Identification of Axial Spondyloarthritis Patients in a Large Dataset: The Development and Validation of Novel Methods

Jessica A. Walsh, Shaobo Pei, Gopi Penmetsa, Jared Lareno Hansen, Grant W. Cannon, Daniel O. Clegg, and Brian C. Sauer

ABSTRACT. ***Objective.*** Observational axial spondyloarthritis (axSpA) research in large datasets has been limited by a lack of adequate methods for identifying patients with axSpA, because there are no billing codes in the United States for most subtypes of axSpA. The objective of this study was to develop methods to accurately identify patients with axSpA in a large dataset.

Methods. The study population included 600 chart-reviewed veterans, with and without axSpA, in the Veterans Health Administration between January 1, 2005, and June 30, 2015. AxSpA identification algorithms were developed with variables anticipated by clinical experts to be predictive of an axSpA diagnosis [demographics, billing codes, healthcare use, medications, laboratory results, and natural language processing (NLP) for key SpA features]. Random Forest and 5-fold cross validation were used for algorithm development and testing in the training subset (n = 451). The algorithms were additionally tested in an independent testing subset (n = 149).

Results. Three algorithms were developed: Full algorithm, High Feasibility algorithm, and Spond NLP algorithm. In the testing subset, the areas under the curve with the receiver-operating characteristic analysis were 0.96, 0.94, and 0.86, for the Full algorithm, High Feasibility algorithm, and Spond NLP algorithm, respectively. Algorithm sensitivities ranged from 85.0% to 95.0%, specificities from 78.0% to 93.6%, and accuracies from 82.6% to 91.3%.

Conclusion. Novel axSpA identification algorithms performed well in classifying patients with axSpA. These algorithms offer a range of performance and feasibility attributes that may be appropriate for a broad array of axSpA studies. Additional research is required to validate the algorithms in other cohorts. (J Rheumatol First Release September 15 2019; doi:10.3899/jrheum.181005)

Key Indexing Terms:
SPONDYLOARTHROPATHY
ANKYLOSING SPONDYLITIS
COHORT STUDIES
DATABASES

Big data research is important for studying uncommon outcomes and diseases in real-world settings[1]. In particular, there are tremendous opportunities to improve knowledge gaps with axial spondyloarthritis (axSpA) with big data research, because axSpA concepts have broadened in recent years[2,3]. With advances in imaging and treatment, it became apparent in the 2000s that a large proportion of people with axSpA phenotypes were unrecognized because their diseases were inconsistent with traditional concepts of axSpA. Despite widespread acceptance of the broader axSpA concepts, big data axSpA research continues to be constrained by outdated axSpA definitions, because International Classification of Diseases, 9th and 10th revisions (ICD-9 and ICD-10) billing codes exist only for the traditionally recognized phenotype of ankylosing spondylitis (AS)[4,5,6,7,8]. Thus, nearly one-half of the 3.2 million Americans with axSpA have been excluded from big data axSpA research[9,10], and there are insufficient data with important outcomes such as mortality, comorbidities, and healthcare use in axSpA populations[8].

Cohort identification is additionally challenging for axSpA because the nomenclature surrounding axSpA is diverse and evolving[11]. The term *axSpA* was introduced in 2009[12], when it became apparent that many patients with axSpA phenotypes were not identified with traditional criteria for SpA in the axial skeleton[13]. Since then, new terms have

From the Salt Lake City Veteran Affairs Medical Center; the University of Utah Medical Center, Salt Lake City, Utah, USA.

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

J.A. Walsh, MD, Salt Lake City Veteran Affairs Medical Center, and University of Utah Medical Center; S. Pei, PhD, Salt Lake City Veteran Affairs Medical Center; G. Penmetsa, MD, University of Utah Medical Center; J.L. Hansen, MStat, Salt Lake City Veteran Affairs Medical Center; G.W. Cannon, MD, Salt Lake City Veteran Affairs Medical Center, and University of Utah Medical Center; D.O. Clegg, MD, Salt Lake City Veteran Affairs Medical Center, and University of Utah Medical Center; B.C. Sauer, PhD, Salt Lake City Veteran Affairs Medical Center, and University of Utah Medical Center.

Address correspondence to Dr. J.A. Walsh, Division of Rheumatology, School of Medicine, 30 North 1900 East, Salt Lake City, Utah 84132, USA. E-mail: jessica.walsh@hsc.utah.edu

Accepted for publication February 20, 2019.

been coined to describe axSpA phenotypes[14], and the use of axSpA terms has varied widely among patients, providers, and other interested parties.

