RMD Open
Rheumatic & Musculoskeletal Diseases

SHORT REPORT

Validation of a machine learning approach to estimate Clinical Disease Activity Index Scores for rheumatoid arthritis

Alison K. Spencer,[1] Jigar Bandaria,[1] Michelle B. Leavy ,[2] Benjamin Gliklich,[3] Zhaohui Su,[4] Gary Curhan,[2] Costas Boussios[1]

To cite: Spencer AK, Bandaria J, Leavy MB, *et al*. Validation of a machine learning approach to estimate Clinical Disease Activity Index Scores for rheumatoid arthritis. *RMD Open* 2021;**7**:e001781. doi:10.1136/rmdopen-2021-001781

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/rmdopen-2021-001781).

Received 15 June 2021
Accepted 29 October 2021

[1]Data Science, OM1 Inc, Boston, Massachusetts, USA
[2]Research, OM1 Inc, Boston, Massachusetts, USA
[3]Research, Noble and Greenough School, Dedham, Massachusetts, USA
[4]Biostatistics, OM1 Inc, Boston, Massachusetts, USA

Correspondence to
Michelle B. Leavy;
mleavy@om1.com

ABSTRACT

Objective Disease activity measures, such as the Clinical Disease Activity Index (CDAI), are important tools for informing treatment decisions and monitoring patient outcomes in rheumatoid arthritis (RA). Yet, documentation of CDAI scores in electronic medical records and other real-world data sources is inconsistent, making it challenging to use these data for research. The purpose of this study was to validate a machine learning model to estimate CDAI scores for patients with RA using clinical notes.

Methods A machine learning model was developed to estimate CDAI score values using clinical notes from a specific rheumatology visit. Data from the OM1 RA Registry were used to create a training cohort of 56 177 encounters and a separate validation cohort of 18 726 encounters, 11 985 of which passed a model-derived confidence filter; all included encounters had both a clinician-recorded CDAI score and a clinical note. Model performance was assessed using the area under the receiver operating characteristic curve (AUC), positive predictive value (PPV) and negative predictive value (NPV), calculated using a binarised version of the outcome. The Spearman's R and Pearson's R values were also calculated.

Results The model had a PPV of 0.80, NPV of 0.84 and AUC of 0.88 when evaluating performance using the binarised version of the outcome. The model had a Spearman's R value of 0.72 and a Pearson's R value of 0.69 when evaluating performance using the continuous CDAI numeric scores.

Conclusion A machine learning model estimates CDAI scores from clinical notes with good performance. Application of the model to real-world data sets may allow estimated CDAI scores to be used for research purposes.

Key messages

What is already known about this subject?

- Composite disease activity measures are useful tools for understanding treatment response and patient outcomes in clinical studies of rheumatoid arthritis (RA).
- Scores are recorded inconsistently in real-world data sources, making it challenging to use these data sources for research.

What does this study add?

- A machine learning model can be used to estimate Clinical Disease Activity Index scores for patients with RA using unstructured clinical notes data.
- Other efforts have used machine learning models to estimate disease activity scores based on notes and laboratory values. This model uses clinical notes only, making it applicable to a larger set of patient encounters.

How might this impact on clinical practice or future developments?

- The model could be used to fill in gaps in real-world data sources, making these data more valuable for RA research.

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic inflammatory condition that can lead to physical impairment and reduced quality of life. Composite disease activity measures are useful tools for understanding treatment response and patient outcomes in clinical studies and for supporting a treat-to-target approach in clinical practice.[1–3] The Clinical Disease Activity Index (CDAI) is a validated composite disease activity measure that is recommended for use in routine clinical practice.[3] While other widely used measures, such as the Disease Activity Score with 28-joint counts (DAS 28) and the Simplified Disease Activity Index (SDAI), require laboratory tests, the CDAI is calculated using information from the physical examination combined with the Patient Global Assessment of Disease Activity and the Provider Global Assessment of Disease Activity.[4 5] As a result, the score is available immediately and can be used to inform treatment decisions during the visit.

