The Journal of Rheumatology 2021;48:685–92
doi:10.3899/jrheum.200570
First Release February 1 2021

The Journal of Rheumatology

Identifying Patients With Axial Spondyloarthritis in Large Datasets: Expanding Possibilities for Observational Research

Jessica A. Walsh[1], Shaobo Pei[1], Gopi K. Penmetsa[1], Rebecca S. Overbury[1], Daniel O. Clegg[1], and Brian C. Sauer[1]

ABSTRACT. ***Objective.*** Observational research of axial spondyloarthritis (axSpA) is limited by a lack of methods for identifying diverse axSpA phenotypes in large datasets. Algorithms were previously designed to identify a broad spectrum of patients with axSpA, including patients not identifiable with diagnosis codes. The study objective was to estimate the performance of axSpA identification methods in the general Veterans Affairs (VA) population.

Methods. A patient sample with known axSpA status (n = 300) was established with chart review. For feasibility, this sample was enriched with veterans with axSpA risk factors. Algorithm performance outcomes included sensitivities, positive predictive values (PPV), and F1 scores (an overall performance metric combining sensitivity and PPV). Performance was estimated with unweighted outcomes for the axSpA-enriched sample and inverse probability weighted (IPW) outcomes for the general VA population. These outcomes were also assessed for traditional identification methods using diagnosis codes for the ankylosing spondylitis (AS) subtype of axSpA.

Results. The mean age was 54.7 and 92% were male. Unweighted F1 scores (0.59–0.74) were higher than IPW F1 scores (0.48–0.65). The full algorithm had the best overall performance ($F1_{IPW}$ 0.65). The Early Algorithm was the most inclusive ($sensitivity_{IPW}$ 0.90, PPV_{IPW} 0.38). The traditional method using ≥ 2 AS diagnosis codes from rheumatology had the highest PPV (PPV_{IPW} 0.84, $sensitivity_{IPW}$ 0.34).

Conclusion. The axSpA identification methods demonstrated a range of performance attributes in the general VA population that may be appropriate for various types of studies. The novel identification algorithms may expand the scope of research by enabling identification of more diverse axSpA populations.

Key Indexing Terms: ankylosing spondylitis, cohort study, epidemiology, spondyloarthritis

Observational research in large populations is important for studying uncommon outcomes and diseases in real-world settings[1]. In the field of axial spondyloarthritis (axSpA), there is a dearth of big data research due to challenges with identifying axSpA patients in large datasets. With advancements in imaging and treatment, it has become apparent that many patients with axSpA were previously unrecognized with traditional concepts of disease[2,3]. Despite our broader understanding of axSpA, big data axSpA research continues to be constrained by outdated axSpA definitions, since International Classification of Diseases, 9th and 10th revisions (ICD-9 and ICD-10, respectively) billing codes exist only for a single axSpA phenotype: ankylosing spondylitis (AS)[4,5,6,7,8]. As such, around half of the approximately 3.3 million Americans with axSpA have been excluded from big data axSpA research[9,10], and data are sparse with important outcomes such as mortality, comorbidities, and treatment patterns in real-world axSpA populations[8,11].

Even within the narrower AS disease spectrum, traditional methods for identifying patients are limited. AS research in large datasets typically relies on AS ICD codes[12,13,14,15]. However, AS ICD codes may not perform well for case identification in the United States[16]. Other methods may improve accuracy, such as supplementing AS ICD codes with chart review; however, studies requiring chart review have limited sample sizes since chart review is impractical for thousands of patients. Further, chart review is not possible in many large datasets since access to patient-level data is restricted. Other approaches may improve specificity, such as requiring an AS medication or a rheumatology visit in conjunction with AS diagnosis codes, but these approaches fail to capture important subsets of AS patients (i.e., untreated patients or patients receiving AS care from a primary care provider).

To address limitations with axSpA patient identification in large datasets, we developed algorithms that use structured and unstructured data to identify axSpA patients in national

This study was supported by the Marriott Daughters Foundation and Rheumatology Research Foundation.

[1]J.A. Walsh, MD, MBA, MSCI, S. Pei, PhD, R.S. Overbury, MD, B.C. Sauer, PhD, G.K. Penmetsa, MD, D.O. Clegg, MD, Salt Lake City Veterans Affairs and University of Utah Medical Centers, Department of Internal Medicine, Divisions of Rheumatology and Epidemiology, Salt Lake City, Utah, USA.

The authors declare no conflicts of interest.

Address correspondence to Dr. J.A. Walsh, Division of Rheumatology, 30 N. 1900 E. Salt Lake City, UT 84109, USA.
Email: Jessica.Walsh@hsc.utah.edu.

Accepted August 13, 2020.

Veterans Health Administration (VHA) datasets[17]. These algorithms performed well in a sample of veterans enriched with risk factors for axSpA (area under the curve [AUC] 0.86–0.96)[18]. A similar approach was applied to a Boston-based population enriched with axSpA patients, and the results were similar (AUC 0.80–0.93)[19]. However, the axSpA-enriched populations are expected to be different from general (unenriched) populations, and the ability of the algorithms to accurately identify patients with axSpA in the general population was unknown. The goal of this study was to estimate the performance of axSpA identification algorithms in the general Veteran population.

