Lessons Learned from the Development of Machine Learning Models to Estimate Validated Measures of Disease Activity and Symptom Severity Using Real-World Data for Four Chronic Conditions

Pedro Alves, Jonathan Gerber, Alison Spencer, Jigar Bandaria, Michelle Leavy, Stefan Weiss, Gary Curhan, Carl Marci, Jessica K. Paulus, Costas Boussios | OM1, Inc, Boston, MA, USA

Disclosures

No outside funding was received for this work. The authors are employees of OM1, which is involved in issues related to the topic of this manuscript.

Background

- Validated measures of disease activity and symptom severity are important tools to monitor disease progression and patient outcomes over time.
- For chronic conditions, these measures are often used to guide decisions related to treatment and to assess response to treatment.
- While these measures are commonly recorded in clinical trials, they are often missing or recorded at inconsistent intervals in real-world data (RWD) sources, such as electronic medical record systems (EMRs).

Objective

To assess the feasibility of developing machine learning models to estimate scores for validated disease activity and outcome measures in four chronic conditions using clinical notes from RWD sources.

Methods

- Machine learning methods were used to develop estimation models for four validated measures:
  - Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) for systemic lupus erythematosus (SLE)
  - Expanded Disability Status Scale (EDSS) for multiple sclerosis
  - Patient Health Questionnaire-9 (PHQ-9) for depression
  - New York Heart Association (NYHA) classification for heart failure
- RWD sources were used to create training and validation cohorts for each model.
- Model performance was assessed using the area under the receiver-operating-characteristic curve (AUC), calculated using a binarized version of each outcome as low versus high at clinically meaningful thresholds.

Results

- Estimated scores produced by the machine learning models were highly correlated with collected disease severity scores.
- The AUC was 0.91 for the SLEDAI model,[1] 0.91 for EDSS,[2] 0.81 for the PHQ-9, and 0.85 for NYHA (**Table 1**).
- Application of the models to patients without recorded scores in RWD sources resulted in increased numbers of patients available for RWD data studies (**Figure 1**).
- Development of the models across a range of chronic conditions resulted in several informative findings for future machine learning estimation models:
  - In each condition area, clinicians documented important information about range of symptoms, symptom severity, disease progression, and medication needs in the clinical notes. This information was sufficient for the models to generate estimated scores, even for measures that are patient-reported, such as the PHQ-9 (**Figure 2**).
  - Model features were reviewed for clinical relevance but were not restricted to features that approximated items on the validated instrument, enabling the models to consider other features, such as medication needs, resulting in improved model performance (**Figure 3**).