Table 1 Demographic and clinical characteristics of training and validation cohorts

		Training cohort (n=21 062)	Validation Cohort (n=11 839)
Age, years	Mean (SD)	62.5 (13.4)	62.5 (13.4)
Sex	Female	78.5%	78.7%
	Male	21.5%	21.3%
Race	White	67.0%	67.2%
	Black	8.7%	9.0%
	Other	22.4%	22.4%
	Unknown	1.9%	1.5%
Duration of follow-up, years	Mean (SD)	5.5 (1)	5.5 (0.9)
Body mass index (BMI)*	Mean (SD)	29.8 (7.1)	29.9 (7.2)
Hypertension	n (%)	720 (3.4%)	235 (2.0%)
Type 2 diabetes	n (%)	751 (3.6%)	372 (3.1%)
Cardiovascular disease	n (%)	202 (1.0%)	83 (0.7%)

*Note, the BMI recorded on the encounter date was used. If BMI was not recorded on the encounter date, the closest recorded BMI before the encounter date was used.

While the CDAI is suitable for use in routine clinical practice, CDAI scores are not documented consistently in electronic medical records (EMRs).[6] This makes it difficult to use EMR data to support real-world research studies of RA treatment patterns and outcomes and to compare real-world outcomes to clinical trial results. While statistical methods may be applied to impute missing CDAI scores, many patients in real-world data sets do not have sufficient longitudinal CDAI scores to allow for imputation. New approaches to estimating CDAI scores at discrete time points using routinely recorded clinical data would increase the utility of real-world data sources for research.

Machine learning algorithms have been developed to estimate disease activity scores for other rheumatological conditions using clinical notes[7] and administrative data.[8] Machine learning algorithms have also been developed for use in RA to predict disease activity at the patient's next rheumatology clinic visit using structured EMR data and laboratory values[9] and to estimate disease activity scores at a specific rheumatology clinic visit using clinical notes and laboratory values.[10] These efforts demonstrate that it is feasible to use routinely recorded clinical and laboratory data to estimate and predict disease activity in RA. Laboratory data may not be available for some rheumatology clinic visits, however, making it important to estimate disease activity measures such as the CDAI using clinical notes only.

The objective of this study was to validate a machine learning model to estimate CDAI scores for RA patients using clinical notes.

METHODS

Participants

Data for this study were drawn from the OM1 RA Registry. The registry contains clinical and administrative data on over 200 900 patients with RA from rheumatology practices across the USA. Data are available from 2013 to 2021 and include clinical data extracted from EMRs (medication history and prescription information, laboratory results and diagnoses as documented by a physician), unstructured physician-documented notes, and unadjudicated medical and pharmacy claims. Disease activity scores, such as the CDAI, DAS 28 and SDAI, are available in the registry when they are documented in the EMR in structured fields. Patients are eligible for the registry if they are followed by a rheumatologist and have either multiple diagnosis codes for RA or prescriptions for a disease-modifying antirheumatic drug plus documented joint exams, RA-specific disease activity measures or RA-specific patient-reported outcomes. All registry data are deidentified.

For this study, RA registry patients with at least one clinician-recorded CDAI score and at least one text-based clinical note associated with the date of the clinician-recorded CDAI score were identified and randomly assigned to either the model training cohort (75%) or the model validation cohort (25%). Prior to random assignment, notes were stratified by score such that the