With other conditions, various approaches have been used for cohort identification, when billing codes were insufficient. The most common include rule-based approaches and natural language processing (NLP)[15]. With a rule-based approach, combinations of structured (coded) data may be used to identify patients with a specific condition. For example, ICD codes, disease-modifying antirheumatic drugs (DMARD), and laboratory data [rheumatoid factor or anticyclic citrullinated peptide antibodies (anti-CCP)] were used to identify patients with rheumatoid arthritis (RA)[16]. With NLP, computers can be trained to identify language in the free text of clinical documents that indicates the presence of specific conditions. For example, computers may be trained to identify variations of the term *rheumatoid arthritis* and to use the surrounding text to classify the terms as "yes" (RA present) or "no" (RA not present).

Our goal was to develop accurate methods for identifying patients with axSpA in large datasets. Given the challenges with insufficient billing codes and evolving axSpA nomenclature, we elected to use a combination of coded and NLP data. In our previous work, we first described the development of 3 NLP algorithms that accurately classified axSpA concepts, including diagnostic language (spond*) and key disease features (sacroiliitis and HLA-B27 positivity)[17]. Second, we described our strategy and processes for establishing an appropriate sample of patients for developing and testing axSpA identification algorithms[18]. In this third stage of developing axSpA identification methods, the objectives were to develop and validate algorithms to accurately classify patients as having or not having axSpA.

MATERIALS AND METHODS

Design, setting, and data sources. This study used historical data from veterans enrolled in the US Veterans Health Administration (VHA). The data source was the Corporate Data Warehouse, a national repository of data from the VHA medical record system and other VHA clinical and administrative systems[19]. The patient Integration Control Number was used to link patients across VHA stations. Data were housed and analyzed within the Veterans Affairs (VA) Informatics and Computing Infrastructure[20]. This research was conducted in compliance with the Helsinki Declaration, with the approval of the University of Utah Institutional Review Board (IRB_00052363).

Population. The study population consisted of 600 veterans enrolled in the VHA between January 1, 2005, and June 30, 2015. A detailed description of this patient sample and the processes for selection and chart review of the sample was previously published[19]. In short, a risk-stratified approach to selecting patients was applied that enriched the population with patients at high risk of axSpA to ensure that a sufficient number of patients in the sample had axSpA. To maximize generalizability, patients at low risk for axSpA were also included. Risk was assigned according to variables that clinical experts anticipated to be associated with high, intermediate, and low risk for axSpA. Veterans with HLA-B27 positivity or ≥ 1 AS ICD-9 code were assigned to the high-risk stratum. Veterans with ≥ 1 ICD-9 code for a non-AS SpA subtype or sacroiliitis were assigned to the moderate-risk stratum. Veterans with ≥ 1 ICD-9 code for a SpA mimic or chronic back pain were assigned to the low-risk stratum. From each stratum, 200 veterans (600 total) were randomly selected into the study sample.

Rheumatologist chart reviewers classified the 600 sampled veterans as having or not having axSpA, according to expert opinion and chart review guidelines (Supplementary Data 1, available with the online version of this article). Of the 600 sampled patients, 162 (27.0%) had axSpA and 438 (73.0%) did not have axSpA. Among the 162 patients with axSpA, 125 (77.2%) were from the high-risk stratum, 34 (20.1%) were from the moderate-risk stratum, and 3 (1.9%) were from the low-risk stratum. Among the 438 patients without axSpA, 75 (17.1%) were from the high-risk stratum, 166 (37.9%) were from the moderate-risk stratum, and 197 (45.0%) were from the low-risk stratum. The sample of 600 chart-reviewed patients was randomly subdivided into a training set (n = 451) for algorithm development and a testing set (n = 149) for independent validation of the algorithms.

Variables. Clinical experts selected 49 variables anticipated to be useful in differentiating people with axSpA from people without axSpA. These variables included both structured data and extracts from unstructured medical notes. Structured data included diagnosis codes (Supplementary Table 1, available with the online version of this article) for SpA and overlapping conditions (AS, undifferentiated SpA, Crohn disease, uveitis, back pain, etc.), laboratory data relevant to SpA [C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), HLA-B27], medications frequently used to treat SpA (biologic and synthetic disease-modifying antirheumatic drugs), healthcare use patterns (no. rheumatology visits, no. visits with other provider types), and comorbidities as measured by the Rheumatic Disease Comorbidity Index (RDCI)[21]. NLP was used to extract information not readily available in structured data, including the concepts of sacroiliitis, HLA-B27 positivity, and terms containing the "spond-" prefix.

Algorithm development and validation. Three algorithms were developed including the Full algorithm, the High Feasibility algorithm, and the Spond algorithm (R code in Supplementary Data 2, available with the online version of this article)[22]. The Full algorithm is the most inclusive and resource-intensive with 3 NLP algorithms and 46 coded variables. The High Feasibility algorithm included only 16 coded variables. The Spond NLP algorithm included only the NLP algorithm for Spond. Random Forest and 5-fold cross validation were used to develop and test the Full algorithm and High Feasibility algorithm[23,24,25,26]. To reduce bias, out-of-bag error estimates were used within the training subset (n = 451)[23]. Random Forest was not necessary for the Spond NLP algorithm, because it was treated as a single variable.

