Replicability and Deployment of a Systemic Lupus Erythematosus Flare Prediction Model from Administrative Claims to Electronic Medical Records

Ellen Stein[a1]; Jae Min[a2]; Ebuwa Igho-Osagie[3]; Laura Evans[1]; Jeff Warnick[1]; John Doole[1]; Lucy Smigiel[1]; Arnold Chan[1]; Junqing Liu[2]; Echo Wang[b2]

[1]TriNetX, LLC, Cambridge, MA, USA; [2]BARDS Epidemiology Real-World Evidence Capability and Analytics, Merck & Co., Inc., Rahway, NJ, USA; [3]Value & Implementation Outcome Research, Merck & Co., Inc., Rahway, NJ, USA

[a]Co-first authors; [b]Corresponding author

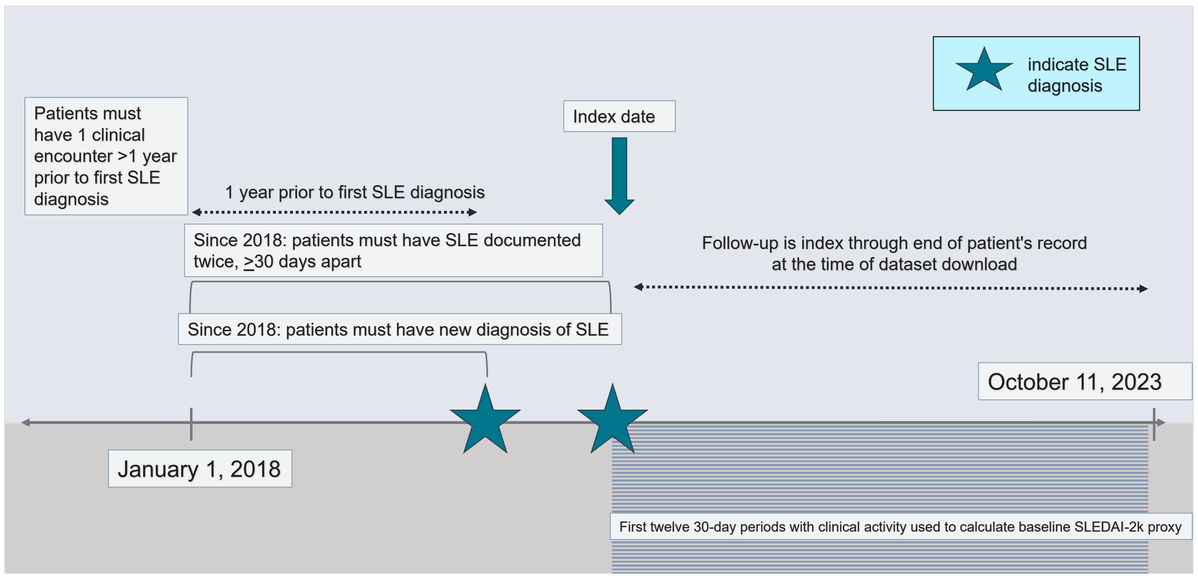
Background and objectives

- Systemic lupus erythematosus (SLE) is an autoimmune disease in which the immune system attacks its own tissues.[1] It can range in severity from mild to life-threatening, and conservative estimates suggest a prevalence of 161,000 patients with definite SLE and 322,000 with definite or probable SLE in the US[2]
- Patients experience periodic acute exacerbation of SLE symptoms, known as "flares."[1] Despite associated morbidity and mortality, flares are not codified in administrative healthcare databases. There is a need to identify its occurrence and quantify its impact on patients and the healthcare system
- Algorithms to identify flaring patients have been published, but their external validity and transportability are unknown; this study aimed to apply, validate, and improve, if necessary, a published flare identification algorithm[3] using TriNetX's Dataworks-USA Network, a large, federated, regularly updated data network containing patient electronic health record (EHR) information

