



Use of machine learning techniques in the development and refinement of a predictive model for early diagnosis of ankylosing spondylitis

Atul Deodhar¹ · Martin Rozycki² · Cody Garges³ · Oodaye Shukla² · Theresa Arndt² · Tara Grabowsky² · Yujin Park⁴

Received: 9 October 2018 / Revised: 28 January 2019 / Accepted: 8 April 2019 / Published online: 1 May 2019

© The Author(s) 2019

Abstract

Objective To develop a predictive mathematical model for the early identification of ankylosing spondylitis (AS) based on the medical and pharmacy claims history of patients with and without AS.

Methods This retrospective study used claims data from Truven databases from January 2006 to September 2015 (Segment 1) and October 2015 to February 2018 (Segment 2). Machine learning identified features differentiating patients with AS from matched controls; selected features were used as inputs in developing Model A/B to identify patients likely to have AS. Model A/B was trained and developed in Segment 1, and patients predicted to have AS in Segment 1 were followed up in Segment 2 to evaluate the predictive capability of Model A/B.

Results Of 228,471 patients in Segment 1 without any history of AS, Model A/B predicted 1923 patients to have AS. Ultimately, 1242 patients received an AS diagnosis in Segment 2; 120 of these were correctly predicted by Model A/B, yielding a positive predictive value (PPV) of 6.24%. The diagnostic accuracy of Model A/B compared favorably with that of a clinical model (PPV, 1.29%) that predicted AS based on spondyloarthritis features described in the Assessment of SpondyloArthritis international Society classification criteria. A simplified linear regression model created to test the operability of Model A/B yielded a lower PPV (2.55%).

Conclusions Model A/B performed better than a clinically based model in predicting a diagnosis of AS among patients in a large claims database; its use may contribute to early recognition of AS and a timely diagnosis.

Keywords Ankylosing spondylitis · Machine learning · Mathematical model · Spondyloarthritis

Cody Garges was an employee of HVH Precision Analytics, LLC, Wayne, PA, USA, at the time of the analysis

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10067-019-04553-x>) contains supplementary material, which is available to authorized users.

✉ Atul Deodhar
deodhara@ohsu.edu

¹ The Oregon Health & Science University, 3181 SW Sam Jackson Park Rd., Portland, OR 97239, USA

² HVH Precision Analytics, Wayne, PA, USA

³ ProMetrics, LLC, King of Prussia, PA, USA

⁴ Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA

Introduction

Ankylosing spondylitis (AS) is a chronic inflammatory disease manifested by progressive spinal stiffness and fusion; this disease primarily involves the sacroiliac joints in the initial stages [1–3]. In the USA, it is estimated that the prevalence of AS ranges from 0.2 to 0.5% [4]. Although AS is thought to affect approximately 350,000 people in the USA, it is an underdiagnosed condition and can take approximately 14 years before it is diagnosed correctly [5]. A number of factors likely contribute to the delayed diagnosis of AS, including the broad prevalence of mechanical back pain in the population, gradual onset of the disease, and lack of specific symptoms or biomarkers unique to AS [6]. Additionally, in the USA, the majority of patients experiencing the onset of low back pain visit general practitioners, orthopedists, or chiropractors. The ability of these providers to accurately diagnose AS is unknown, and there are no clear guidelines to refer

patients with suspected AS to a rheumatologist [6, 7]. A lack of diagnostic criteria is another factor that may delay identification of patients with AS. The modified New York AS classification criteria become useful for diagnosis only in later stages of the disease, when the patient has experienced symptoms for many years with possible loss of function [8]. Delayed diagnosis and treatment contribute to the economic, physical, and psychological burdens on patients and their caregivers [6–9].

Multiple studies have identified factors that play a role in delaying diagnosis and developed recommendations for improving referrals [6, 10, 11]; however, results were mixed, with many studies limited by the use of relatively small databases based on electronic health records at a small number of healthcare institutions. Small sample sizes can be overcome through use of a large administrative claims database, although earlier studies using databases were hampered by the application of limited filtering and selection criteria [12, 13]. Predictive systems driven by machine learning techniques, which evolve based on empirical data, are ideal tools for recognizing the patterns obscured by the volume of claims in such a database. Machine learning models can efficiently mine patient information—including diagnoses, medication and treatment plans, and laboratory and test results—from these large databases and provide valuable insight into disease management [14, 15].

