joint erosions on plain radiographs were also recorded. When there was disagreement on the interpretation of clinical criteria it was resolved by a combined review of the record by the 2 reviewers and consensus.

Statistical Analysis and Sample Size Estimates

Our analysis was based on the calculation of PPV for each of the algorithms of interest. Sample size estimates were based on data from Singh et al.³ Randomization and sample size calculations were based on methods and data from Singh et al.⁹ To compare RA and non-RA patients in those with RA codes without DMARDS, using a rate of 0.325 in the RA patients and a rate of 0.675 in non-RA patients, $\alpha=0.05$, $\text{power}=0.90$, a sample size of 47 per group was estimated. The number of patients needed to detect RA versus non-RA in those with RA codes and DMARDS was 15 per group using a rate of 0.811 for RA and 0.189 for non-RA patients, $\alpha=0.05$ and $\text{power}=0.90$. To compare sensitivity and specificity of RA diagnosis as influenced by presence or absence of rheumatology clinic visits, a moderate effect size (able to detect about 6% of variance) was assumed because no prior data were available to calculate sample size. Assuming a moderate effect size 64 subjects in each group was needed. On the basis of this estimate we planned to select a random sample of 100 subjects per group. However only 50 patients were available in 1 group and an additional 6 records could not be found for review.

Other studies have validated random samples of patients without stratification, but we had concerns that without increased sampling in some of the groups (e.g. no visits to rheumatologist and no DMARD therapy) the data would not be robust. Therefore, after review and diagnostic validation within the various strata, PPV, sensitivity and specificity of each group were calculated. This analysis assumed then that the population under study was comprised of all individuals with at least 2 visits to the VAMC 6 months apart coded as RA.

Results

Our initial search identified 1779 total subjects with two RA claim codes 6 months apart. Of these 579 (32.5%) had never seen a VAMC rheumatologist; 988 (56.1%) had not received DMARD for at least 180 days. We randomly selected a validation sample of 543 patients (see methods) divided into 6 groups. The mean age of the patients was 62 years with 91% males. With regards to the medical record review of ACR criteria for RA, there was high agreement between the two reviewers (Kappa 0.97, 95% confidence interval: 0.94-1.0). Figure 1 shows the results of case ascertainment for the stratified random sample. The

number in brackets represents the percentage of patients with RA according to our gold standard, and therefore the PPV for each group.

Patients who received a DMARD for at least 180 days

Of the patients who had received DMARD but had never been seen by a VAMC rheumatologist, 71/95 (PPV=74.7%) fulfilled our RA criteria (see Figure 2). Of these 71 patients, 68 were diagnosed with RA and prescribed DMARD by private rheumatologists. Of those who were assigned a RA code at a rheumatology visit, 75/100 (PPV=75.0%) fulfilled our RA diagnosis criteria by chart review. The remaining patients in this group had other conditions which required DMARD. In the group that had never been assigned the RA code at a rheumatology visit but were on DMARD, only 2/50 (PPV=4.0%) met our criteria for RA. The remaining had psoriatic arthritis, systemic lupus erythematosus (SLE), or other diseases requiring DMARD.

Patients who had not received a DMARD for at least 180 days

Of the patients who had never seen a VAMC rheumatologist and were not receiving DMARD, 15/100 (PPV=15.0%) fulfilled our criteria for RA (see Figure 2). Of those who had a RA code for at least one rheumatology visit, 5/99 (PPV=5.1%) fulfilled our RA diagnosis. Of those without a RA code for a rheumatology visits, none fulfilled our RA diagnosis criteria.

In the patient groups which had never seen a rheumatologist, the overwhelming diagnosis was osteoarthritis (83.0%) as identified on chart review. For those that had at least one rheumatology visit but did not fulfill our RA criteria in their records, the major diagnosis was again osteoarthritis (53.0%).

Last rheumatology visit analysis

Further analysis of those visiting the rheumatology clinic was done based on whether the code for RA was assigned at the last recorded VAMC rheumatology visit or not. If the RA code was in the last documented VAMC rheumatology visit in the group who were on DMARD, 64/70 (PPV=91.4%) fulfilled our RA diagnosis criteria. The remaining 7 patients had conditions that required DMARD such as ankylosing spondylitis and psoriatic arthritis.