The development of the NLP algorithms was described in a previous publication[18]. In short, terms representing the concepts of SpA, sacroiliitis, and HLA-B27 positivity were explored by clinical experts, and sections of text containing clinically meaningful terms (snippets) were extracted from clinical notes. With annotation, clinical experts reviewed the snippets and classified them according to whether they represented the intended axSpA concept. With supervised machine learning on the annotated snippets, computers were trained to replicate the clinical experts' snippet classifications (Library for Support Vector Machines, version 3.21)[27]. The accuracies of the NLP algorithms in an independent dataset of annotated snippets were 91.0% for Spond, 92.0% for sacroiliitis, and 99.0% for HLA-B27 positivity. The Spond NLP was used in 2 ways. In the Full algorithm, it served as a variable, in combination with the 46 coded variables and the 2 other NLP variables. Since our preliminary assessment of the Spond NLP suggested that it performed well independently of other variables, it was also validated in this study population as a standalone instrument for classifying patients as having or not having axSpA.

For the High Feasibility algorithm, variables were selected according to importance rankings determined with Random Forest Mean Decrease Gini scores that take into account the magnitude of effect and frequency of each variable in the population[28]. The 3 NLP variables were excluded because they were more resource-intensive to apply than the coded variables. The remaining 46 coded variables were ranked from highest to lowest impor-

tance. Error rates were calculated for 46 candidate algorithms that ranged in size from 1 coded variable to 46 coded variables. The lowest error rate was balanced with the lowest number of coded variables to select the candidate algorithm that served as the High Feasibility algorithm.

Statistics. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were used to assess the performances of the algorithms and the AS ICD-9 codes. Performance was also measured with concordance (accuracy), discordance (percentage of population with false-positive and false-negative classifications), and receiver-operating characteristic (ROC) curves[29]. CI were determined with bootstrapping, with sampling with replacement of the observed data 500 times[30].

RESULTS

Population. Among the 600 veterans selected for chart review, 162 (27%) were classified as having axSpA (Yes axSpA), and the remaining 438 (73%) were classified as not having axSpA (No axSpA). Compared to the No axSpA group, the Yes axSpA group had a younger mean age, a higher percentage of males, a lower percentage of CCP positivity, a higher percentage with HLA-B27 testing, and a higher percentage with DMARD exposure. Demographic details and other characteristics of the subjects are in Table 1.

Variable selection for the High Feasibility algorithm. The variables with the highest Random Forest Mean Decrease Gini importance scores were selected for the High Feasibility algorithm (Figure 1). The number of AS ICD-9 codes and number of rheumatology visits during the study period had the highest importance scores. Additional variables ranked among the top 16 most important included number of tests for inflammatory markers (CRP, ESR), number of provider visits (any specialty), exposure to a biologic DMARD, age, geographic region, duration of VA system use, HLA-B27 result, and number of ICD-9 codes indicative of various axial skeleton conditions.

Table 1. Demographics, characteristics, and healthcare use in sample of veterans at risk for axSpA.

Variables	Yes axSpA, n = 162				No axSpA, n = 438			
	N or Mean	SD or %	95% CI		N or Mean	SD or %	95% CI	
Demographics								
Age at cohort entry, yrs	56.2	13.5	54.1	58.3	59.8	13.3	58.6	61.1
Sex, male	155	95.7	91.4	97.9	391	89.3	86.0	91.8
Race								
White	128	79.0	72.1	84.6	327	74.7	70.4	78.5
Black	18	11.1	7.1	16.9	67	15.3	12.2	19.0
Other	2	1.2	0.3	4.4	6	1.4	0.6	3.0
Unknown	14	8.6	5.2	14.0	38	8.7	6.4	11.7
Ethnicity								
Non-Hispanic	144	88.9	83.1	92.9	394	90.0	86.8	92.4
Hispanic	7	4.3	2.1	8.7	21	4.8	3.2	7.2
Unknown	11	6.8	3.8	11.8	23	5.3	3.5	7.8
Geographic region at cohort entry								
Southeast	57	35.2	28.3	42.8	167	38.1	33.7	42.8
North Atlantic	35	21.6	16.0	28.6	91	20.8	17.2	24.8
Midwest	30	18.5	13.3	25.2	70	16.0	12.9	19.7
Continental	22	13.6	9.1	19.7	62	14.2	11.2	17.7
Pacific	18	11.1	7.1	16.9	48	11.0	8.4	14.2
Laboratory tests								
C-reactive protein, mean (mg/l)	19.4	29.4	14.2	24.5	18.3	39.3	12.7	24.0
ESR, mean, mm/h	25.3	26.1	20.9	29.7	22.1	23.9	19.2	25.0
RF tested	65	40.1	32.9	47.8	162	37.0	32.6	41.6
RF-positive, among tested patients	9	13.9	7.5	24.3	28	17.3	12.2	23.9
CCP tested	33	20.4	14.9	27.2	68	15.5	12.4	19.2
CCP-positive, among tested patients	0	0.0	0.0	0.0	12	17.7	10.4	28.4
HLA-B27 tested	72	44.4	37.0	52.1	81	18.5	15.1	22.4
HLA-B27–positive, among tested patients	59	81.9	71.5	89.1	66	81.5	71.7	88.4
HLA-B27–positive, among tested patients not selected into population for HLA-B27 positivity*	21	61.8	45.0	76.1	4	21.1	11.8	48.8
DMARD								
≥ 1 biologic during study period	72	44.4	37.0	52.1	40	9.1	6.8	12.2
≥ 1 nonbiologic during study period	52	32.1	25.4	39.6	81	18.5	15.1	22.4
Healthcare use within the VA								
Active system use duration during study period, yrs	9.3	2.0	9.0	9.6	9.0	2.4	8.7	9.2
No. provider visits/yr during active system use	43.6	39.3	37.5	49.6	44.5	42.0	40.6	48.5