We used mutual information (MI) [16] to identify features that differentiated AS from a control population to better understand the journey of patients with AS prior to the receipt of an AS diagnosis. Since patients with AS are routinely misdiagnosed in routine clinical practice [17], we hypothesized that analyzing this process by examining claims data may help distinguish true cases of AS. We used machine learning algorithms to identify features that differentiate patients with AS from a control population; these were then used to predict an AS diagnosis. The overall objective of this analysis was to develop and refine a predictive mathematical model for AS using features observed in the medical and claims history of patients with and without a diagnosis of AS to aid in the earlier identification of the disease.

Materials and methods

Data source

This retrospective cohort study used administrative claims data from > 182 million patients from January 2006 to February 2018 in the Truven Health MarketScan® Commercial and Medicare Supplemental databases. The MarketScan databases represent one of the largest collections of deidentified US patient data available for healthcare research, with 23.9 and 1.9 million covered lives in the Commercial and Medicare

data sets, respectively, in 2016. These databases provide the opportunity to simultaneously sample from various sources (e.g., employers, states, and health plans) and feature representation from > 350 unique carriers.

Patient data consisted of medical and pharmacy claims history, including diagnoses received (represented by the *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)* 720.0 codes), procedures undergone (represented by *Current Procedural Terminology* codes), and prescription drugs received as specified by Therapeutic Detail Code and Therapeutic Class variables from Truven's RED BOOK. Additionally, Clinical Classifications Software was used to categorize services and procedures into clinically meaningful diagnosis and procedure categories [18].

This study was conducted in accordance with the Guidelines for Good Pharmacoepidemiology Practices of the International Society for Pharmacoepidemiology, the Strengthening the Reporting of Observational Studies in Epidemiology guidelines, and the ethical principles in the Declaration of Helsinki. All database records are fully compliant with US patient confidentiality requirements, including the Health Insurance Portability and Accountability Act of 1996. As this study is a retrospective analysis of deidentified claims data in compliance with US patient confidentiality requirements, and did not involve the collection, use, or transmittal of individually identifiable data, institutional review board approval was not required to conduct this study.

Segment 1

Patient data were initially collected between January 2006 and September 2015, the period designated as Segment 1. In Segment 1, a cohort of 6325 patients with AS was identified; these patients had received ≥ 2 ICD-9-CM diagnosis codes of AS given by rheumatologists ≥ 30 days apart and had ≥ 12 months of continuous enrollment in the Truven databases prior to receiving their first AS diagnosis. In addition, 14,832,350 controls in Segment 1 were identified. Each control was matched by age and sex to each patient with AS; additionally, each control had no history of AS and had been enrolled ≥ 1 year prior to the diagnosis date of the patient with AS. The “index date” of the control was defined as the diagnosis date of the AS patient to which they were matched. The entire cohort of patients with AS and 49,198 of the controls were considered the training set; these patients were used to build and train the initial predictive model, Model A, to predict a diagnosis of AS. The application of Model A to the overall control population not used for training revealed that many patients were erroneously predicted to have AS (false positives). Therefore, a second machine learning model, Model B, was trained using the same set of 6325 patients with AS but using

50,000 false positive patients from the overall control population as controls.

Segment 2

The time frame of October 2015 to February 2018, or Segment 2, was the observation window to ascertain whether patients predicted to have AS in Segment 1 (patients scoring above our model score prediction threshold of 0.5; more detail follow) subsequently received an *ICD-10-CM* AS diagnosis code (M45.x or M08.1). A summary of the characteristics of patients and controls in Segments 1 and 2 are shown in Table 1.

Development and refinement of predictive models (Models A and B)

Figure 1 a and b present a high-level overview of the development, validation, and application of the AS-predictive risk models.

Model A

Mutual information [16], a predictor selection technique, was used to extract claims data from patients with AS and the matched control population. Mutual information