Positive predictive value, sensitivity and specificity calculations

Using the population of interest which comprised of individuals having 2 or more visits with RA codes at least 6 months apart, we calculated PPV, sensitivity and specificity of the each group. Table 2 shows the results for all patients combined, as well as for the subgroup who received DMARD for at least 6 months. For all patients combined (with or without DMARD), having 2 RA codes with at least one by a rheumatologist gave a sensitivity of 47.6% with a specificity of 68.3%. Using an algorithm with a last rheumatologist visit having a RA code increased the specificity to 91.5%, but decreased sensitivity.

Specificity but not sensitivity increased when the requirement of DMARD therapy was added to the algorithm. Two RA codes with at least one assigned by a rheumatologist gave a sensitivity of 44.6% and a specificity of 93.3%. By requiring that the last visit by a rheumatologist be coded as RA, the specificity increased to 98.4% (PPV 91.4%), but over 60% of cases would be missed (sensitivity 38.1%). Using 2 RA codes by any provider substantially increased detection of cases (sensitivity 88.1%), but lowered the specificity of the results to 74.1%.

Discussion

The identification of patients with RA in administrative databases is contingent upon an accurate clinical diagnosis. This is a major limitation for secondary data analyses which rely heavily on these codes to define study cohorts. Comparative effectiveness research relies frequently on the use of administrative databases. Hence, the reliability of diagnostic codes in these studies is paramount to ensure the validity of study findings. A few studies have examined how to increase the PPV of ICD codes for RA by adding various criteria such as concomitant use of DMARD and positive RF.^{3,4} However, these studies have proposed diagnostic algorithms that only include patients seen by rheumatologists and may have missed many patients without documented rheumatology visits who truly have RA. We conducted a validation study of ICD RA codes in the VHA system including all patients with at least 2 RA codes, independently of whether they were seen by a rheumatologist. To our knowledge this is the only study validating this approach and it is also the largest validation study within the VHA system. In addition, we also examined the accuracy of rheumatology RA codes at any visit versus RA codes at the last visit, allowing for potential changes in disease diagnosis by a specialist. This can be relevant because at earlier visits a tentative diagnosis of RA may be made, resulting in an administrative RA code being assigned to those visits but later confirmed or rejected as additional clinical data becomes available. Thus the code assigned at the later visit is expected to be more accurate.

Our results are in agreement with two other studies reporting that adding a DMARD increases PPV^{3,4}. However, our PPV for patients not receiving DMARD are lower than those reported in the literature which have ranges from 56 to 66.^{3,4} This could possibly be attributed to the increasing and earlier use of DMARD in RA patients. Our values which are lower than that of Singh et al may be explained by the fact that we also included PCP diagnosed patients in our analysis, unlike Singh et al, and from those of Kim et al possibly because theirs was a MEDICAID database with different coding practices.

The highest PPV (91.4%) was found in the group who had received DMARD for at least 180 days, and had a RA code at the last rheumatology visit. This likely reflects that as the diagnostic process continues based on multiple visits and testing, and the coding also follows. This algorithm also resulted in the highest specificity (98.4%) but at a loss of sensitivity, as more than 60% of the patients with RA could have been missed using this case definition. When using an algorithm with 2 RA codes at least one by a rheumatologist at any time in patients receiving DMARD the results showed a PPV of 75% with sensitivity of 44.6% and specificity of 93.3%. Using the RA diagnostic codes alone without including the requirement of DMARD therapy yielded inadequate results which could result in the inclusion of false positive and false negative cases and misclassification. This approach should therefore be used with caution for analytical studies. In our stratified sample 20 (6.7%) of the 298 patients who were not on DMARD had RA according to our gold standard criteria. Some of these patients may have received DMARD outside the VA system, may have had contraindications to one or more of these drugs or possibly have milder disease. A diagnostic algorithm including therapy with DMARD will miss these patients but this bias appears less important than the misclassification that would occur by including patients not having received therapy.

It is interesting that in the group who received DMARD who had not been assigned a RA code at any rheumatology visit, only 2/50 (4.0%) were diagnosed as having RA by our criteria (see Figure 1). For those who had not received DMARD and did not have a RA code by a rheumatologist, none had RA by our gold standard criteria and yet were identified as having RA by other providers. It was our anecdotal observation during the review of medical records that the past medical history section is often not updated and may continue

to carry a diagnosis of RA without further mention in the visit record about RA in the assessment and plan.