Figure 1. Availability of Clinician-Recorded and Estimated Scores in RWD Source

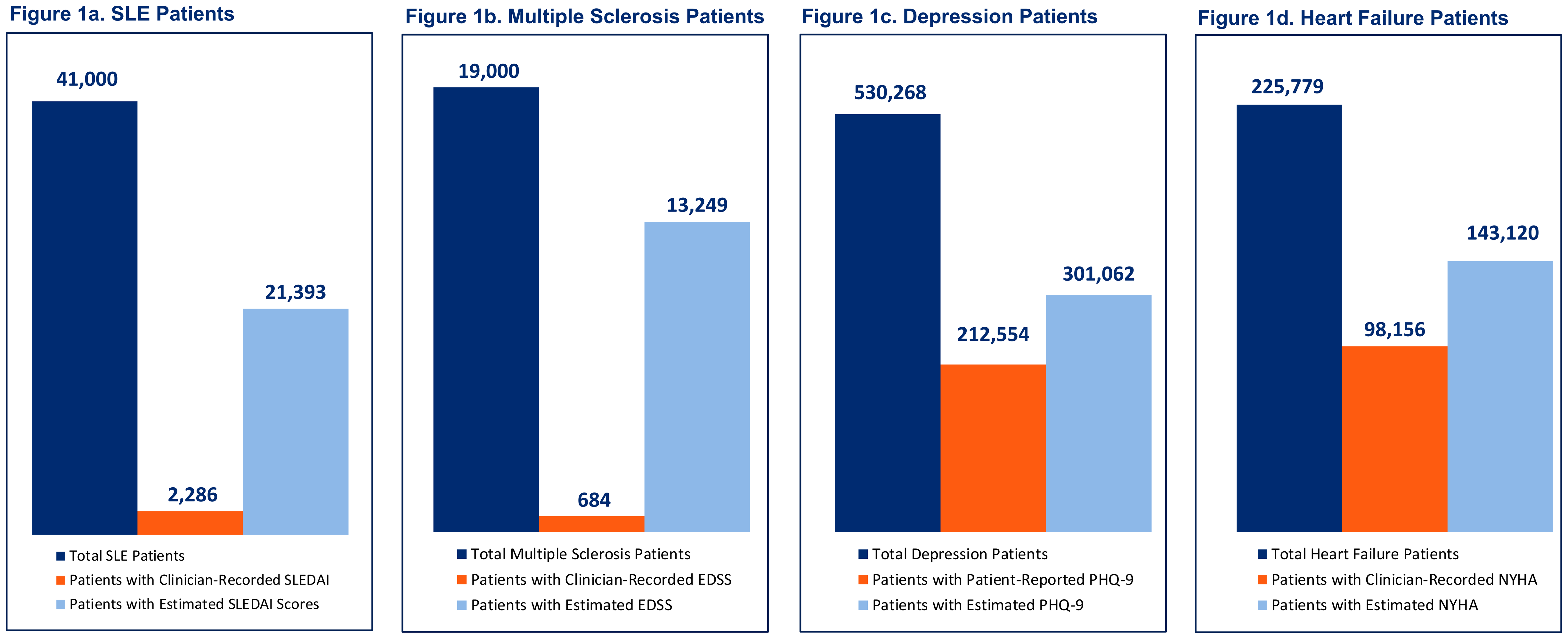

Figure 2. Example of Relevant Information for Estimating EDSS in Clinical Notes

46 yo presents for f/u for MS. She is currently on Ocrevus as DMT, last infusion was in Feb 2019. She continues to struggle with low energy and poor balance. She does take Gabapentin daily for nerve pain. To review, she has previously failed Gilenya due to elevated LFT's and Copaxone due to chest pain and nausea. Last MRI Brain was completed in March 2018 which showed several small hyperintensities involving the periventricular regions bilaterally. MRI C-Spine was showed no new lesions. Ocrevus follow up labs showed CD19/20 at 0%. No falls.

Figure 3. Examples of EDSS Model Features and Overlap with EDSS Instrument Items

Examples of EDSS Machine Learning Model Features	*Functional Systems Assessed in the EDSS*
Walker use Scooter usage Weakness	Pyramidal
Balance problems Spasms Unstable walk	Cerebellar
Speech problems	Brainstem
Pain Stiffness Edema	*Not directly assessed in EDSS*

Table 1. AUC for Estimated Validated Measures

Machine Learning Model	AUC
SLEDAI model	0.91
EDSS model	0.91
PHQ-9 model	0.81
NYHA model	0.85

AUC = area under the receiver-operating-characteristic curve; EDSS = Expanded Disability Status Scale; NYHA = New York Heart Association; PHQ-9 = Patient Health Questionnaire-9; RWD = real-world data; SLEDAI = Systemic Lupus Erythematosus Disease Activity Index

Conclusions

- Application of the models to RWD sources increases patients available for analyses of disease progression, treatment response, and patient outcomes by 1.5 to 19 times.
- Use of machine learning techniques allowed for development of models that use features that align with those in the validated instrument as well as novel features.
- Future research should focus on exploring the strengths and limitations of the models when applied to RWD studies of treatment and outcomes.

Conference

Presented at the International Conference on Pharmacoepidemiology & Therapeutic Risk Management, August 23-27, 2023. Halifax, Nova Scotia, Canada.

References

[1] 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(2):e001586..

[2] Alves P, Green E, Leavy M, et al. Validation of a machine learning approach to estimate expanded disability status scale scores for multiple sclerosis. Mult Scler J Exp Transl Clin. 2022;8(2):20552173221108635.