Table 1 Summary of the characteristics of the cohorts in our study

Cohort	Description	<i>N</i>	Males (%)	Age (SD) at AS diagnosis or control index date	Year of AS diagnosis or control index date (range)*
Segment 1—AS population	<ul style="list-style-type: none"> •Patient population used in the training of Models A and B •Patients with ≥ 2 diagnosis codes of AS given by rheumatologists •Patients with ≥ 12 months of continuous enrollment in the Truven database prior to their first AS diagnosis 	6325	50.2	45.1 (14.5)	Jan 2006–Sep 2015
Segment 1—overall matched control population	<ul style="list-style-type: none"> •Patients without any AS diagnosis code who were matched by age and sex to patients with AS •Patients with ≥ 12 months of continuous enrollment in the Truven database prior to their index date (corresponding to the diagnosis date of matched patients with AS) 	14,832,350	50.6	44.8 (16.1)	Jan 2006–Sep 2015
Segment 1—Model A matched control population	<ul style="list-style-type: none"> •Subset of the Segment 1 overall control population used to train Model A 	49,198	50.1	45.0 (14.5)	Jan 2006–Sep 2015
Segment 1—Model B matched control population	<ul style="list-style-type: none"> •Subset of the Segment 1 overall control population that was used to train Model B •Patients falsely predicted to have AS when Model A was applied to the Segment 1 overall matched control population •Mutually exclusive from Model A control population 	50,000	44.8	48.8 (14.2)	Jan 2006–Sep 2015
Segment 2—population where Model A and B were applied	<ul style="list-style-type: none"> •Patients without any diagnosis code of AS with ≥ 12 months of continuous enrollment in the Truven database prior to Sep 2015 <ul style="list-style-type: none"> - Risk scores are calculated by applying Models A and B using 1-year data from Sep 2014 to Sep 2015 •Patients with continuous enrollment in the Truven database from Oct 2015 to Feb 2018 <ul style="list-style-type: none"> - All patients were observed in this period to see if they were diagnosed with AS •These patients were not used in the training of Models A and B 	227,229	36.2	53.4 (15.5)	Oct 2015–Feb 2018
Segment 2—patients newly diagnosed with AS	<ul style="list-style-type: none"> •Subset of the Segment 2 population where Models A and B were applied •These patients received their first AS diagnosis after October 1, 2015, in the Truven database 	1242	37.2	20.1 (14.4)	Oct 2015–Feb 2018

AS ankylosing spondylitis

*The index date of the control was defined as the diagnosis date of the matched patient with AS

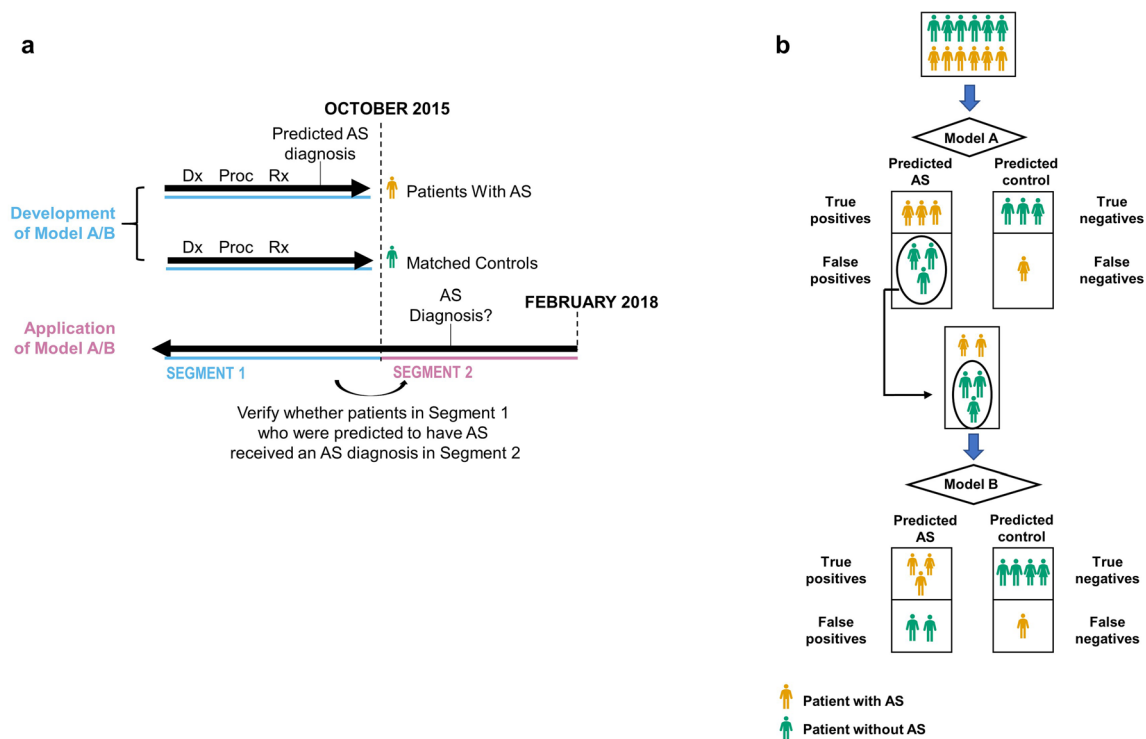


Fig. 1 Overview of model development, where (a) Model A/B was first trained using a subset of patients in Segment 1 and subsequently used to predict an AS diagnosis among the remaining patient population in Segment 1, and (b) patients predicted as likely to have AS were tracked

into Segment 2 to determine whether they eventually were diagnosed with AS. AS ankylosing spondylitis, Dx diagnoses, Proc procedures, Rx prescriptions

identified positive and negative predictors of AS. Model A was first trained using the Segment 1 cohort of 6325 patients with AS and 49,198 matched controls using these predictors. Please see the [Supplementary Appendix](#) for technical details on model development. Model A was subsequently tested on all controls in Segment 1 who were not used for training.