In the group who had never seen a VAMC rheumatologist and were on DMARD, the majority (71.6%) were patients seeing non VAMC rheumatologists but getting their DMARD prescription filled at the VAMC through the PCP perhaps because of lower costs. When patients were not on DMARD and were only being managed by the PCPs, only 15% of those with a RA code actually had RA based on our criteria. In these 15 patients, 10 were managed by non-VAMC rheumatologists (they may be receiving DMARD from non-VA sources) but 5 patients who were managed by their VAMC PCP with steroids or NSAIDs alone. Herein is a serious issue either with the diagnosis or management as DMARD should ideally be introduced as early as possible. Nevertheless, it does reflect either a problem in diagnosis or the appropriate management of RA.

Our results suggest that PCPs may require better education on the diagnosis and management of RA. Of the patients referred to rheumatology, with a RA code 53% had osteoarthritis. While we did not undertake an analysis to see the concordance between the referring diagnosis and the actual diagnosis that was made at the rheumatology visit, we can safely estimate that it was low. This agrees with the existing evidence that diagnostic agreement between PCPs and rheumatologists is generally low.⁹ Indeed codes are more accurately placed when the disease is well known and has a clear definition and is also a function of the clinicians' experience.¹⁰ More education and training on differentiating between RA and other conditions and appropriate workup may be required. Also more referrals to rheumatology may be warranted for management and diagnosis of conditions like inflammatory arthropathy, spondyloarthropathies, psoriatic arthritis and SLE.

Our study is not without limitations. The data only applies to the VAMC population which is overwhelmingly male and with multiple co-morbid conditions. The latter may explain the lack of documentation of RA by the PCPs as their notes may be taken up by other more pressing and acute matters in a limited encounter. This problem of poor documentation may also be more pronounced in the VAMC database where both clerical staff and physician can enter these codes, the coding is not directly claims-based and thus lacks incentive for proper documentation, and is devoid of inherent cross-check mechanisms.³ Using documentation of outside rheumatologists prescribing DMARD as one of our RA criteria for PCP-visits only subjects could have resulted in selection bias. However, only 22 of the 195 patients who had only seen the PCP would fulfill our RA diagnosis, if we had used only ACR RA criteria and/or positive anti-CCP test to make RA diagnosis with PCP clinical notes. The reliance of RA identification based solely on administrative coding will depend on the specific research questions being asked and can be improved by addition of a rheumatologist diagnosis and/or DMARD prescription. Some studies may require an accompanying chart review to confirm the diagnosis as suggested by others.^{3,4,6} However, our results show that an algorithm including two RA codes, at least one by a rheumatologist, and therapy with DMARD yields adequate PPV, sensitivity and specificity for population-based studies examining for instance trends (e.g mortality, therapeutic patterns) in longer term cohort studies. For some studies, particularly analytical research aiming to establish associations the specificity and PPV of the sample need to be higher. In this case, adding as a criterion a RA code at the last rheumatology visit substantially improves performance. The use of RA codes alone without additional criteria could strongly compromise the validity of RA studies based on administrative databases. Wrong patient selection based on incorrect coding has previously led to misleading conclusions on disease associations in RA.¹¹

Our study has ascertained the validity of various diagnostic algorithms to identify cases with RA using administrative databases. While these results may not be generalizable to settings

other than the VHA, they provide alternatives for identification that could be used for different research objectives. By adding pharmacy claims to other administrative databases, the accuracy of RA case identification can be greatly improved. Furthermore, we have identified algorithms that have excellent PPV and specificity and could allow for the use of these databases in analytical research. Nevertheless, it is clear that for many research questions the accuracy of the diagnosis as ascertained by these may not be sufficient and review of medical records might be required. Our results also suggest that training of providers and clinic staff in coding practices could improve the utility of administrative databases. Finally, our findings also point to the need for enhanced education for PCPs on the diagnosis and management of RA.

Acknowledgments

The research reported here was supported by the Department of Veteran Affairs, Veterans Health Administration, Health Services Research and Development Service, (VA.SCV.1010./000-00.B_N). This work was also partly supported by the Veterans Affairs Health Services Research and Development Service Houston Center of Excellence (HFP90-020). Dr. Suarez-Almazor has a K24 career award from the National Institute for Arthritis, Musculoskeletal and Skin Diseases.