MATERIALS AND METHODS

Data sources. This study used historical data from veterans enrolled in the VHA. The data source was the Corporate Data Warehouse (CDW), a national repository of data from the VHA medical record system (VistA) and other VHA clinical and administrative systems[2]. Data were housed and analyzed within the Veterans Affairs Informatics and Computing Infrastructure (VINCI)[20]. This research was conducted in compliance with the Declaration of Helsinki, with the approval of the University of Utah Institutional Review Board (IRB_00052363).

AxSpA identification methods. Five methods for identifying axSpA in large datasets were assessed (Table 1). These included 3 novel methods developed by our team: the Full Algorithm, the High Feasibility Algorithm, and the Early Algorithm. We also assessed 2 traditional methods that have been used in axSpA epidemiologic studies including ≥ 2 AS ICD codes ≥ 7 days apart from any source (AS codes, any specialty) and ≥ 2 AS ICD codes ≥ 7 days apart from a rheumatology encounter (AS codes, rheumatology). Details about algorithm development were previously published[18,21]. In brief, the Full Algorithm is the most comprehensive, with 3 natural language processing (NLP) models[22] and 46 coded variables. The High Feasibility Algorithm included the 16 top-ranked coded variables. The Early Algorithm is similar to the Full Algorithm, except SpA ICD codes and rheumatology visits were not included, to enhance identification of earlier disease or less

Table 1. Methods for identifying axial SpA.

Method	No. Unique Variables	Variables
Full algorithm[a]	49	· NLP models: SpA, sacroiliitis, HLA-B27 · Demographics: age, geographic region, race, ethnicity, sex · Laboratory results: CCP, HLA-B27, RF · Healthcare utilization: no. rheumatology visits, no. visits with any provider, no. CRP tests, exposure to ≥ 1 biologic DMARD, no. ESR tests, duration in VHA system, exposure to ≥ 1 nonbiologic DMARD, exposure to allopurinol or colchicine · ICD codes[b]: AS, unspecified inflammatory SpA, osteoarthrosis and allied disorders, psoriasis, arthropathy associated with Reiter & nonspecific urethritis, sciatica, low back pain, backache unspecified, osteoarthrosis of the spine (includes DISH), intervertebral disk disorders, RA, sacroiliitis not otherwise classified, cervicalgia, other specified inflammatory SpA, UC, spinal stenosis lumbar thoracic cervical, neuritis or radiculitis, acute anterior uveitis, connective tissue disease, psoriatic arthritis, gout, thoracic pain, CD, arthropathy associated with CD or UC, polymyalgia rheumatica, dorsalgias unspecified, Paget, sarcoidosis, vasculitis · Comorbidity index: Charlson Comorbidity Index[3,4]
Early algorithm[a,c]	42	· NLP models: SpA, sacroiliitis, HLA-B27 · Demographics: age, geographic region, race, ethnicity, sex · Laboratory results: CCP, HLA-B27, RF · Healthcare utilization: no. visits with any provider, no. CRP tests, exposure to ≥ 1 biologic DMARD, no. ESR tests, duration in VHA system, exposure to ≥ 1 nonbiologic DMARD, exposure to allopurinol or colchicine · ICD codes[b]: osteoarthrosis and allied disorders, psoriasis, sciatica, low back pain, backache unspecified, osteoarthrosis of the spine (includes DISH), intervertebral disk disorders, RA, sacroiliitis not otherwise classified, cervicalgia, UC, spinal stenosis lumbar thoracic cervical, neuritis or radiculitis, acute anterior uveitis, connective tissue disease, gout, thoracic pain, CD, polymyalgia rheumatica, dorsalgias unspecified, Paget, sarcoidosis, vasculitis · Comorbidity index: Charlson Comorbidity Index
High feasibility algorithm[a,d]	16	· Demographics: age, geographic region · Laboratory results: HLA-B27 · Healthcare utilization: no. rheumatology visits, no. visits with any provider, no. CRP tests, exposure to ≥ 1 biologic DMARD, no. ESR tests, duration in VHA system · ICD codes[b]: AS, unspecified inflammatory SpA, osteoarthrosis and allied disorders, low back pain, osteoarthrosis of the spine (includes DISH), cervicalgia · Comorbidity Index: Charlson Comorbidity Index
AS codes, any specialty	1	≥ 2 AS diagnosis codes ≥ 7 days apart from any specialty
AS codes, rheumatology	1	≥ 2 AS diagnosis codes ≥ 7 days apart from rheumatology

[a] Algorithms developed with random forest. [b] Number of ICD-9 and ICD-10 codes used for each ICD variable. [c] Same as full algorithm except SpA ICD codes & rheumatology visits excluded. [d] Top 16 ranked variables after excluding NLP models; variables ranked with random forest mean Gini scores. AS: ankylosing spondylitis; CCP: cyclic citrullinated peptide antibody; CD: Crohn disease; CRP: C-reactive protein; DISH: diffuse idiopathic skeletal hyperostosis; DMARD: disease-modifying antirheumatic drug; ESR: erythrocyte sedimentation rate; ICD: International Classification of Diseases; NLP: natural language processing; RA: rheumatoid arthritis; RF: rheumatoid factor; SpA: spondyloarthritis; UC: ulcerative colitis; VHA: Veterans Health Administration.

classic phenotypes. Random forest and 5-fold cross-validation were used to develop and test the algorithms[23,24,25,26].