* The 100 patients who were selected into the study population specifically because of a positive HLA-B27 test were excluded. RF: rheumatoid factor; CCP: cyclic citrullinated peptide antibody; DMARD: disease-modifying antirheumatic drugs; ESR: erythrocyte sedimentation rate; axSpA: axial spondyloarthritis; VA: (US) Veterans Affairs.

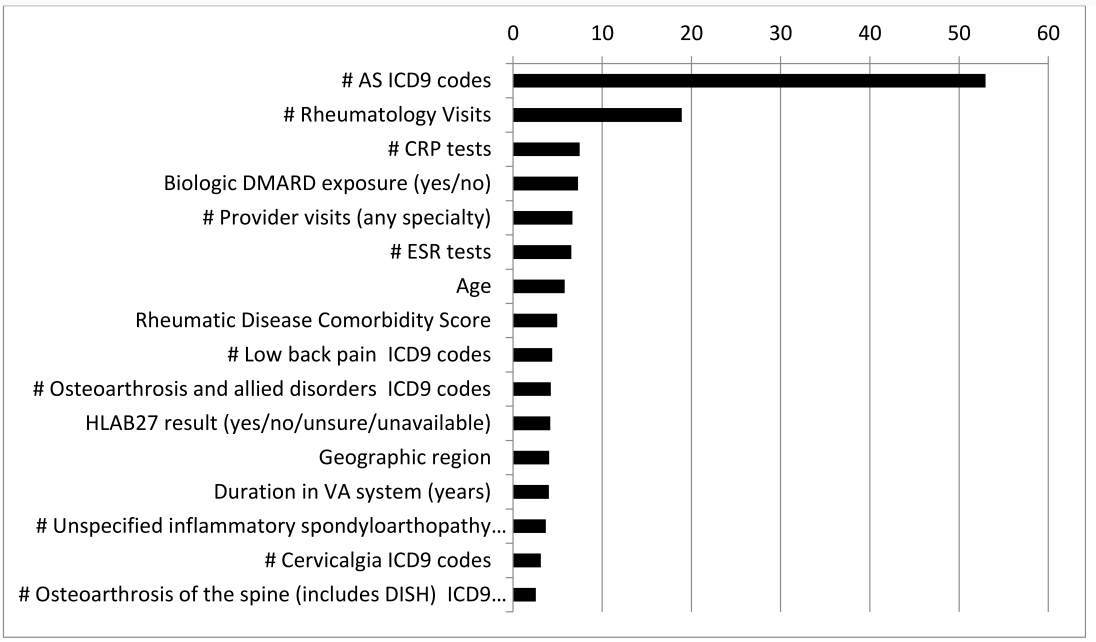


Figure 1. Variables in High Feasibility algorithm according to Random Forest importance scores. ICD9: International Classification of Diseases, 9th revision; CRP: C-reactive protein; DMARD: disease-modifying antirheumatic drug; ESR: erythrocyte sedimentation rate; VA: Veterans Affairs; DISH: diffuse idiopathic skeletal hyperostosis.

Performance of axSpA identification methods: traditional methods. In the subset of patients at risk for axSpA who were not selected to the study population specifically because of an AS ICD-9 code (n = 500), the sensitivity, specificity, PPV, and NPV of an AS ICD-9 code for axSpA were 57.3%, 96.9%, 76.8%, and 92.8%, respectively.