References

1. Malik A, Dinnella JE, Kwok CK, Schumacher HR. Poor validation of medical record ICD-9 diagnoses of gout in a veterans affairs database. *J Rheumatol.* 2009; 36:1283–6. [PubMed: 19447931]
2. Singh JA, Holmgren AR, Krug H, Noorbaloochi S. Accuracy of the diagnoses of spondylarthritides in veterans affairs medical center databases. *Arthritis Rheum.* 2007; 57:648–55. [PubMed: 17471541]
3. Singh JA, Holmgren AR, Noorbaloochi S. Accuracy of Veterans Administration databases for a diagnosis of rheumatoid arthritis. *Arthritis Rheum.* 2004; 51:952–7. [PubMed: 15593102]
4. Kim SY, Servi A, Polinski JM, et al. Validation of rheumatoid arthritis diagnoses in health care utilization data. *Arthritis Res Ther.* 2011; 13:R32. [PubMed: 21345216]
5. Katz JN, Barrett J, Liang MH, et al. Sensitivity and positive predictive value of Medicare Part B physician claims for rheumatologic diagnoses and procedures. *Arthritis Rheum.* 1997; 40:1594–600. [PubMed: 9324013]
6. Losina E, Barrett J, Baron JA, Katz JN. Accuracy of Medicare claims data for rheumatologic diagnoses in total hip replacement recipients. *J Clin Epidemiol.* 2003; 56:515–9. [PubMed: 12873645]
7. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* 1988; 31:315–24. [PubMed: 3358796]
8. Lee DM, Schur PH. Clinical utility of the anti-CCP assay in patients with rheumatic diseases. *Ann Rheum Dis.* 2003; 62:870–4. [PubMed: 12922961]
9. Gamez-Nava JI, Gonzalez-Lopez L, Davis P, Suarez-Almazor ME. Referral and diagnosis of common rheumatic diseases by primary care physicians. *Br J Rheumatol.* 1998; 37:1215–9. [PubMed: 9851272]
10. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. *Health Serv Res.* 2005; 40:1620–39. [PubMed: 16178999]
11. Tennis P, Bombardier C, Malcolm E, Downey W. Validity of rheumatoid arthritis diagnoses listed in the Saskatchewan Hospital Separations Database. *J Clin Epidemiol.* 1993; 46:675–83. [PubMed: 8326356]

Significance and Innovation

1. Use of administrative database is important to address long term outcomes in chronic diseases like rheumatoid arthritis (RA)
2. Having robust data extraction algorithms with high positive predictive value is vital for conducting retrospective observational studies using administrative databases
3. Previous studies have looked only at rheumatology clinic visits but did not include primary care visits which accounts for a large part of patients' clinical records.
4. Having 2 RA diagnosis codes at any clinic visits is not sufficient. It is important that the RA diagnosis code was entered at the most recent rheumatology clinic visit. This criterion together with prescription of disease modifying anti-rheumatic drugs results provides the highest positive predictive value for RA data extraction.