Population. To assess performance of the identification methods, an independent sample of 300 veterans with known axSpA status (axSpA vs no axSpA) was established with chart review. Since it was not feasible to chart review the tens of thousands of patients that would have been required to randomly capture a sufficient number of axSpA patients, we enriched this chart review sample with veterans at elevated risk for axSpA.

We used the Full Algorithm to estimate axSpA risk. Random forest cutoff scores from the Full Algorithm were used to quantify axSpA risk and assign each veteran to a risk quartile (i.e., cutoff scores of 0.75–1.00 for the highest risk quartile, 0.50–0.75 for the second highest, 0.25–0.50 for the third highest, and 0.0–0.25 for the lowest). Veterans were selected from each risk quartile for chart review. Rheumatologist chart reviewers used annotation software (eHOST[27]) to review and classify the 300 sampled veterans as having or not having axSpA, according to expert opinion and our previously published axSpA chart review guidelines[18,21].

Inverse probability weighting for estimating algorithm performance in the general population. The performance estimates in axSpA-enriched populations are subject to partial verification bias[20]. Verification bias occurs when the reference standard has not been carried out in all patients because of ethical or practical reasons[28] (i.e., the algorithms were developed in an axSpA-enriched sample rather than the general Veterans Affairs [VA] population). Verification is known to lead to biased accuracy estimates. To correct for verification bias and understand how the algorithms are expected to function in the general VA population, we applied inverse probability weighting (IPW). IPW enables calculating statistics standardized to a target population (general VHA population) that is different from the population in which the data were collected (axSpA-enriched sample)[29]. General population-level statistics were calculated by applying sampling weights based on the distribution of patients within each risk quartile (Supplementary Table 1, available with the online version of this article) and bootstrapping to obtain 95% CI.

Other statistics. Means, percentages, SD, and 95% CI were used to compare the subset of randomly selected veterans to the sample selected for chart review. Algorithm performance was evaluated with sensitivity, positive predictive values (PPV), and F1 scores, with and without IPW. The F1 score is an overall measure of performance that considers both the PPV (precision) and sensitivity (recall)[30] $[F1 = 2 * \frac{PPV*sensitivity}{PPV+sensitivity}]$. F1 is useful for assessing performance in low-prevalence conditions. F1 scores were used to select the optimal cutoff scores for determining a positive vs negative outcome with the algorithms. As expected in a low-prevalence disease, specificity, negative predictive value, accuracy, and receiver-operator curve analysis were not useful for differentiating between axSpA identification methods, since the differences between algorithms with each of these outcomes was very small.

RESULTS

Population. We randomly selected 150,000 veterans participating in the national VHA system between January 1, 2007, and June 30, 2017 (Figure 1). After excluding Veterans with < 2 outpatient encounters ≥ 7 days apart, there were 79,826 remaining veterans. They were classified into quartiles for