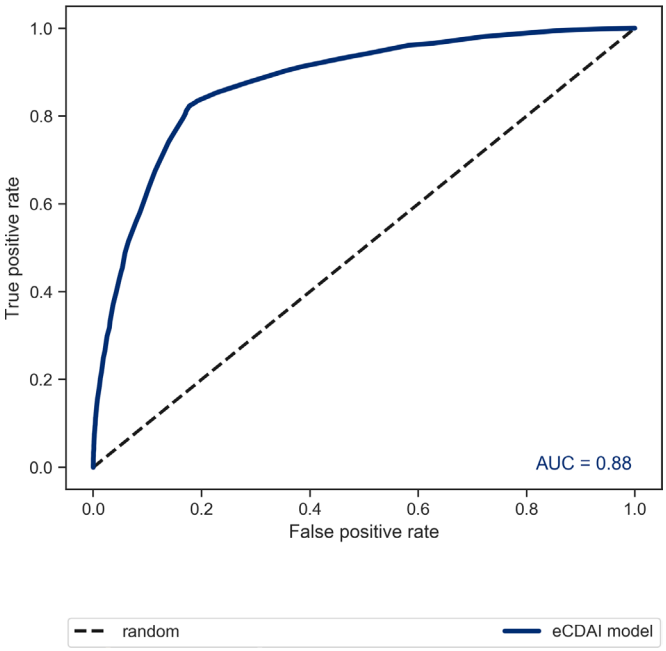

Figure 1 The AUC was calculated using a binarised version of the outcome in which the negative class is defined as those notes with CDAI scores less than or equal to 10.0 (the threshold at which CDAI scores are considered to transition from 'low' to 'moderate' disease activity), and the positive class is defined as those records with scores greater than or equal to 10.1. For this model, the AUC=0.88. AUC, area under the receiver operating characteristic curve; CDAI, Clinical Disease Activity Index; eCDAI, estimated CDAI.

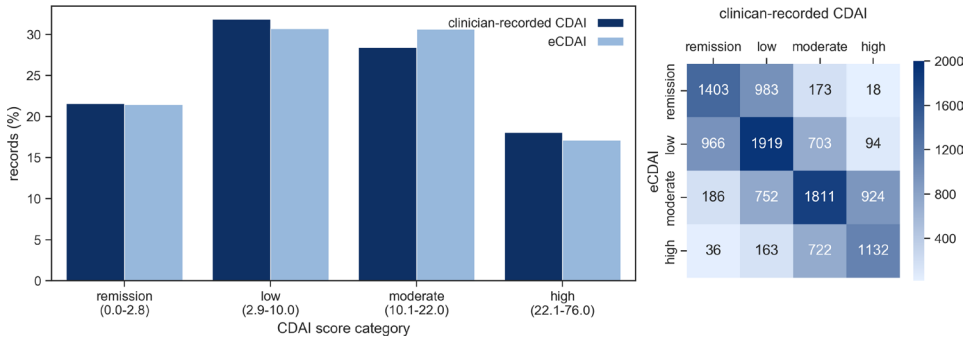

Figure 2 Distribution of categories and confusion matrix of estimated and clinician-recorded CDAI scores in validation cohort. The figure presents the distribution of eCDAI scores and the distribution of clinician-recorded CDAI scores categorised into remission (0–2.8), mild (2.9–10.0), moderate (10.1–22.0) and high disease activity (22.1–76.0) (left). The confusion matrix (right) shows the model's performance at estimating CDAI scores in each of these four categories. CDAI, Clinical Disease Activity Index; eCDAI, estimated CDAI.

score distribution across the model training and model validation cohorts was similar.

Modeling strategy

Dependent variable

The model is trained to estimate a numeric CDAI score, which can range from 0 to 76.0. The trained model generates an estimated CDAI (eCDAI) score for a specific encounter on a specific date. CDAI scores are mapped commonly to the following disease activity levels: remission (0.0–2.8); low disease activity (2.9–10.0); moderate disease activity (10.1–22.0) and high disease activity (22.1–76.0).[11]

Explanatory variables: Model features were derived from the clinical notes. Only notes that had clinical evaluation components were included to ensure that there was sufficient clinical detail to generate an eCDAI score. The explanatory terms and phrases included those relating to counts of tender or swollen joints, patient's or physician's assessment of disease activity, as well as terms that describe improvement, worsening or change in these variables.