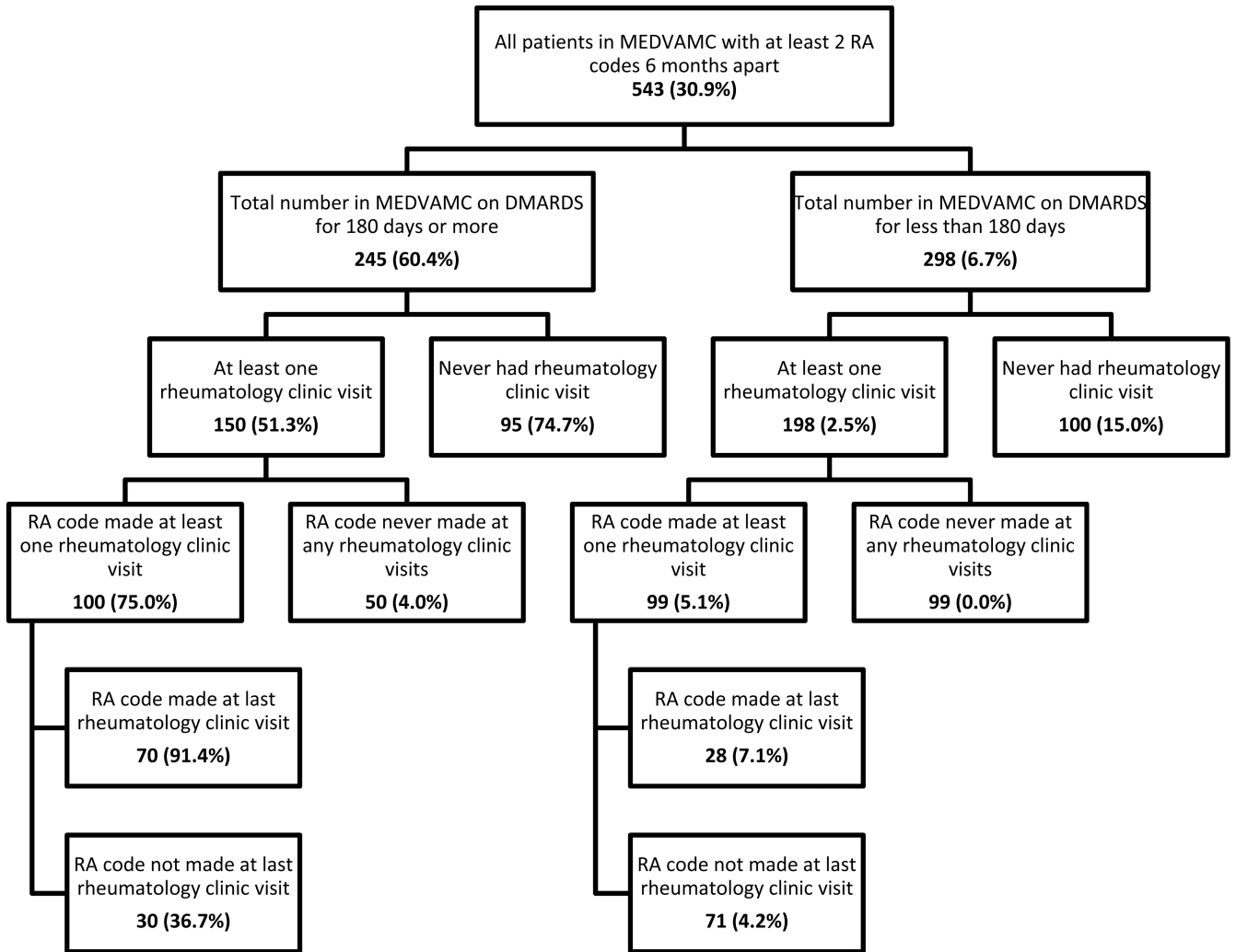


Figure 1. Sample of patients receiving care at the local VA medical center who have at least 2 RA codes at a minimum of 6 months apart. The number in each box represents those tested positive for RA by the individual algorithms. The percentages of patients who fulfilled RA diagnosis by medical record review for each category are found in the brackets. This number corresponds to the positive predictive value of the individual box's algorithm which is defined as ratio of the number of RA diagnosed by our gold standard (medical record review) in that box to the total number tested positive for RA by the algorithm.