Model B

To improve the accuracy of the machine learning model, a second model (Model B) was trained using the same set of 6325 patients with AS, as well as a control set of 50,000 false positives—i.e., controls whom Model A had erroneously predicted as having AS. Because this control population was considered especially challenging by the machine learning model to accurately predict as having AS, its use for the training of Model B provided a better sense of how the model would perform in more difficult cases. This model served to distinguish true patients with AS from the subset of patients who may be diagnosed with AS.

Model A/B

In order to evaluate the likelihood that a patient has AS, the models were applied sequentially: Model A was first used to eliminate patients who likely did not have AS—i.e., select for patients who scored above the model cutoff threshold of 0.5—and subsequently were processed by Model B to further remove false positives and provide a final risk score. The serial application of Models A and B, our machine learning model, is henceforth referred to as Model A/B.

Evaluation of the optimized predictive models (Model A/B)

Patients predicted to have AS in Segment 1 by Model A/B were followed up through Segment 2 to see whether they truly had AS, i.e., received an *ICD-10-CM* AS diagnosis code (M45.x or M08.1). The overall performance of the model was measured by its diagnostic accuracy. Metrics of diagnostic accuracy include estimates of sensitivity, the potential to recognize true patients with AS; specificity, the ability to identify patients without AS; positive predictive value (PPV), the

proportion of patients predicted to have AS who truly have AS; and area under the receiver operating characteristic curve (AUC) [19], a global measure of diagnostic accuracy that estimates the overall ability to discriminate between 2 groups. An AUC of 0.5 is equivalent to a coin toss, and an AUC of 1 represents a perfect model [19]. AUC is commonly used in the machine learning community to evaluate classifier performance [19].

Lastly, to assess the operability of Model A/B, we created a simpler, linear regression model that operates using the same machine learning codes.

Comparative performance of Model A/B

The performance of Model A/B was also evaluated against that of a clinical model and an “all-AS” model. The clinical model was built to predict a diagnosis of AS based on the spondyloarthritis features described in the Assessment of SpondyloArthritis international Society classification criteria (aged < 45 years; presence of back and joint pain; and presence of ≥ 1 clinical feature of psoriasis, uveitis, inflammatory bowel disease, and/or enthesitis and ≥ 1 nonsteroidal anti-inflammatory drug treatment) [20]. The all-AS model was built to predict that all patients have AS. These predictive models were applied to the same patients identified and tracked from Segment 1 to Segment 2 using Model A/B.

Results

Development and refinement of predictive risk models (Models A and B)

As summarized in Table 1, a total of 6325 patients with AS were identified in the MarketScan databases based on the patient selection criteria of ≥ 2 ICD-9-CM AS diagnosis codes given by a rheumatologist, ≥ 30 days apart, as well as continuous enrollment in the Truven databases for 12 months prior to the first AS code. Concurrently, 14,832,350 demographically matched participants without any ICD-9-CM AS code served as controls (“Segment 1 overall control population”). An overview of the diagnosis, procedure, and prescription codes used in Models A and B from 0 to 12 months before diagnosis is shown in Fig. S1a and b, respectively.

Model A was trained using patients with AS and 49,198 controls selected from Segment 1 overall control population (AUC = 0.81; Fig. S2a). Please see the Supplementary Appendix for technical details on model development. Model A then scored all patients from the Segment 1 overall control population and predicted patients from the control population as having AS. To further improve the diagnostic accuracy of our machine learning model, Model B was built

using 50,000 random patients from the Segment 1 overall control population who scored > 0.5 in Model A (i.e., erroneously identified as patients with AS and hence more challenging to diagnose) as a new control population (AUC = 0.79; Fig. S2b).