Performance of axSpA identification methods: novel methods. In the testing subset, the areas under the curve with the ROC analysis for the Full algorithm, High Feasibility algorithm, and Spond NLP algorithm were 0.96, 0.94, and 0.86, respectively (Figure 2). The sensitivity, specificity, PPV, and NPV were similar in the training and testing subsets

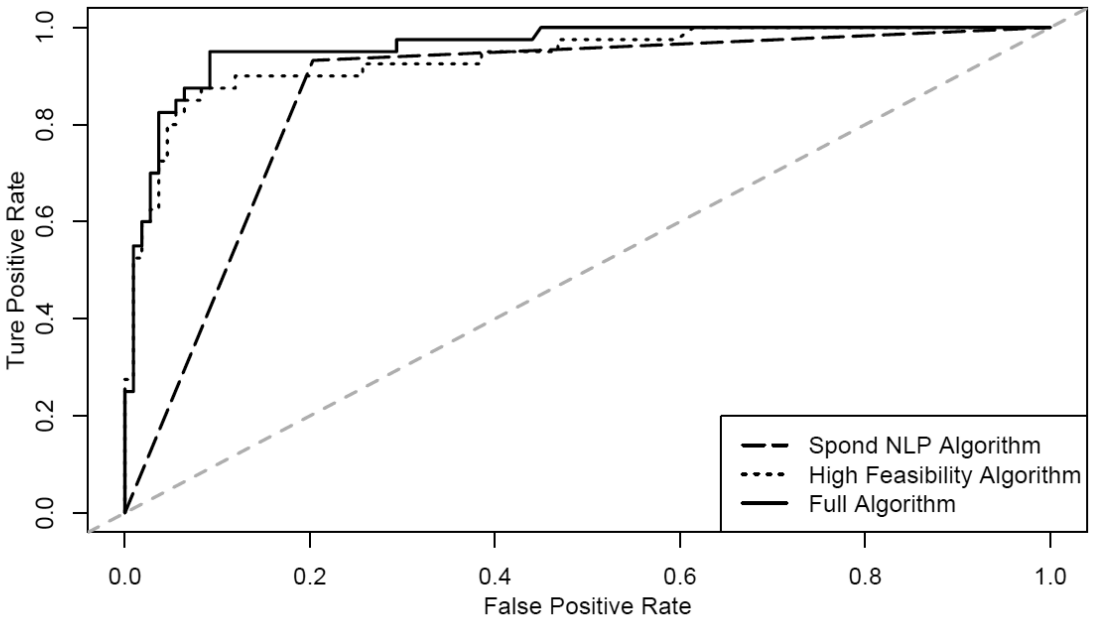


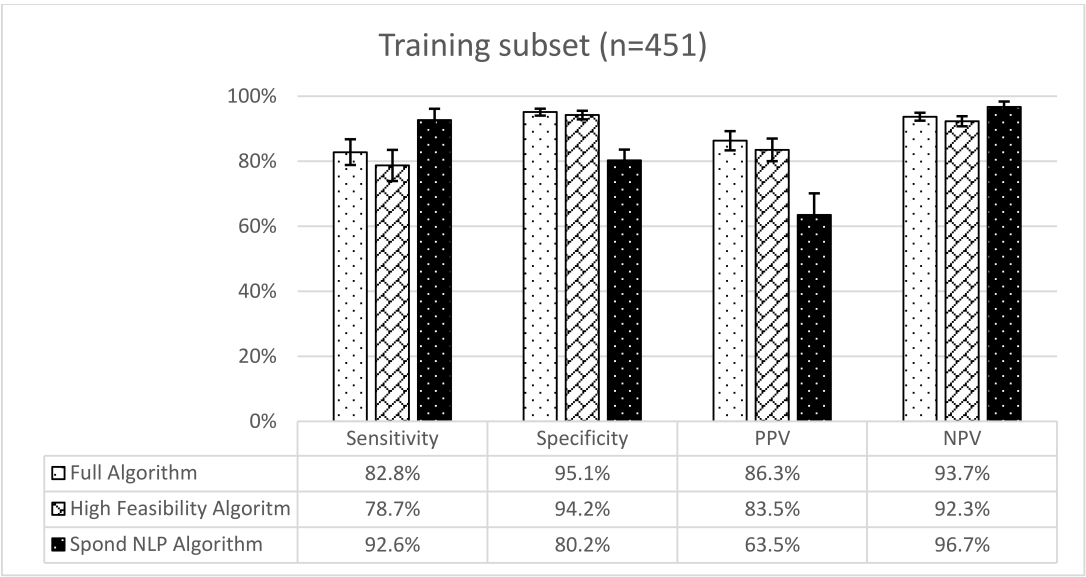
Figure 2. Receiver-operating characteristic curves in the testing subset (n = 149). Area under the curve: Spond NLP algorithm 0.86, High Feasibility algorithm 0.94, Full algorithm 0.96. NLP: natural language processing.