Methods

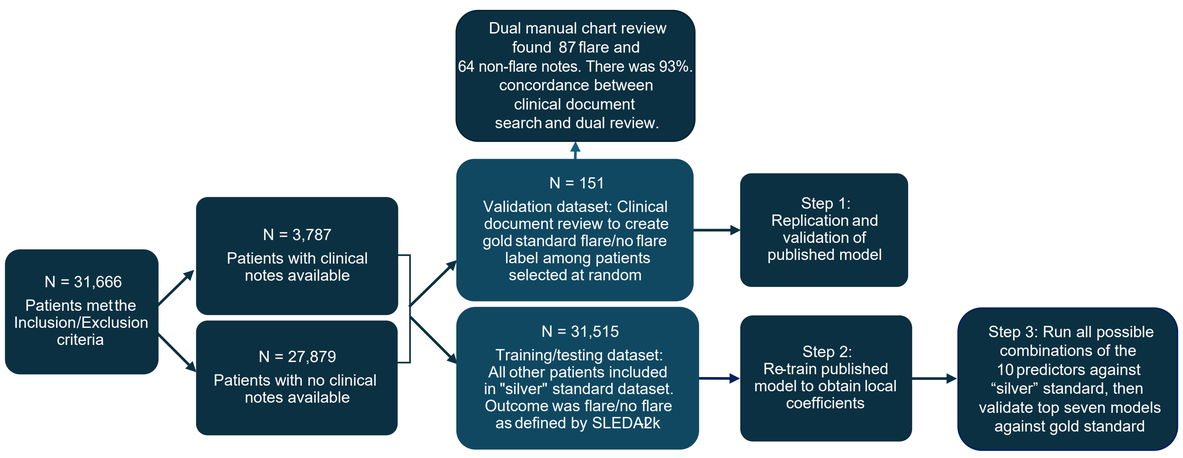
- This study used the TriNetX Dataworks-USA de-identified EHR database from 60 healthcare organizations across the US
- Study population included adult patients newly diagnosed with SLE between January 1, 2018, and October 11, 2023, with at least one historical encounter prior to diagnosis (**Figure 1**)
- A validation dataset was created by identifying 151 patients with flares and patients without flares using clinical document review. Dual clinical document review was completed to determine flare status, any discrepancies were resolved, and inter-rater reliability statistics were calculated (**Figure 2**)
- The remaining de-identified EHR data not selected for chart review was randomly divided into 80%/20% training/testing datasets
- **Modeling step 1:** The published algorithm[2] which used 10 predictors and proxy SLEDAI-2k scores as outcome was replicated. The cutoff value of being at-risk of flaring was determined using the gold standard dataset
- **Modeling step 2:** Due to the poor modeling performance, a second model was trained, where all coefficients of the published predictors were determined using the TriNetX Dataworks-USA data
- **Modeling step 3:** To further improve model performance, regression models of all possible combinations of the 10 predictors were trained and tested. Models with the best performance predicting the SLEDAI-2k and containing clinically relevant predictors were selected for validation

Figure 1. Study design[a] and process



[a]Inclusion/exclusion criteria for this project.

Figure 2. Study process flow chart



Inclusion criteria

- Patients in the validation cohort had the ICD-10 code for SLE (all child codes for M32 except M32.0) documented at least twice during the study period, at least 30 days apart. Index date was the second occurrence of an SLE diagnosis in the study period
- Patients had to be 18 years or older at the time of the index date
- Patients had no SLE diagnosis documented prior to January 1, 2018
- Patients had at least one instance of a clinical encounter documented at least one year prior to the first diagnosis of SLE in the Dataworks-USA network
- For validation dataset, patients had at least one visit chart note available from the index date through the date of the dataset download (October 22, 2023)

Exclusion criteria

- Months where the patient has no observed medical activity during the study period will be excluded from the flare/no flare classification and will not be used in the analysis
- Primary cancer diagnosis (except basal cell carcinoma or squamous cell carcinoma) within 12 months after index date