Evaluation of Model A/B

Model A/B evaluated 228,471 patients in Segment 1 without any history of AS who were followed in Segment 2. Based on a 0.5 model score prediction threshold, the model predicted that 1923 patients in Segment 1 would develop AS in Segment 2. Of these 1923 patients, 120 had received ≥ 1 ICD-10-CM AS diagnosis code, yielding an AUC of 0.629 and a PPV of 6.24% (120/1923 patients). The sensitivity of Model A/B was 9.66% (120/1242 patients), as there were a total of 1242 patients with ≥ 1 ICD-10-CM AS diagnosis code in Segment 2 and only 120 of 1242 patients were predicted by the model to have AS (Table 2). Figure 2 shows the relationship between PPV and sensitivity as the model cutoff threshold changes. When the model cutoff threshold was increased, the PPV of the model also increased but at the cost of reduced sensitivity. Importantly, the sensitivity of the model decreased with increasing time to AS diagnosis (Fig. S3). For example, patients 5 months away from diagnosis could be predicted to receive an AS diagnosis with 12% sensitivity, and those 25 months away could be predicted with 5% sensitivity. The decrease in sensitivity is expected, as patients who are farther from diagnosis may be less symptomatic and hence have lower signal for the model to detect.

The linear regression model, which is a simplified version of Model A/B (machine learning model), demonstrated higher sensitivity and AUC than Model A/B, although its PPV was orders of magnitude lower. This can be seen in Fig. 2, where the superior performance of Model A/B is apparent at high PPVs.

The diagnostic accuracy of the all-AS and clinical models was also evaluated, along with that of Model A/B in Segment 2. The all-AS model demonstrated a sensitivity of 100%, a PPV of 0.54%, and an AUC of 0.5, indicating no predictive value (Table 2). The clinical model, based on features more likely to be associated with a diagnosis of AS, yielded a PPV of 1.29% and sensitivity of 18.84% (Table 2).

Discussion

The need for earlier referral of a patient with possible AS to a rheumatologist has been recognized for quite some time [11]. Several strategies for timely referral of patients with potential AS to rheumatologists are established in Europe [21–26]. Identifying the population of patients for screening and

Table 2 Measures of diagnostic accuracy of various predictive models

Model	Description	AUC	Sensitivity (%)	FPR (%)	PPV (%)
Model A/B	Machine learning model	0.63	9.66	0.79	6.24
Simple linear regression model	Simplified version of Model A/B	0.71	22.38	4.67	2.55
Clinical model	The following criteria were used: age < 45 years <i>and</i> back/joint pain <i>and</i> psoriasis or uveitis or inflammatory bowel disease <i>and</i> NSAID use	0.55	18.84	7.88	1.29
All-AS	Predicts everyone in database to have AS	0.50	100	100	0.54

AS ankylosing spondylitis, ASAS Assessment of SpondyloArthritis international Society, AUC area under the receiver operating characteristic curve, FPR false positive rate, NSAID nonsteroidal anti-inflammatory drug, PPV positive predictive value

referral may be useful to avoid overwhelming rheumatologists with high volumes of patients with back symptoms.

Regarding time to diagnosis, some patients receive a diagnosis of AS within a year [7], while it may take up to 10 years for others [7, 27]. Educating providers in nonrheumatology settings on how to identify critical, and perhaps elusive, elements for quicker AS diagnosis is challenging. Our exploratory study using a large, US administrative claims database was novel, since we attempted to use a machine learning approach to describe and analyze sequences of health events (e.g., diagnosis, procedures, treatments) in patients with AS to identify predictors that led to an AS diagnosis. More specifically, our objective was to determine the data elements that had statistical relevance to a future AS diagnosis. For example, uveitis and back pain are more common in patients with AS than in the general population, whereas a finger injury may be

equally common in patients with and without AS. Additionally, patients with AS may tend to visit their provider more frequently as they attempt to obtain a diagnosis for their disease. In this example, uveitis, back pain, and frequent doctor visits would have more relevance to an AS diagnosis than a finger injury. Here, we developed, trained, and refined a series of predictive risk models to identify specific features that differentiated patients with AS from controls and prospectively applied the models to a population of patients in Segment 2 to predict a diagnosis of AS. We then compared the performance of Model A/B with that of traditional statistical and clinically based models. The latter two models lack the power to measure and apply these predictors to determine a diagnosis of AS.

Studies within the USA to identify patients with early AS have been hampered by small sample sizes and/or by applying limited filtering and selection criteria to larger databases [6, 7]. The explosive growth in collected data over the past few decades has stimulated the development of sophisticated computational tools designed not only to tabulate and sort through vast databases but also to “learn” about and extract insights from that data [12, 13]. Such machine learning software can identify subtle patterns and connections that older analytical systems were not designed to do. This capability has already been implemented in criminal investigation, business, and the military [13, 28–31]. In healthcare, predictive models have been used to detect patients in claims databases with depression, cataracts, rheumatoid arthritis, and type 2 diabetes mellitus [32–35]. Google recently used machine learning software to identify tumors on mammograms [36] and detect diabetic retinopathy in retinal fundus photographs [37]. The value of machine learning in healthcare is evident in the timely processing of large volumes of data beyond human capability, and the subsequent insights generated can be used to aid in clinical decision-making processes. Our study is directly applicable to patients with AS, and, to our knowledge, is the first such study attempted in a patient population with AS.