(Figure 3). In the testing subset, the sensitivity, specificity, PPV, and NPV of the Full algorithm were 87.5%, 91.7%, 79.5%, and 95.2%, respectively. For the High Feasibility algorithm, the sensitivity, specificity, PPV, and NPV were 85.0%, 93.6%, 82.9%, and 94.4%, respectively. For the Spond NLP algorithm, the sensitivity, specificity, PPV, and NPV were 95.0%, 78.0%, 61.3%, and 97.7%, respectively.

The classification concordance (accuracy) was also similar in the training and testing subsets (Figure 4). In the testing subset, concordance was achieved with 90.6%, 91.3%, and 82.6% of patients with the Full algorithm, High Feasibility algorithm, and the Spond NLP algorithm, respectively. The percentages of the population with false-positive and false-negative classifications were similar (3.4%–6.0% testing subset) for the Full algorithm and the High Feasibility algorithm. With the Spond NLP algorithm, the percentage of false positives was higher than the percent of false negatives (16.1% vs 1.3% testing subset).

DISCUSSION

We developed novel methods for identifying patients with axSpA. These methods will enable the development of more inclusive axSpA cohorts than traditional cohort identification methods and may be used to study a variety of poorly understood outcomes in axSpA, such as mortality, comorbidities, treatment patterns, healthcare use, and costs.



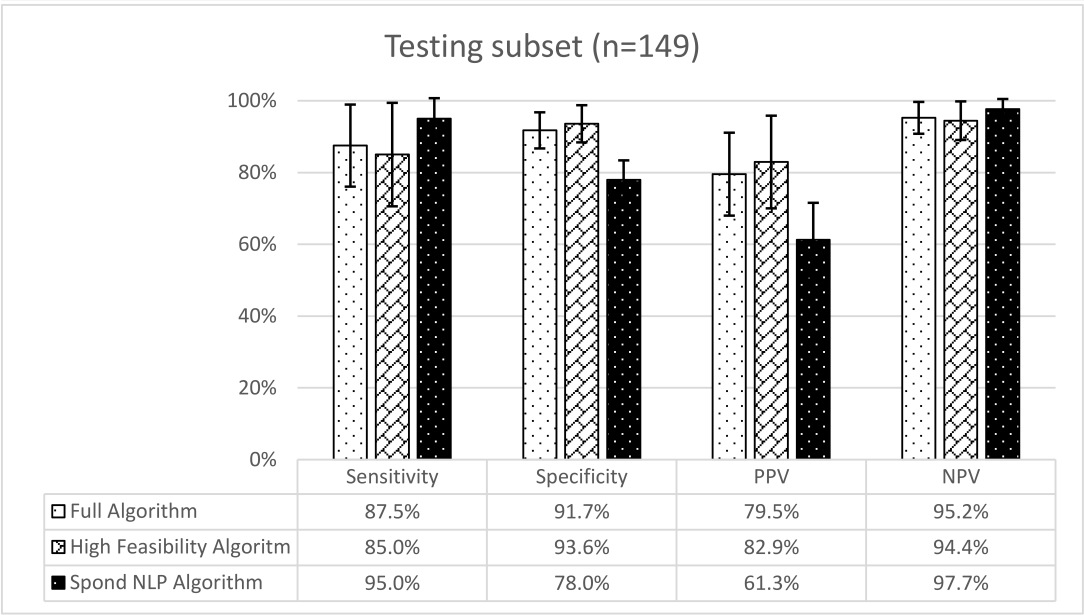


Figure 3. Sensitivity, specificity, PPV, and NPV of axial spondyloarthritis identification methods. PPV: positive predictive value; NPV: negative predictive value; NLP: natural language processing.