Results

- Overall, 31,666 patients met the inclusion criteria for this study
- The replicated algorithm (modeling step 1) did not yield high performance (sensitivity of 0.11, specificity of 0.97, PPV of 0.82, c-statistic of 0.54, and Brier statistic of 0.91) (**Table 1**)
- Modeling step 2 did not yield high performance (sensitivity of 0.55, specificity of 0.65, PPV of 0.67, c-statistic of 0.60, and Brier statistic of 0.53) (**Table 1**)
- In modeling step 3, the seven models with the highest performance metrics were selected for validation and yielded positive predictive value (PPV) (0.63-0.70), sensitivity (0.44-0.72), specificity (0.47-0.76), and c-statistic (0.59-0.60)
- The best-performing model selected based on highest PPV (0.70) had sensitivity of 0.44, specificity of 0.65, and c-statistic of 0.6 (**Table 1**)
- Model predictor coefficients after retraining (**steps 2** and **3**) were not comparable to those of the published model (**Table 2**)

Table 1. Gold standard validation results comparison for published paper and study results. Step 3 model represents model 384, the model found to have the highest specificity among all models produced

	Published paper[3]	Step 1: Validation only	Step 2: Model retraining and validation	Step 3: Final, improved model
Sensitivity	Not published	0.11	0.55	0.44
Specificity	Not published	0.97	0.65	0.76
PPV	Not published	0.82	0.67	0.70
c-statistic	0.75-0.76	0.54	0.60	0.60
Brier Score	0.07	0.91	0.53	0.65

Table 2. Predictor score point and coefficient values for published model, as well as models produced in modeling steps 1, 2, and 3

	Points			
	Published paper, Points[3]	Step 1: Validation only, Points	Step 2: Model retraining and validation, Points (coefficient)	Step 3: Final improved model, Points (coefficient)
Predictor				
CBC lab	2	2	10 (0.596)	13 (0.597)
ER visit	5	5	10 (0.617)	13 (0.574)
Fibromyalgia	2	2	22 (1.342)	31 (1.385)
Hypertension	2	2	13 (0.781)	18 (0.801)
IP admission	9	9	8 (0.496)	12 (0.524)
MRI	5	5	13 (0.790)	0 (0)
OP visit (each)	1	1	1 (0.061)	2 (0.075)
OP visit (once)	6	6	3 (0.213)	0 (0)
Rheum visit (each)	3	3	7 (0.420)	9 (0.418)
X-ray	2	2	0 (-0.002)	1 (0.044)
Cutoff	Not published	10	28	23

Conclusion

- Our improved models only had moderate predictive performance, possibly because the published claims-based algorithm was not fully replicable and transportable in our de-identified EHR data
- The published model may be heavily influenced by healthcare utilization patterns, given that several predictors were based on clinical encounter history. Future studies using EHR data would ideally curate more clinical features for the model
- Creating the gold standard (dual clinician review) dataset was time-consuming and expensive; thus, the sample size was relatively small
- In the model improvement phase, the data- and expert opinion-driven approach resulted in seven moderately sensitive models that could be utilized to identify patients with SLE-related flares to better understand SLE exacerbations
- Future work should consider different modeling approaches suitable for repeated events and time-to-event analyses
- While challenges remain, establishing external validity and replicability for predictive models are critical for real-world applications

References

1. CDC – Division of Population Health, National Center for Chronic Disease Prevention and Health Promotion. Systemic lupus erythematosus (SLE). https://www.cdc.gov/lupus/facts/detailed.html. Accessed October 10, 2023.
2. Helmick CG, et al. *Arthritis Rheum.* 2008;58(1):15-25.
3. Goetz I, et al. *Curr Med Res Opin.* 2022;38(9):1641-1649.
4. Gladman DD, et al. *J Rheumatol.* 2002;29(2):288-291.

Presented at ISPOR Europe; Barcelona, Spain; 17-20 November 2024