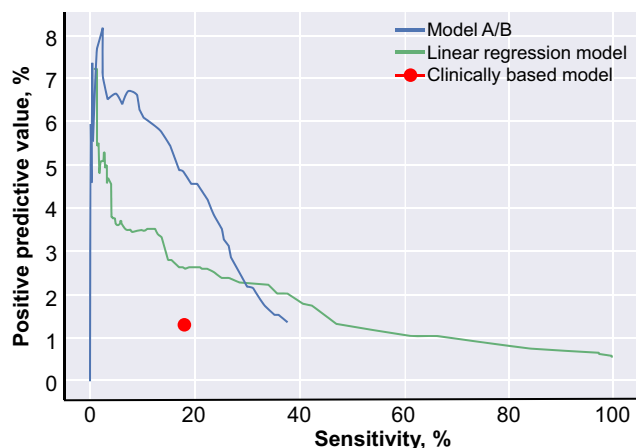


Fig. 2 Comparison of sensitivity and PPV measures across all model thresholds. Sensitivity and PPV are a function of the model cutoff threshold. If the threshold is increased, fewer patients are predicted to have AS, therefore decreasing sensitivity and increasing PPV. At high PPVs, Model A/B outperformed the linear regression and clinical models. AS ankylosing spondylitis, PPV positive predictive value

Although Model A/B has a low PPV, it is important to note that it is still higher than the PPV of both clinical and linear regression models. The linear regression model has a higher AUC and sensitivity than our machine learning model, but it also has a higher false positive rate. There are no fixed standards of diagnostic accuracy; unlike sensitivity or specificity, metrics such as PPV are dependent on the prevalence of the disease [38]. Maximizing the predictive capability of any machine learning model comes with the trade-off of decreased sensitivity. We attempted to improve on this by challenging the algorithm using “AS mimics,” i.e., patients with AS characteristics but who did not have a diagnosis of AS. At the same time, we acknowledge that certain predictors used by our model, or by machine learning models in general, may not be clinically rational; our model should not completely replace the medical or scientific judgment of trained professionals. The use of machine learning algorithms in combination with targeted clinical evaluations may be more effective in obtaining an early AS diagnosis. As a next step, our model needs to be applied to another large database in order to predict a diagnosis of AS, and, as an added verification step, confirmation of the diagnosis needs to be performed by a rheumatologist.

There were limitations with our study. As with any analysis of claims-based data, patient diagnoses may have been coded incorrectly by healthcare providers. The diagnosis of AS in our study was dependent on *ICD-9-CM* codes, which were originally developed for research purposes; however, in the USA, *ICD-9-CM* codes are often used for billing and may not reflect an accurate diagnosis. In this study, there was no independent confirmation of patient diagnoses. Additionally, age was not considered a feature in our study. As the age of onset of inflammatory back pain is typically < 45 years in patients with AS, our predictive models may perform better in a younger population.

In summary, we developed and refined predictive models for AS diagnosis in this analysis of a US administrative claims database. As Model A/B continues to be refined and optimized, its ability to distinguish patients with AS correctly from the general population will improve. The various iterations of the AS-predictive models underscore the importance of balancing big data analytics with real-world clinical observations. These predictive models may be vital for a timely diagnosis of AS despite their low PPV. Further validation of these models needs to be performed in a separate commercially insured patient population database, where a diagnosis of AS can be verified by an in-person medical assessment.

Acknowledgment Support for third-party writing assistance for this manuscript, furnished by Kheng Bekdache, PhD, of Health Interactions, Inc., was provided by Novartis Pharmaceuticals Corporation, East Hanover, NJ.

Funding This study was funded by Novartis Pharmaceuticals Corporation, East Hanover, NJ.