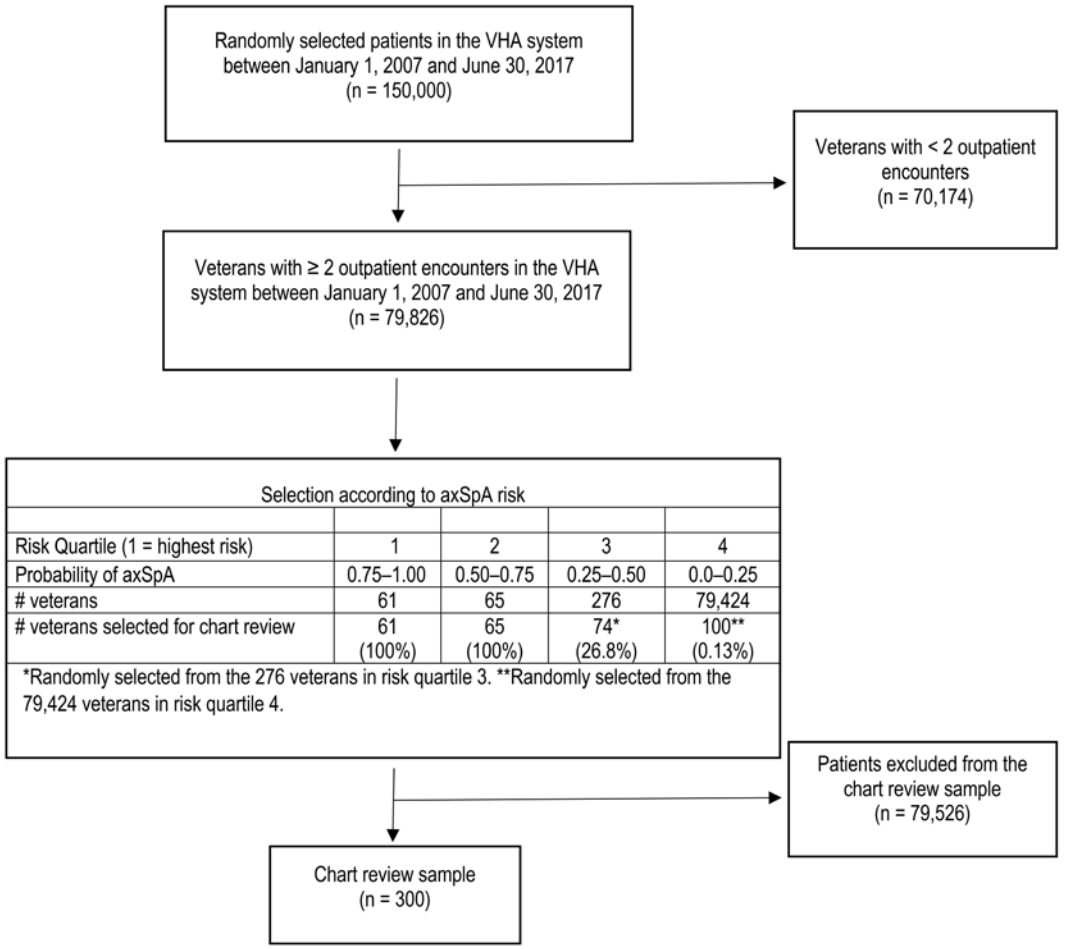


Figure 1. Selection of patient sample for chart review. axSpA: axial spondyloarthritis; VHA: Veterans Health Administration.

axSpA risk determined by probabilities calculated from the Full Algorithm. Sixty-one patients were classified in the first (highest risk) quartile, 65 in the second quartile, 276 in the third quartile, and 79,424 in the fourth quartile. All patients in the first and second quartiles were selected for chart review. From the third quartile, 74 of the 276 patients were randomly selected for chart review, and from the fourth quartile, 100 of 79,424 patients were randomly selected for chart review. In total, 300 patients were selected for chart review.

Patient features were compared between veterans from the general population (with ≥ 2 encounters in the VHA system ≥ 7 days apart) vs patients selected to the chart review study sample (Table 2). Compared to the general VA population, the chart review patients were younger (54.7 vs 58.1 yrs) and had a higher comorbidity burden (Charlson comorbidity index 2.6 vs 2.1). The percentage of veterans with a clinically available positive HLA-B27 test result was higher in the chart review sample than the general VA population (8.0% vs 0.09%). Low back pain was more common in the chart review sample than the general VA population (12.0% vs 7.1%), and exposure to disease-modifying antirheumatic drugs (DMARD) was higher in the chart review sample (nonbiologic: 21.3% vs 1.6%, biologic: 22.0% vs 0.5%). The chart review patients were enrolled in the VA for a longer duration (7.7 vs 6.5 yrs). The chart review patients had more encounters within the VA system per year than the general VA population (37.9 vs 26.9), where VA system encounters were defined as provider visits and ancillary services (pharmacy, laboratory, radiology, physical therapy, social work, nutrition, chiropractic services, etc.). Per chart review, 30.3% of veterans in the chart review patients were classified as having axSpA.

Table 2. Characteristics and axSpA classifications of the general VHA population and the chart review sample.

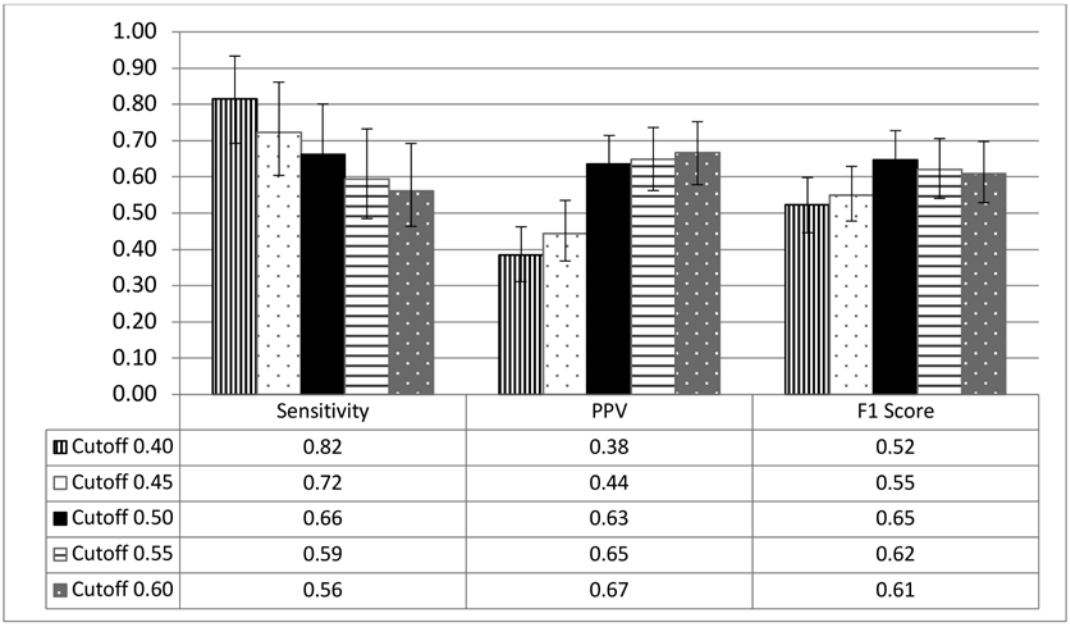
	Patients From General Population With ≥ 2 Encounters in VHA System, n = 79,826				Chart Review Sample, axSpA-enriched Population, n = 300			
	N or mean	SD or %	95% CI		N or mean	SD or %	95% CI	
Age, yrs, at the cohort entry*	58.1	17.8	58.0	58.2	54.7	17.7	52.7	56.7
Sex, male	74,359	93.2	93.0	93.3	277	92.3	88.8	94.8
Race								
White	55,830	69.9	69.6	70.3	212	70.7	65.3	75.5
Black	11,269	14.1	13.9	14.4	42	14.0	10.5	18.4
Other	4490	5.6	5.5	5.8	23	7.7	5.2	11.2
Unknown	8237	10.3	10.1	10.5	23	7.7	5.2	11.2
Ethnicity								
Non-Hispanic	67,589	84.7	84.4	84.9	262	87.3	83.1	90.6
Hispanic	4326	5.4	5.3	5.6	16	5.3	3.3	8.5
Unknown	7911	9.9	9.7	10.1	22	7.3	4.9	10.9
Geographic region at cohort entry								
Continental	13,151	16.5	16.2	16.7	49	16.3	12.6	20.9
Midwest	17,521	22.0	21.7	22.2	62	20.7	16.5	25.6
North Atlantic	19,137	24.0	23.7	24.3	82	27.3	22.6	32.6
Pacific	14,391	18.0	17.8	18.3	57	19.0	15.0	23.8
Southeast	15,626	19.6	19.3	19.9	50	16.7	12.9	21.3
Health characteristics								
Charlson Comorbidity Index*	2.1	2.8	2.1	2.2	2.6	2.9	2.3	2.9
HLA-B27–positive[a,*]	74	0.09	0.07	0.12	24	8.0	5.4	11.6
Low back pain[b,*]	2.29	7.1	2.2	2.3	5.67	12.0	4.3	7.0
Exposure to ≥ 1 nonbiologic DMARD*	1292	1.6	1.5	1.7	64	21.3	17.1	26.3
Exposure to ≥ 1 biologic DMARD*	410	0.5	0.5	0.6	66	22.0	17.7	27.0
VA system utilization								
Duration in the VA system*	6.5	3.5	6.5	6.5	7.7	3.1	7.3	8.0
No. encounters per year, mean[c,*]	26.9	44.9	26.6	27.2	37.9	38.4	33.6	42.3
Classified as axSpA								
Chart review	NA	NA	NA	NA	91	30.3	25.4	35.8
Full algorithm*	126	0.16	0.13	0.19	126	42.0	36.6	47.7
High feasibility algorithm*	140	0.18	0.15	0.21	122	40.7	35.3	46.3
Early algorithm*	409	0.51	0.47	0.56	158	52.7	47.0	58.3
≥ 2 AS codes, any specialty*	105	0.13	0.11	0.16	105	35.0	29.8	40.6
≥ 2 AS codes, rheumatology*	49	0.06	0.05	0.08	49	16.3	12.6	20.9