Modelling strategy

The CDAI estimation model used in this study was an ordinal regression based multivariable model with features generated from the clinical notes. The modelling approach comprised three phases: (1) Processing the body of clinical notes to de-noise and standardise their content; (2) formation of the set of explanatory variables by identifying the terms and phrases that clinicians use to describe CDAI-related signs and symptoms and (3) development and validation of the model. Text preprocessing included standard n-gram/bag-of-words modelling approaches including tokenisation, lemmatisation and stop word removal. Variables were selected using a combination of Term frequency Inverse Document Frequency and evaluation of clinical relevance to the CDAI score. A more detailed description of these preprocessing steps can be found in a separate publication.[7] These variables were then used to train the final ordinal regression model on notes in the training cohort, and the model was validated on the validation cohort.

Model performance

The area under the receiver operating characteristic curve (AUC), positive predictive value (PPV) and negative predictive value (NPV) were used to measure the performance of the model when estimating a binary variable. For this study, we assessed performance of the model by using a binarised version of the outcome in which the negative class is defined as those notes with scores less than or equal to 10.0 (the threshold at which CDAI scores are considered to transition from 'low' to 'moderate' disease activity), and the positive class is defined as those records with scores greater than or equal to 10.1. The Spearman's R and Pearson's R values were used to evaluate the eCDAI scores vs the clinician-recorded CDAI scores on the continuous scale.

As a final step, the predictors were reviewed for clinical suitability, and the distribution of eCDAI scores was compared with the distribution of clinician-recorded CDAI scores.

RESULTS

Participants and characteristics

The model training cohort consisted of 56177 encounters from 21 062 patients with RA, while the validation cohort consisted of 18726 encounters from 11839 patients. Of the 18726 encounters, 11985 of which passed a model-derived confidence filter. Demographics and clinical characteristics of patients in the training cohort and validation cohort are presented in table 1. The median age was 62.5 years in both cohorts as of the first encounter date, and the majority of patients were female in both cohorts (78.5% and 78.7% in the training and validation cohorts, respectively).

Model performance

The model had a PPV of 0.80, NPV of 0.84, and AUC of 0.88 when evaluating performance using the binarised version of the outcome in the validation cohort (figure 1). The model

had a Spearman's R value of 0.72 and Pearson's r value of 0.69 when evaluating performance using the continuous eCDAI and CDAI.

The distribution of eCDAI scores was compared with the distribution of clinician-recorded CDAI scores categorised into remission (0–2.8), mild (2.9–10.0), moderate (10.1–22.0) and high disease activity (22.1–76.0) (figure 2).

DISCUSSION

The CDAI is a useful tool for informing treatment decisions and monitoring patient outcomes over time. Yet, capture and documentation of CDAI scores in routine clinical practice remains inconsistent, with one large study finding that only 50% of patients had a recorded disease activity score.[6] Even patients with disease activity scores documented at some visits often lack measurements for other visits, making it more difficult to use real-world data sources to study disease activity changes over time.[6] This study demonstrated that a machine learning model can use clinical notes to generate eCDAI scores for specific patient encounters. The model performed well when estimating numeric CDAI scores using clinical notes from a large, real-world data set.

Efforts to use machine learning approaches in rheumatology have increased in recent years, but there appear to be few other efforts to use these approaches to estimate disease activity scores for a specific patient encounter.[12] Other efforts to use machine learning to predict or estimate RA disease activity measures required laboratory values.[9 10] By using notes only, the approach described here is applicable to a larger set of patient encounters. While clinician-recorded CDAI scores remain the gold standard, the use of a machine learning model to generate eCDAI scores for other encounters will increase the number of CDAI scores available for retrospective analysis and may provide a more complete view of patient outcomes over time.

This approach has some limitations. The model relies on information recorded in the clinical notes, and notes with insufficient clinical detail were excluded. Second, the model was trained and validated using clinical notes from rheumatology practices in the USA. Documentation practices may vary across care settings, and further work is needed to assess the model's performance in other data sources.