Compliance with ethical standards

Conflict of interest A. Deodhar has received research grants from AbbVie, Eli Lilly and Company, GSK, Janssen, Novartis, Pfizer, and UCB and has participated in advisory boards sponsored by AbbVie, Amgen, Bristol-Myers Squibb, Eli Lilly and Company, Janssen, Novartis, Pfizer, and UCB; C. Garges is currently an employee of ProMetrics, LLC, and was employed at HVH Precision Analytics, LLC, at the time of the analysis; M. Rozycki, O. Shukla, T. Arndt, and T. Grabowsky are employees of HVH Precision Analytics, LLC; Y. Park is an employee of Novartis Pharmaceuticals Corporation.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Braun J, Sieper J (2007) Ankylosing spondylitis. *Lancet* 369:1379–1390
- Rudwaleit M, Landewé R, van der Heijde D, Listing J, Brandt J, Braun J, Burgos-Vargas R, Collantes-Estevez E, Davis J, Dijkman B, Dougados M, Emery P, van der Horst-Bruinsma IE, Inman R, Khan MA, Leirisalo-Repo M, van der Linden S, Maksymowych WP, Mielants H, Olivieri I, Sturrock R, de Vlam K, Sieper J (2009) The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. *Ann Rheum Dis* 68:770–776
- van der Linden S, Valkenburg HA, Cats A (1984) Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum* 27:361–368
- Reveille JD (2011) Epidemiology of spondyloarthritis in North America. *Am J Med Sci* 341:284–286
- Feldtkeller E, Khan MA, van der Heijde D, van der Linden S, Braun J (2003) Age at disease onset and diagnosis delay in HLA-B27 negative vs positive patients with ankylosing spondylitis. *Rheumatol Int* 23:61–66
- Deodhar A, Mease PJ, Reveille JD, Curtis JR, Chen S, Malhotra K, Pangan AL (2016) Frequency of axial spondyloarthritis diagnosis among patients seen by US rheumatologists for evaluation of chronic back pain. *Arthritis Rheum* 68:1669–1676
- Deodhar A, Mittal M, Reilly P, Bao Y, Manthena S, Anderson J, Joshi A (2016) Ankylosing spondylitis diagnosis in US patients with back pain: identifying providers involved and factors associated with rheumatology referral delay. *Clin Rheumatol* 35:1769–1776
- van der Heijde D, Sieper J, Elewaut D, Deodhar A, Pangan AL, Dorr AP (2014) Referral patterns, diagnosis, and disease management of patients with axial spondyloarthritis: results of an international survey. *J Clin Rheumatol* 20:411–417
- Jordan CL, Rhon DI (2012) Differential diagnosis and management of ankylosing spondylitis masked as adhesive capsulitis: a resident’s case problem. *J Orthop Sports Phys Ther* 42:842–852
- Oostveen J, Prevo R, den Boer J, van de Laar M (1999) Early detection of sacroiliitis on magnetic resonance imaging and subsequent development of sacroiliitis on plain radiography. A prospective, longitudinal study. *J Rheumatol* 26:1953–1958