*P value < 0.05. [a] Among veterans with clinically available HLA-B27 test results. [b] Defined as ≥ 1 low back pain ICD code (724.2x, M54.5x). [c] Encounters included outpatient provider visits and ancillary services (pharmacy, laboratory, radiology, physical therapy, social work, nutrition, chiropractic services, etc.). AS: ankylosing spondylitis; axSpA: axial spondyloarthritis; DMARD: disease-modifying antirheumatic drug; ICD: International Classification of Diseases; NA: not applicable; VA: Veterans Affairs; VHA: Veterans Health Administration.

According to the 5 axSpA identification methods, 0.06–0.51% were classified as having axSpA in the general VA population vs 16.3–52.7% in the chart review sample.

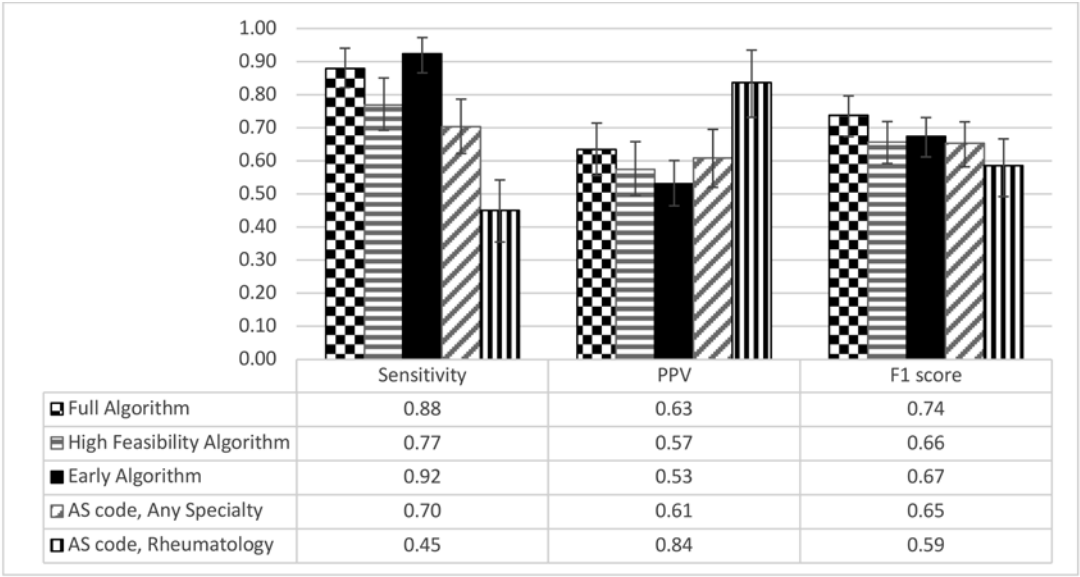
Performance of identification methods. For the Full Algorithm, Early Algorithm, and High Feasibility Algorithm, random forest cutoff scores (for classifying patients as having or not having axSpA) were evaluated with sensitivity, PPV, and F1 scores (exemplified with the Full Algorithm in Figure 2). We selected optimal cutoff scores of 0.50 for all 3 algorithms. These scores were applied to subsequent analyses.