Real-world data are an important source of information on RA treatments and outcomes. This study addressed the lack of clinician-recorded disease activity measures in real-world data sources by developing a model to estimate CDAI scores using unstructured clinical notes. Application of the model to real-world data sources could make these data more useful for understanding real-world treatment patterns, treatment effectiveness and patient outcomes.

Contributors AS: study concepts and design, manuscript and figure preparation, manuscript editing, final approval of the article; JB: study concepts and design, manuscript and figure preparation, manuscript editing, final approval of the article; ML: study concepts and design, manuscript and figure preparation, manuscript editing, final approval of the article; BG: study concepts and design, manuscript and figure preparation, manuscript editing, final approval of the article; ZS: study concepts and design, manuscript and figure preparation, manuscript editing, final approval of the article; GC: study concepts and design, manuscript and figure preparation, manuscript editing, final approval of the article; and CB: study concepts and design, manuscript and figure preparation, manuscript editing, final approval of the article.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests The authors indicated are employees of OM1, which is involved in issues related to the topic of this manuscript.

Patient consent for publication Not applicable.

Ethics approval This study was submitted for Institutional Review Board approval and determined to be exempt.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement This study was conducted using deidentified participant data compiled from multiple sources. Restrictions apply to the availability of these data, which were used under certain permissions for this study. Please contact the corresponding author with any questions.

ORCID iD

Michelle B. Leavy http://orcid.org/0000-0003-1927-7248

REFERENCES

1 van der Heijde DM, van 't Hof MA, van Riel PL, *et al*. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis* 1990;49:916–20.
2 Anderson J, Caplan L, Yazdany J, *et al*. Rheumatoid arthritis disease activity measures: American college of rheumatology recommendations for use in clinical practice. *Arthritis Care Res* 2012;64:640–7.
3 England BR, Tiong BK, Bergman MJ, *et al*. 2019 update of the American College of rheumatology recommended rheumatoid arthritis disease activity measures. *Arthritis Care Res* 2019;71:1540–55.
4 Gaujoux-Viala C, Mouterde G, Baillet A, *et al*. Evaluating disease activity in rheumatoid arthritis: which composite index is best? a systematic literature analysis of studies comparing the psychometric properties of the DAS, DAS28, SDAI and CDAI. *Joint Bone Spine* 2012;79:149–55.
5 Rintelen B, Sautner J, Haindl PM, *et al*. Comparison of three rheumatoid arthritis disease activity scores in clinical routine. *Scand J Rheumatol* 2009;38:336–41.
6 Gandrup J, Li J, Izadi Z, *et al*. Three quality improvement initiatives and performance of rheumatoid arthritis disease activity measures in electronic health records: results from an interrupted time series study. *Arthritis Care Res* 2020;72:283–91.
7 Alves P, Bandaria J, Leavy MB, *et al*. Validation of a machine learning approach to estimate systemic lupus erythematosus disease activity index score categories and application in a real-world dataset. *RMD Open* 2021;7:e001586.
8 Speyer CB, Li D, Guan H, *et al*. Comparison of an administrative algorithm for SLE disease severity to clinical SLE disease activity index scores. *Rheumatol Int* 2020;40:257–61.
9 Norgeot B, Glicksberg BS, Trupin L, *et al*. Assessment of a deep learning model based on electronic health record data to forecast clinical outcomes in patients with rheumatoid arthritis. *JAMA Netw Open* 2019;2:e190606.
10 Lin C, Karlson EW, Canhao H, *et al*. Automatic prediction of rheumatoid arthritis disease activity from the electronic medical records. *PLoS One* 2013;8:e69932.
11 Singh JA, Furst DE, Bharat A, *et al*. 2012 update of the 2008 American college of rheumatology recommendations for the use of disease-modifying antirheumatic drugs and biologic agents in the treatment of rheumatoid arthritis. *Arthritis Care Res* 2012;64:625–39.
12 Hügle M, Omoumi P, van Laar JM, *et al*. Applied machine learning and artificial intelligence in rheumatology. *Rheumatol Adv Pract* 2020;4:rkaa005.