11. Danve A, Deodhar A (2015) Screening and referral for axial spondyloarthritis—need of the hour. *Clin Rheumatol* 34:987–993
12. Gamez-Nava JI, Gonzalez-Lopez L, Davis P, Suarez-Almazor ME (1998) Referral and diagnosis of common rheumatic diseases by primary care physicians. *Br J Rheumatol* 37:1215–1219
13. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G (2014) Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)* 33:1123–1131
14. Andreau-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ (2015) Big data for health. *IEEE J Biomed Health Inform* 19:1193–1208
15. Yang GZ, Leff DR (2015) Big data for precision medicine. *Engineering* 1:277–279
16. Ross BC (2014) Mutual information between discrete and continuous data sets. *PLoS One* 9:e87357
17. Curtis JR, Harrold LR, Asgari MM, Deodhar A, Salman C, Gelfand JM, Wu JJ, Herrinton LJ (2016) Diagnostic prevalence of ankylosing spondylitis using computerized health care data, 1996 to 2009: underrecognition in a US health care setting. *Perm J* 20:4–10
18. Healthcare Cost and Utilization Project (HCUP). Clinical classifications software (CCS) for ICD-9-CM. <https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>. Accessed August 8, 2018
19. Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27:861–874
20. van der Heijde D, Ramiro S, Landewé R, Baraliakos X, Van den Bosch F, Sepriano A, Regel A, Ciurea A, Dagfinrud H, Dougados M, van Gaalen F, Géher P, van der Horst-Bruinsma I, Inman RD, Jongkees M, Kiltz U, Kvien TK, Machado PM, Marzo-Ortega H, Molto A, Navarro-Compán V, Ozgocmen S, Pimentel-Santos FM, Reveille J, Rudwaleit M, Sieper J, Sampaio-Barros P, Wiek D, Braun J (2017) 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. *Ann Rheum Dis* 76:978–991
21. Brandt HC, Spiller I, Song IH, Vahldiek JL, Rudwaleit M, Sieper J (2007) Performance of referral recommendations in patients with chronic back pain and suspected axial spondyloarthritis. *Ann Rheum Dis* 66:1479–1484
22. Hermann J, Giessauf H, Schaffler G, Ofner P, Graninger W (2009) Early spondyloarthritis: usefulness of clinical screening. *Rheumatology (Oxford)* 48:812–816
23. Braun A, Saracbası E, Grifka J, Schnitker J, Braun J (2011) Identifying patients with axial spondyloarthritis in primary care: how useful are items indicative of inflammatory back pain? *Ann Rheum Dis* 70:1782–1787
24. Podubnyy D, Vahldiek J, Spiller I, Buss B, Listing J, Rudwaleit M, Sieper J (2011) Evaluation of 2 screening strategies for early identification of patients with axial spondyloarthritis in primary care. *J Rheumatol* 38:2452–2460
25. Sieper J, Srinivasan S, Zamani O, Mielants H, Choquette D, Pavelka K, Loft AG, Géher P, Danda D, Reitblat T, Cantini F, Ancuta C, Erdes S, Raffayová H, Keat A, Gaston JS, Praprotnik S, Vastesaeger N (2013) Comparison of two referral strategies for diagnosis of axial spondyloarthritis: the Recognising and Diagnosing Ankylosing Spondylitis Reliably (RADAR) study. *Ann Rheum Dis* 72:1621–1627
26. Van den Berg R, de Hooge M, van Gaalen F, Reijnierse M, Huizinga T, van der Heijde D (2013) Percentage of patients with spondyloarthritis in patients referred because of chronic back pain and performance of classification criteria: experience from the Spondyloarthritis Caught Early (SPACE) cohort [published correction appears in *rheumatology (Oxford)* 2015;54:1336]. *Rheumatology (Oxford)* 52:1492–1499
27. Ogdie A, Nowell WB, Reynolds R, Gavigan K, Venkatachalam S, de la Cruz M, Flood E, Schwartz EJ, Romero B, Park Y (2018) Diagnosis journey of patients with ankylosing spondylitis in the United States. *Ann Rheum Dis* 77:631–632
28. Kvanecz DA, Sredzinski MN, Tadlock CG (2016) Predictive analytics: a case study in machine-learning and claims databases. *Am J Pharm Benefits* 8:214–219
29. Custom Strategies: IBM government analytics forum. Putting predictive analytics to work for the army—an executive perspective <http://www.govexec.com/govexec-sponsored/2015/04/putting-predictive-analytics-work-army-executive-perspective/111406/>. Accessed August 8, 2018
30. Klein A (2018) The Washington post. Police enlist war tech in crime fight https://www.washingtonpost.com/local/police-enlist-war-tech-in-crime-fight/2013/02/18/0a9e18e2-6bc6-11e2-ada0-5ca5fa7ebe79_story.html. Accessed August 8, 2018
31. Wood C (2018) Emergency Management. How does the military use big data? <http://www.emergencymgmt.com/safety/Military-Use-Big-Data.html>. Accessed August 8, 2018
32. Carroll RJ, Eyler AE, Denny JC (2011) Naive electronic health record phenotype identification for rheumatoid arthritis. *AMIA Annu Symp Proc* 2011:189–196
33. Huang SH, LePendu P, Iyer SV, Tai-Seale M, Carrell D, Shah NH (2014) Toward personalizing treatment for depression: predicting diagnosis and severity. *J Am Med Inform Assoc* 21:1069–1075
34. Peissig PL, Rasmussen LV, Berg RL, Linneman JG, McCarty CA, Waudby C, Chen L, Denny JC, Wilke RA, Pathak J, Carrell D, Kho AN, Starren JB (2012) Importance of multi-modal approaches to effectively identify cataract cases from electronic health records. *J Am Med Inform Assoc* 19:225–234
35. Wei WQ, Leibson CL, Ransom JE, Kho AN, Caraballo PJ, Chai HS, Yawn BP, Pacheco JA, Chute CG (2012) Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus. *J Am Med Inform Assoc* 19:219–224
36. Stumpe M, Peng L (2018). Google AI blog. Assisting pathologists in detecting cancer with deep learning. <https://research.googleblog.com/2017/03/assisting-pathologists-in-detecting.html>. Accessed August 8, 2018
37. Guishan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR (2016) Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 16:2402–2410
38. Simundic AM (2009) Measures of diagnostic accuracy: basic definitions. *EJIFCC* 19:203–211

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.