Chart Diagnosis Per Provider Type

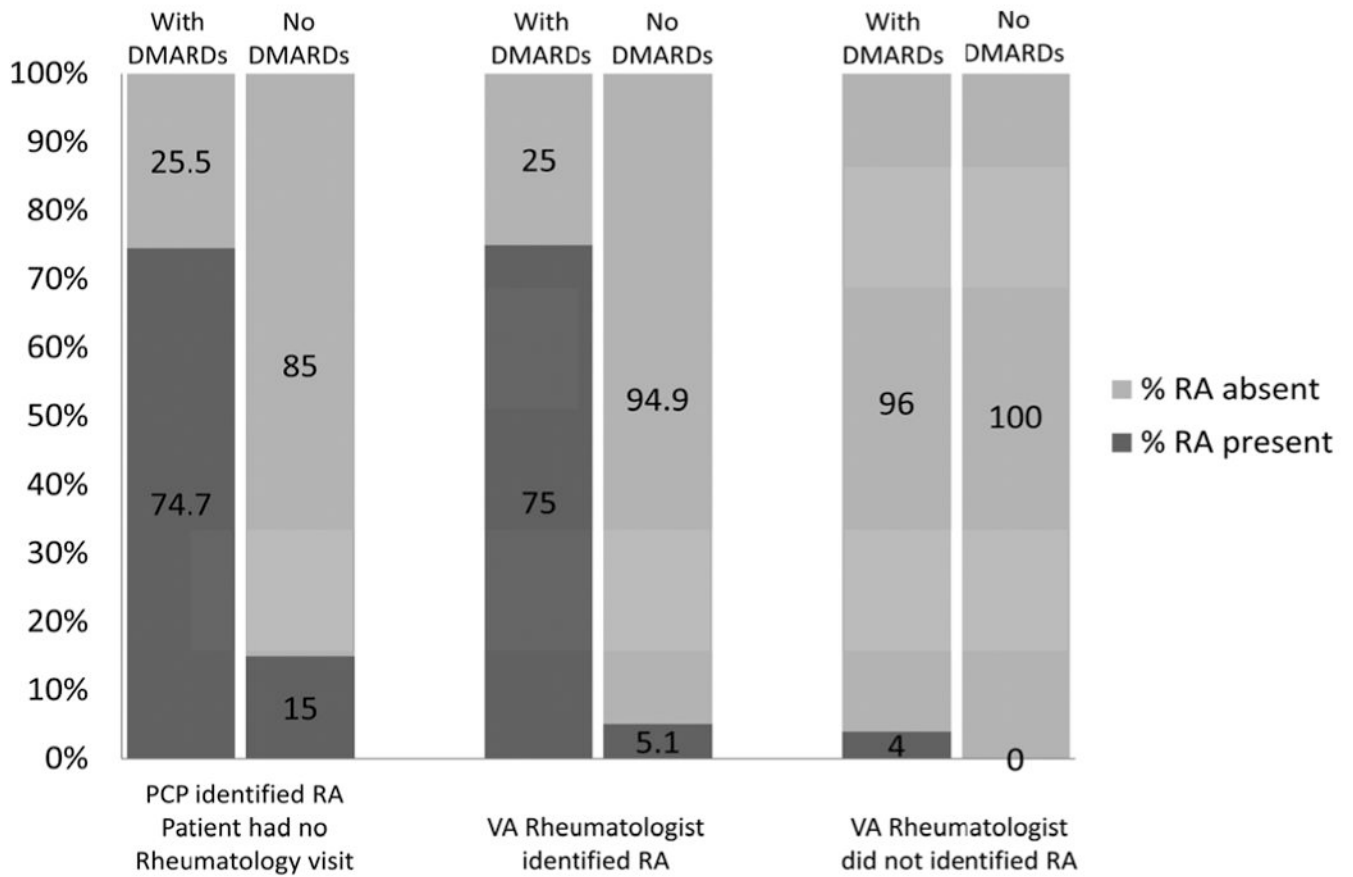


Figure 2.

RA medical record diagnosis per provide type in our sample of 543 patients. For each group, columns towards the y-axis are patients who had 180 days or more of DMARD. The numbers in the darker shaded columns represent the percentages who are diagnosed with RA by medical records review in each group.

Table 1
Sampling of patients receiving DMARD for 180 days or more versus less than 180 days

	CHARTS REVIEWED IN PATIENTS RECEIVING DMARD FOR	
	180 DAYS OR MORE	LESS THAN 180 DAYS
Patients with only VAMC primary care visits (n=100)	95 5 of the records could not be located, suggesting that although patients had been assigned our facility code they might have been transferred to another facility	100
Patients with one or more VAMC rheumatology clinic visits (n=200). These were further subdivided into those with:		
a) RA diagnostic code for at least one rheumatology clinic visit (n=100)	100	99 One record could not be found
b) No RA diagnosis code at any rheumatology clinic (n=100)	50 Only 50 patients were found in this category and all were included	99 One record could not be found

Table 2
Total population number and sample statistics: positive predictive value (PPV), sensitivity and specificity of different algorithms for random sample of patients with at least 2 RA codes

	N Total Population	Estimated number of patients with RA PPV from random sample × Total Population	Random Sample Statistics		
			PPV, % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
ALL PATIENTS (WITH OR WITHOUT DMARD THERAPY FOR 180 DAYS)					
2 RA codes (any visit)	1779	550	30.9 (27.7-34.2)	NA*	NA*
2 RA codes with at least one RA code in a rheumatologist visit	1014	408	40.2 (34.1-46.3)	47.6 (41.4-53.8)	68.3 (62.5-74.1)
2 RA codes with at least one rheumatology visit and with the last rheumatologist visit with a RA code	541	364	67.3 (58.9-75.8)	39.3 (30.5-48.0)	91.5 (86.5-96.5)
PATIENTS WITH DMARD THERAPY FOR 180 DAYS					
2 RA codes (any visit)	791	478	60.4 (55.3-65.5)	88.1 (84.7-91.5)	74.1 (69.6-78.7)
2 RA codes with at least one RA code in a rheumatologist visit	621	466	75.0 (67.2-82.8)	44.6 (35.7-53.6)	93.3 (88.9-97.8)
2 RA codes with at least one rheumatology visit and with the last rheumatologist visit with a RA code	430	393	91.4 (85.4-97.4)	38.1 (27.7-48.5)	98.4 (95.7-100.0)

* Sensitivity and specificity cannot be estimated for all