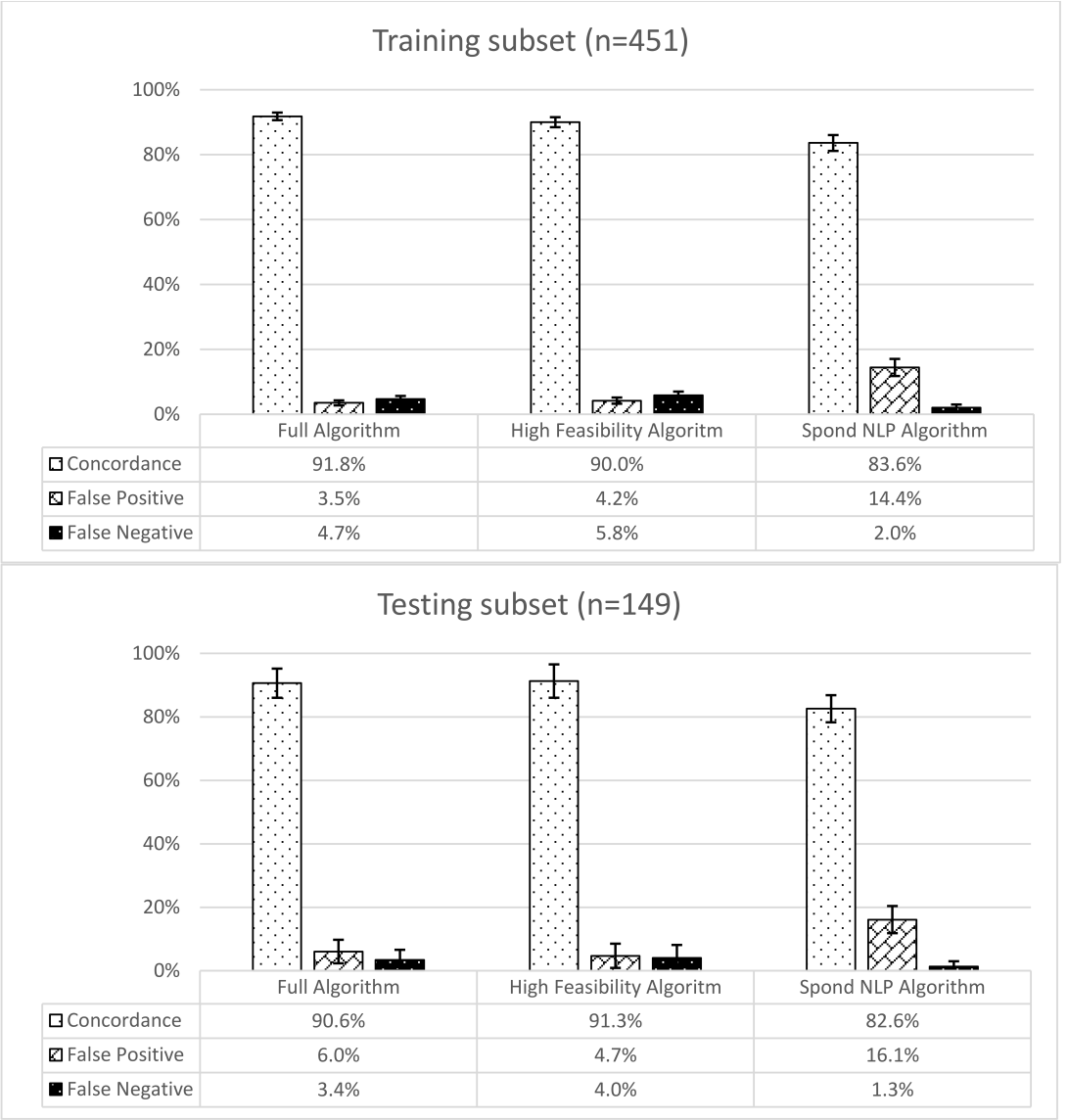


Figure 4. Algorithm concordance (accuracy) with axial spondyloarthritis classification by chart review. NLP: natural language processing.

With other conditions, cohort identification methods have been applied that similarly used combinations of structured (coded) data and unstructured data from clinical notes. For example, coded data and simple queries in the text of clinical documents were used to identify dialysis patients, with precision (PPV) of 78.4% and recall (sensitivity) of 100%[31]. Structured data and unstructured data were also successfully used to identify patients eligible for clinical trials (ROC = 0.95)[32]. In another study, NLP methods were compared to structured data extraction methods for identifying patients undergoing diabetic dialysis; the NLP methods were more sensitive than the coded data extraction methods (89.4% vs 54.9%)[16]. These studies demonstrated that methods using unstructured data (simple queries and NLP) are useful in cohort identification and may be particularly powerful when combined with structured data extraction methods.

The algorithms developed in our study differ in their performance and feasibility and may be used for different purposes. The Full algorithm and the High Feasibility algorithm had high specificity and may be implemented when low false-positive rates are desired (i.e., treatment considerations). Conversely, the Spond NLP model was more sensitive and may be best when more inclusive identification is desired (i.e., screening for people at high risk of axSpA). The High Feasibility algorithm is the simplest to use and may be applied and tested relatively easily in different datasets.

Strengths of our study include the use of robust, well-characterized data. The study sample was classified and

phenotyped by rheumatologist chart reviewers specializing in SpA[19]. The NLP variables were highly accurate and previously validated in a population of veterans[18]. Methods for extracting, cleaning, and interpreting the coded variables used in this project were also evaluated and refined for the study population[19,33,34,35,36,37].

Limitations of our study include the limited generalizability of the algorithms to other datasets. While the chart-reviewed population in which the algorithms were developed was designed to maximize both feasibility and generalizability, this population is not identical to the general veteran population. Currently the algorithms can be applied to VA populations at risk for axSpA who are selected in the same manner that the chart-reviewed population was selected (with specific variables associated with high, moderate, and low risk of axSpA). To bypass this initial selection step and make the algorithms generalizable to the general veteran population, the algorithms may be tested and refined in that population. Likewise, the VA population is different from other large data populations and the algorithms will require testing prior to use in non-VA datasets.

