Development of Scalable, Electronic Health Record (EHR)-Based Screening for Undiagnosed Systemic Mastocytosis: PREDICT-SM

Poster Number 566

Daniel Herman,[1] Ranran Zhang,[1] Sayeda Humaira,[1] Yachana Panchal[1], Marvi Rizwan,[2] Olajumoke O. Fadugba,[2] Daniel Shaheen[3], Elizabeth Hexner[4]

[1]Department of Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; [2]Division of Pulmonary, Allergy and Critical Care , Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; [3]Blueprint Medicines Corporation, Cambridge, MA; [4]Divison of Hematology and Oncology, Department of Medicine, University of Pennsylvania, Philadelphia, PA.

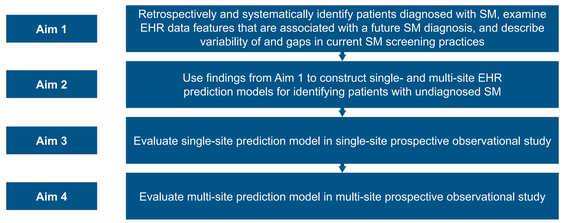
Background

- Systemic mastocytosis (SM) is a rare, clonal mast cell (MC) disease driven by the *KIT* D816V mutation (~95% cases).[1,2] Uncontrolled proliferation and activation of MCs in patients with SM can cause severe, debilitating, and unpredictable symptoms which may result in reduced quality of life and early mortality in advanced SM[2,3]
- The heterogeneity of phenotypic presentation coupled with the low specificity of symptoms spread across multiple organ systems can make diagnosing SM challenging.[4] Specifically, diagnosis can be delayed up to 7–9 years from the onset of symptoms in patients with non-advanced SM[5]
- Earlier diagnosis of SM could decrease SM-associated symptoms, improve quality of life, and decrease silent secondary organ damage. Adoption of electronic health records (EHRs) along with rapid improvement in computational methods has created opportunities to apply machine learning and artificial intelligence (AI) to clinical data from multiple health systems to learn how to identify patients with rare diseases[6–8]
- Here we describe and report the initial findings of PREDICT-SM, a study that aims to develop a pragmatic, accurate, and scalable approach to screen for undiagnosed SM by applying AI tools to EHR data

Study objectives

- PREDICT-SM comprises of 4 specific aims reflected in **Figure 1**
  - **Aim 1:** Identify patients with SM across multiple healthcare systems and describe their EHR data
  - **Aim 2:** Establish EHR prediction models for patients with undiagnosed SM
  - **Aim 3:** Prospectively evaluate single-site EHR prediction model for undiagnosed SM
  - **Aim 4:** Prospectively evaluate multi-site prediction model for undiagnosed SM

Figure 1. Overall strategy for PREDICT-SM

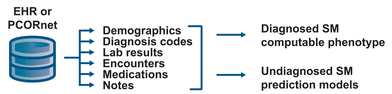


EHR, electronic health record; SM, systemic mastocytosis.

Study design

- For Aims 1 and and 2, we plan to use longitudinal EHR data from 13 health systems
- Data will be collected for patients who were ≥18 years of age as of January 1, 2022, and had outpatient encounters in primary care, dermatology, emergency, gastroenterology, or allergy/immunology clinics at least every 2 years on average
- EHR data features potentially associated with undiagnosed SM will be extracted from among a broad sampling of demographic information, diagnosis codes, problem lists, procedures, encounters with specialists, medications prescribed, laboratory test orders and results, and SM-associated symptoms and symptoms extracted from notes (**Figure 2**)

Figure 2. Key predictors extracted from EHR data



PCORnet, Patient-Centered Outcomes Research Network.