The unweighted sensitivities of the 5 identification methods ranged from 0.45 to 0.92 (Figure 3). The unweighted PPV ranged from 0.53 to 0.84, and the unweighted F1 scores ranged from 0.59 to 0.74. With IPW, the weights assigned to each risk quartile were proportional to the number of patients from the general veteran population in each risk quartile (first quartile 1.0, second quartile 1.0, third quartile 3.7, fourth quartile 794.2; Supplementary Table 1, available with the online version of this article). After applying IPW, the sensitivities of the 5 identification methods ranged from 0.34 to 0.90 (Figure 4). The Early Algorithm had statistically higher sensitivity (nonoverlapping 95% CI) compared to the High Feasibility Algorithm, AS codes



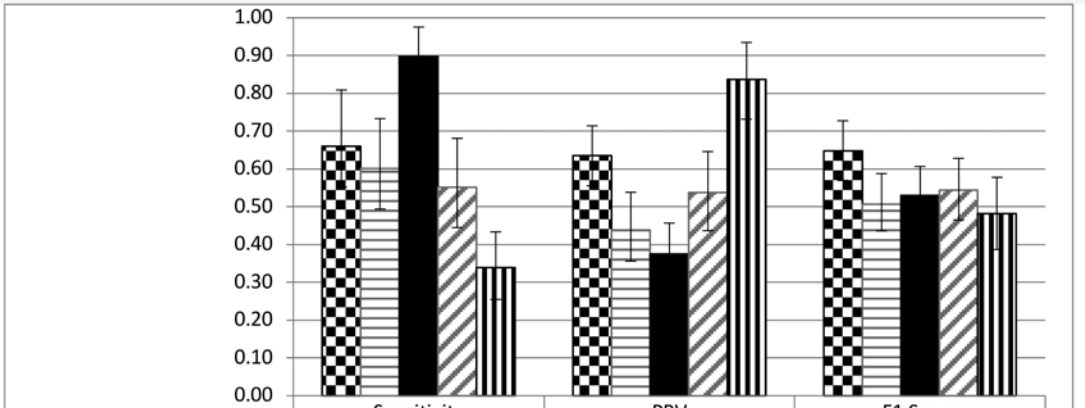
	Sensitivity	PPV	F1 Score
Cutoff 0.40	0.82	0.38	0.52
Cutoff 0.45	0.72	0.44	0.55
Cutoff 0.50	0.66	0.63	0.65
Cutoff 0.55	0.59	0.65	0.62
Cutoff 0.60	0.56	0.67	0.61

Figure 2. Random forest cutoff scores for full algorithm. PPV: positive predictive value.



	Sensitivity	PPV	F1 score
Full Algorithm	0.88	0.63	0.74
High Feasibility Algorithm	0.77	0.57	0.66
Early Algorithm	0.92	0.53	0.67
AS code, Any Specialty	0.70	0.61	0.65
AS code, Rheumatology	0.45	0.84	0.59

Figure 3. Performance of identification methods in the axSpA-enriched population, as measured by unweighted sensitivity, PPV, and F1 scores. AS: ankylosing spondylitis; axSpA: axial spondyloarthritis; PPV: positive predictive value.



	Sensitivity	PPV	F1 Score
Full Algorithm	0.66	0.63	0.65
High Feasibility Algorithm	0.60	0.44	0.51
Early Disease Algorithm	0.90	0.38	0.53
AS codes, Any Specialty	0.55	0.54	0.54
AS codes, Rheumatology	0.34	0.84	0.48

Figure 4. Performance estimates of identification methods for the general veteran population, as measured with inverse probability weighted sensitivity, PPV, and F1 scores. AS: ankylosing spondylitis; PPV: positive predictive value.

(any specialty), and AS Codes (rheumatology). The PPV ranged from 0.38–0.84. The AS codes (rheumatology) had statistically higher PPV than the other 4 identification methods. F1 scores ranged from 0.48 to 0.65. The Full Algorithm had the numerically highest F1 score, but the F1 scores for all methods were statistically similar.

DISCUSSION

We developed novel methods for identifying axSpA patients in national VHA datasets. Previous validation work with axSpA identification algorithms demonstrated excellent performance in axSpA-enriched populations[18,19]. This study is important for understanding how the algorithms are estimated to perform in the general veteran population with future studies. The differences in baseline population characteristics (axSpA-enriched vs general VHA) and outcomes (unweighted vs IPW) demonstrate the need to account for verification bias that can occur when the reference population differs from the target population. We addressed verification bias with IPW and found lower algorithm performance estimates with IPW outcomes than unweighted outcomes. While the unweighted outcomes reflect algorithm performance in the axSpA-enriched population (reference population), the IPW outcomes reflect algorithm performance in the general VA population (target population).