Another limitation is that the algorithms with NLP (Full algorithm and Spond algorithm) are relatively resource-intensive when applied on a large scale, requiring both bioinformatics expertise and computing resources. Further, the PPV in this study population were likely overestimated relative to the general population and the NPV were likely underestimated, because disease prevalence influences predictive values and the study population was enriched for axSpA. The Full and High Feasibility algorithms were also limited in that they were not inclusive of ICD-10 codes, because ICD-10 codes were not yet implemented in clinical practice during the study period.

Novel methods for accurately identifying patients with axSpA in a large dataset were developed. Compared to traditional methods of cohort identification, the novel methods were more inclusive (sensitive) and more representative of the broadening concepts of axSpA. Additional research is required to apply these methods to other populations to facilitate a wide range of previously impractical big data research in axSpA.

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 2015;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. Braun J. Axial spondyloarthritis: thoughts about nomenclature and treatment targets. Clin Exp Rheumatol 2012;4 Suppl 73:S132-5.
12. Rudwaleit M, van der Heijde D, Landewé R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis 2009;68:777–83.
13. Poddubnyy D, Sieper J. Similarities and differences between nonradiographic and radiographic axial spondyloarthritis: a clinical, epidemiological and therapeutic assessment. Curr Opin Rheumatol 2014;26:377-83.
14. Slobodin G, Eshed I. Non-radiographic axial spondyloarthritis. Isr Med Assoc J 2015;17:770-6.
15. Sarmiento RF, Dernoncourt F. Improving patient cohort identification using natural language processing. In: Secondary analysis of electronic health records. New York: Springer, Cham; 2016:405–17.
16. Liao KP, Cai T, Gainer V, Goryachev S, Zeng-treitler Q, Raychaudhuri S, et al. Electronic medical records for discovery research in rheumatoid arthritis. Arthritis Care Res 2010;62:1120-7.
17. 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.
18. Walsh JA, Pei S, Penmetsa G, Hansen JL, 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.
19. Fihn S, Francis J, Clancy C, Neilson C, Nelson K, Rumsfeld J, et al. Insights from advanced analytics at the Veteran Health Administration. Health Affairs 2014;33:1203-11.
20. U.S. Department of Veterans Affairs. VA Informatics and Computing Infrastructure (VINCI). [Internet. Accessed June 10, 2019.] Available from: www.hsrd.research.va.gov/for_researchers/vinci
21. England BR, Sayles H, Mikuls TR, Johnson DS, Michaud K. Validation of the rheumatic disease comorbidity index. Arthritis Care Res 2015;67:865-72.
22. R Core Team. The R Project for Statistical Computing. [Internet. Accessed June 10, 2019.] www.R-project.org
23. Breiman L. Random forests. In: Machine learning. New York: Springer US; 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. Chang CC, Lin CJ. LIBSVM — a library for support vector machines. [Internet. Accessed June 10, 2019.] Available from: www.csie.ntu.edu.tw/~cjlin/libsvm
28. Wang H, Yang F, Luo Z. An experimental study of the intrinsic stability of random forest variable importance measures. BMC Bioinformatics 2016;17:60.
29. Carter JV, Pan J, Rai SN, Galandiuk S. ROC-ing along: evaluation and interpretation of receiver operating characteristic curves. Surgery 2016;159:1638-45.
30. Varian H. Bootstrap tutorial. Math J 2005;9:768-75.
31. Abhyankar S, Demner-Fushman D, Callaghan FM, McDonald CJ. Combining structured and unstructured data to identify a cohort of ICU patients who received dialysis. J Am Med Inform Assoc 2014;21:801-7.
32. Miotto R, Weng C. Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials. J Am Med Inform Assoc 2015;22:e141–50.
33. Walsh JA, Pei S, Burningham Z, Penmetsa G, Cannon GW, Clegg DO, et al. Use of disease-modifying antirheumatic drugs for inflammatory arthritis in US veterans: effect of specialty care and geographic distance. J Rheumatol 2018;45:430-6.
34. Walsh JA, Zhou X, Clegg DO, Teng C, Cannon GW, Sauer B, et al. Mortality in American veterans with the HLA-B27 gene. J Rheumatol 2015;42:638-44.
35. Cannon GW, Mikuls TR, Hayden CL, Ying J, Curtis JR, Reimold AM, et al. Merging Veterans Affairs rheumatoid arthritis registry and pharmacy data to assess methotrexate adherence and disease activity in clinical practice. Arthritis Care Res 2011;63:1680-90.
36. Nelson SD, Lu CC, Teng CC, Leng J, Cannon GW, He T, et al. The use of natural language processing of infusion notes to identify outpatient infusions. Pharmacoepidemiol Drug Saf 2015;24:86-92.
37. Sauer B, Teng CC, Burningham Z, Cannon G. Errata to NLP study of infusion notes to identify outpatient infusions in the VA. Pharmacoepidemiol Drug Saf 2015;24:1225-6.