- Models will be trained using a set of patients with SM confirmed by chart review, following the 2022 World Health Organization (WHO) diagnostic and subclassification criteria for SM (**Figure 3**)[9]
- To inform model performance goals, we will perform qualitative interviews of clinical providers and patients diagnosed with or evaluated for SM

Figure 3. WHO diagnostic and subclassification criteria for SM

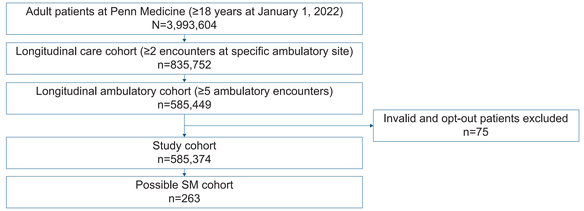
Diagnosis of systemic mastocytosis per WHO criteria requires 1 major and ≥1 minor criterion or ≥3 minor criteria
Major criterion
Multifocal dense mast cells infiltrates (≥15 mast cells in aggregates) are detected in sections of bone marrow and/or sections of other extracutaneous organ(s)
Minor criteria
In bone marrow biopsy sections or biopsy sections from other extracutaneous organs, >25% of the mast cells in the infiltrate appear spindle-shaped or have atypical morphologic features; or >25% of all mast cells in bone marrow aspirate smears are immature or have atypical features
Presence of an activating point mutation in *KIT* at codon 816 or in other critical regions of *KIT*[a] in bone marrow, blood, or another extracutaneous organ
Mast cells in bone marrow, blood, or other extracutaneous organ(s) aberrantly express at least 1 of CD2, CD25, and/or CD30 (confirmed by either flow cytometry or immunohistochemistry)
Serum total tryptase persistently >20 ng/mL (if the patient has an associated myeloid neoplasm, this parameter is not valid)

[a]Any *KIT* mutation where published solid evidence for its transforming behavior is available.
These guidelines have been used in accordance with the Terms of Use and Conditions. © International Agency for Research on Cancer. Verstovek S et. al. 2023. Systemic Mastocytosis. https://tumourclassification.iarc.who.int/chaptercontent/63/20. Date accessed February 3, 2023.

Results

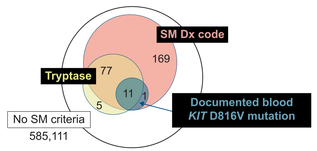
- In preliminary analyses of adult patients (age ≥18 years) from Penn Medicine as of January 1, 2022 (N=3,993,604), we identified a longitudinal ambulatory cohort (N=585,374) with ≥5 outpatient encounters including ≥2 encounters at a single primary care, dermatology, gastroenterology, or allergy/immunology practice between 2012 and 2021 (**Figure 4**)

Figure 4. Penn Medicine study patients



- The possible SM cohort included 263 patients, including 258 with ≥2 SM diagnosis codes (ICD10 D47.0*), 93 with documentation of tryptase ≥20 ng/mL or tryptase above assay upper reference limit and an SM diagnosis code, and 12 with documentation of a blood *KIT* D816V mutation (**Figure 5**)
- Blood *KIT* mutation testing was infrequently identified across this time period

Figure 5. Identification of patients with possible SM



Patients with ≥2 SM diagnostic codes (SM Dx code), elevated serum tryptase (Tryptase) and/or documented blood *KIT* D816V mutation (Documented blood *KIT* D816V positive) were included in the possible SM cohort.

- As expected, for many of the patients with possible SM, particularly those with only documentation of diagnosis codes, a diagnosis of SM was considered but either later excluded or not definitively established. We are currently reviewing these patients' medical records to ascertain which actually have been diagnosed with SM
- Demographics for the primary care cohort and patients with possible SM are shown in **Table 1**

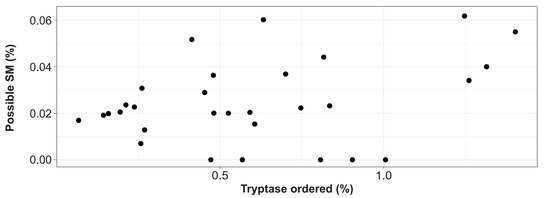
Table 1. Demographics for full EHR cohort