The 5 axSpA identification methods evaluated with IPW for use in the general VA population demonstrate various performance attributes. While there is no consensus on optimal methods for axSpA cohort identification, the use of ≥ 2 AS diagnosis codes from any specialty is commonly used[14,15,31,32]. This method has the advantages of being simple and inclusive of patients who may not be attending rheumatology clinics. However, our data suggest that the performance of this method is suboptimal for the general VA population; the 55% sensitivity demonstrated that approximately one-half of axSpA veterans remained unidentified by this method, and the 54% PPV demonstrated that approximately one-half of veterans identified by this method did not have axSpA. A higher PPV (66%) was reported in 102 chart-reviewed patients in a Kaiser Permanente study, but sensitivity could not be determined, since patients without AS diagnosis codes were not reviewed[15].

The highest PPV (84%) occurred with the other traditional method evaluated in this study (≥ 2 AS diagnosis codes from a rheumatology clinic). This PPV was in the range of other studies reporting a PPV of 81%[29] and 100%[33]. The sensitivity estimate, available in the latter study, was higher than in our study (82% vs 34%, respectively). This difference was likely due to the inclusion of all axSpA phenotypes in our study vs only patients classified as AS in the previous study. The method using ≥ 2 AS codes from rheumatology may be appropriate for studies for which a high confidence in axSpA diagnosis is prioritized over inclusivity of diverse axSpA patients. An example of such a study may involve comparing outcomes in axSpA vs mechanical back pain patients.

In contrast to the method requiring AS diagnosis codes from rheumatology, the Early Algorithm had a high sensitivity (90%) and lower PPV (38%). This was expected since the Early Algorithm was designed to enhance identification of patients with early disease and less classic phenotypes. This identification method is unique in that neither SpA diagnosis codes nor rheumatology encounters are considered in the determination of axSpA risk. The Early Algorithm may be particularly useful for screening studies designed to identify undiagnosed or untreated axSpA patients.

The Full Algorithm had the numerically highest overall performance (F1 scores: 0.65 vs 0.48–0.54). This was expected since the Full Algorithm included the largest number of variables. The

Full Algorithm may be considered the best method for studies in which sensitivity and PPV are equally prioritized.

The High Feasibility Algorithm was designed to be less resource-intensive than the Full Algorithm and more accurate than traditional axSpA identification methods. Unfortunately, the High Feasibility Algorithm did not perform better than traditional methods. Since the High Feasibility Algorithm is more resource intensive than the traditional methods, it is unlikely to be used for axSpA identification in large datasets.

Strengths of this study include the development of a method for identifying axSpA patients who are at high risk for being missed with alternative identification methods. In particular, the Early Algorithm does not consider SpA diagnosis codes or rheumatology visits when calculating axSpA risk for classification assignments. Thus, this method may enable previously impractical research in patients with early disease or less classic phenotypes. Other strengths include extensively studied methodology for the algorithm development, including deep evaluations of NLP model performance, comparisons of different methods for algorithm development (e.g., random forest vs lasso vs *K*-nearest neighbor, vs a combination of these methods called SuperLearner), and well-characterized chart review processes completed by rheumatologists specializing in axSpA[18,21,26]. The methods developed with this research will enable further refinement of the algorithms (i.e., additional NLP variables) and identification of patients at high risk of undiagnosed or unaddressed axSpA who may be recommended for a rheumatologic evaluation.

In summary, there are important unmet needs for better axSpA identification methods for real-world research across the broad spectrum of axSpA patients. There are growing opportunities to advance observational research with large data resources and bioinformatics advancements. This study demonstrates that these resources can be successfully used to improve research methods in axSpA. Further, the novel identification algorithms evaluated in this study may expand the scope of observational research in diverse axSpA populations.

ONLINE SUPPLEMENT

Supplementary material accompanies the online version of this article.

REFERENCES

1. Lee CH, Yoon HY. Medical big data: promise and challenges. Kidney Res Clin Pract 2017;36:3-11.
2. Garg N, van den Bosch F, Deodhar A. The concept of spondyloarthritis: where are we now? Best Pract Res Clin Rheumatol 2014;28:663-72.
3. Lubrano E, De Socio A, Perrotta FM. Unmet needs in axial spondyloarthritis. Clin Rev Allergy Immunol 2018;55:332-9.
4. Walsh JA, Adejoro O, Chastek B, Park Y. Treatment patterns of biologics in US patients with ankylosing spondylitis: descriptive analyses from a claims database. J Comp Eff Res 2018;7:369-80.
5. Deodhar A, Mittal M, Reilly P, Bao Y, Manthena S, Anderson J, et al. Ankylosing spondylitis diagnosis in US patients with back pain: identifying providers involved and factors associated with rheumatology referral delay. Clin Rheumatol 2016;35:1769-76.
6. Lu MC, Koo M, Lai NS. Incident spine surgery in patients with ankylosing spondylitis: a secondary cohort analysis of a nationwide, population-based health claims database. Arthritis Care Res 2018;70:1416-20.
7. Wysham KD, Murray SG, Hills N, Yelin E, Gensler LS. Cervical Spinal Fracture and Other Diagnoses Associated With Mortality in Hospitalized Ankylosing Spondylitis Patients. Arthritis Care Res 2017;69:271-7.
8. Wang R, Ward MM. Epidemiology of axial spondyloarthritis: an update. Curr Opin Rheumatol 2018;30:137-43.
9. Baraliakos X, Braun J. Non-radiographic axial spondyloarthritis and ankylosing spondylitis: what are the similarities and differences? RMD Open;1 Suppl 1:e000053.
10. Reveille JD, Weisman MH. The epidemiology of back pain, axial spondyloarthritis and HLA-B27 in the United States. Am J Med Sci 2013;345:431-6.
11. Walsh JA, Zhou X, Clegg DO, Teng C, Cannon GW, Sauer B. Mortality in American Veterans with the HLA-B27 gene. J Rheumatol 2015;42:638-44.
12. Walsh J, Hunter T, Schroeder K, Sandoval D, Bolce R. Trends in diagnostic prevalence and treatment patterns of male and female ankylosing spondylitis patients in the United States, 2006-2016. BMC Rheumatol 2019;3:39.
13. Walsh JA, Song X, Kim G, Park Y. Evaluation of the comorbidity burden in patients with ankylosing spondylitis treated with tumour necrosis factor inhibitors using a large administrative data set. J Pharm Health Serv Res 2018;9:115-21.
14. Sloan VS, Sheahan A, Stark JL, Suruki RY. Opioid use in patients with ankylosing spondylitis is common in the United States: outcomes of a retrospective cohort study. J Rheumatol 2019;46:1450-7.
15. Curtis JR, Harrold LR, Asgari MM, Deodhar A, Salman C, Gelfand JM, et al. Diagnostic prevalence of ankylosing spondylitis using computerized health care data, 1996 to 2009: underrecognition in a US health care setting. Perm J 2016;20:15-151.
16. Zhao SS, Ermann J, Xu C, Lyu H, Tedeschi SK, Liao KP, et al. Comparison of comorbidities and treatment between ankylosing spondylitis and non-radiographic axial spondyloarthritis in the United States. Rheumatology 2019;58:2025-30.
17. Walsh JA, Rozycki M, Yi E, Park Y. Application of machine learning in the diagnosis of axial spondyloarthritis. Curr Opin Rheumatol 2019;31:362-36.
18. Walsh JA, Pei S, Penmetsa G, Hansen JL, Cannon GW, Clegg DO, et al.Identification of axial spondyloarthritis patients in a large dataset: the development and validation of novel methods. J Rheumatol 2020;47:42-9.
19. Zhao SS, Hong C, Cai T, Xu C, Huang J, Ermann J, et al. Incorporating natural language processing to improve classification of axial spondyloarthritis using electronic health records. Rheumatology 2020;59:1059-65.
20. O'Sullivan JW, Banerjee A, Heneghan C, Pluddemann A. Verification bias. BMJ Evid Based Med 2018;23:54-5.
21. Walsh JA, Pei S, Penmetsa GK, Leng J, Cannon GW, Clegg DO, et al. Cohort identification of axial spondyloarthritis in a large healthcare dataset: current and future methods. BMC Musculoskelet Disord 2018;19:317.
22. Walsh JA, Shao Y, Leng J, He T, Teng CC, Redd D, et al. Identifying axial spondyloarthritis in electronic medical records of US veterans. Arthritis Care Res 2017;69:1414-20.
23. Breiman, L. (2001) Random forests. Machine Learning 2001; 45:5-32.
24. Touw WG, Bayjanov JR, Overmars L, Backus L, Boekhorst J, Wels M, et al. Data mining in the life sciences with Random Forest: a walk in the park or lost in the jungle? Brief Bioinform 2013; 14:315-26.

25. Lin WJ, Chen JJ. Class-imbalanced classifiers for high-dimensional data. Brief Bioinform 2013;14:13-26.
26. Rodríguez JD, Pérez A, Lozano JA. Sensitivity analysis of kappa-fold cross validation in prediction error estimation. IEEE Trans Pattern Anal Mach Intell 2010;32:569-75.
27. Leng CJ, South B, Shuying S. eHOST: The Extensible Human Oracle Suite of Tools. [Internet. Accessed December 15, 2020]. Available from: orbit.nlm.nih.gov/browse-repository/software/nlp-information-extraction/62-ehost-the-extensible-human-oracle-suite-of-tools
28. de Groot JA, Janssen KJM, Zwinderman AH, Bossuyt PMM, Reitsma JB, Karel Moons KGM. Correcting for partial verification bias: a comparison of methods. Ann Epidemiol 2011;21:139-48.
29. Peters CB, Hansen JL, Halwani A, Cho ME, Leng J, Huynh T, et al. Validation of algorithms used to identify red blood cell transfusion related admissions in veteran patients with end stage renal disease. EGEMS 2019;7:23.
30. Sasaki Y. The truth of the F-measure. Teach Tutor mater 2007;1:1-5.
31. Walsh JA, Pei S, Penmetsa GK, Sauer BC, Patil V, Walker JH, et al. Treatment patterns with disease-modifying antirheumatic drugs in U.S. veterans with newly diagnosed rheumatoid arthritis, psoriatic arthritis, or ankylosing spondylitis. J Manag Care Spec Pharm 2019;25:1218-28.
32. Wu LC, Leong PY, Yeo KJ, Li TY, Wang YH, Chiou JY, et al. Celecoxib and sulfasalazine had negative association with coronary artery diseases in patients with ankylosing spondylitis: a nation-wide, population-based case-control study. Medicine 2016;95:e4792.
33. Singh JA, Holmgren AR, Krug H, Noorbaloochi S. Accuracy of the diagnoses of spondylarthritides in veterans affairs medical center databases. Arthritis Rheum 2007; 15;57:648-55.
34. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005;43:1130-9.