Characteristic, n (%)	All patients (N=585,374)	Non-SM controls (n=585,111)	Possible SM (n=263)
Gender			
Female	347,630 (59)	347,446 (59)	184 (70)
Male	237,725 (41)	237,647 (41)	78 (30)
Unknown	2 (<1)	2 (<1)	0
Non-binary	17 (<1)	16 (<1)	1 (<1)
Race			
American Indian/Alaskan Native	572 (<1)	571 (<1)	1 (<1)
Asian	22,307 (4)	22,303 (4)	4 (2)
Black	111,539 (19)	111,527 (19)	12 (5)
Multiple	5,368 (1)	5,362 (1)	6 (2)
Native Hawaiian or other Pacific Islander	584 (<1)	584 (<1)	0
Unknown	51,921 (9)	51910 (9)	11 (4)
White	393,083 (67)	392,854 (67)	229 (87)
Ethnicity			
Hispanic	20,437 (3)	20,430 (3)	7 (3)
Non-Hispanic	557,310 (95)	557,056 (95)	254 (97)
Unknown	7,627 (1)	7,625 (1)	2 (1)
Insurance			
Commercial	61,435 (10)	61,405 (10)	30 (11)
Managed care	278,000 (47)	277,851 (47)	149 (57)
Managed Medicaid	43,070 (7)	43,059 (7)	11 (4)
Managed Medicare	60,399 (10)	60,376 (10)	23 (9)
Medicaid	1,538 (<1)	1,538 (<1)	0
Medicare	120,111 (21)	120,067 (21)	44 (17)
Self-pay	65 (<1)	65 (<1)	0
Unknown	20,756 (4)	20,750 (4)	6 (2)
Tryptase ordered			
No	581,129 (99)	581,072 (99)	57 (22)
Yes	4,245 (1)	4,039 (1)	206 (78)
Tryptase result			
Median (90% CI)	4.60 (2.50–9.90)	4.60 (2.40–9.00)	8.80 (3.15–65.0)
Missing	581,484 (99)	581,417 (99)	67 (25)

CI, confidence interval.

- The possible SM cohort had a 1.9-fold higher median tryptase concentration, consistent with a mixture of both patients with SM and patients evaluated for SM but found to be unaffected
- As compared to the remainder of the longitudinal care cohort, patients in the possible SM cohort were more likely identified as female (odds ratio [OR]=1.6; P=4x10^{-4}) and white (OR=3.3; P=2x10^{-13}) (**Table 1**) patients

Figure 6. SM screening and identification across 30 primary care practices



- Across the largest 30 primary care practices, we compared the percentages of patients who had tryptase measured and who met criteria for possible SM (**Figure 6**)
- There was considerable variability in the fraction of patients with tryptase measured
- The frequency of possible SM appeared similarly variable (median=0.021%; interquartile range, 0.016–0.038%) across sites
- The relationship between tryptase and possible SM frequencies across practices was weak (R^2=0.09; P=0.06)
- It is currently unclear whether the absence of a strong relationship is due to the multiple indications for tryptase testing, the complexity of SM symptomatology, the multifactorial process of SM evaluations, or the preliminary SM cohort that contains many patients who have SM diagnosis codes but not SM

Summary

- This study highlighted the importance of the need to address observed demographic biases in patients diagnosed with SM to ensure equity of trained prediction models
- There was considerable variability in how often tryptase testing was performed, but this variability did not appear to translate into large differences in the frequency of SM diagnosis across practice sites
- Preliminary findings suggest there is potentially a substantial set of patients with SM from which to learn an EHR prediction model, which has the potential to detect SM-associated symptoms earlier, support providers in diagnosis of SM, and ultimately decrease SM morbidity and mortality
- Preliminary findings suggest underutilization of high-sensitivity peripheral blood mutational testing (*KIT* D816V)

References

1. Kristensen T et al. *Am J Hematol*. 2014;89:493–498.
2. Ungerstedt J et al. *Cancers*. 2022;14:3942.
3. Valent P et al. *Int J Mol Sci*. 2019;20:2976.
4. Valent P et al. *Blood*. 2017;129:1420–1427.
5. Jennings SV et al. *Immunol Allergy Clin North Am*. 2018;38:505–525.
6. Cohen AM et al. *PLoS ONE*. 2020. 15:e0235574.
7. La Cava W et al. *MedRxiv*. 2020.
8. Zhang L et al. J *Am Med Inform Assoc*.2020;27:119–126.
9. ©International Agency for Research on Cancer. Verstovek S et al. Systemic Mastocytosis. 2023. https:tumourclassification.iarc.who.int/chaptercontent/63/20 [beta online ahead of print; accessed February 3, 2023].

Acknowledgements

Medical writing support was provided by Kyle Wiid, Mselenge Mdegela and Jeremy Kennard, and editorial support was provided by Travis Taylor, all of Paragon, Knutsford, UK, supported by Blueprint Medicines Corporation, Cambridge, MA, according to Good Publication Practice guidelines.

Disclosures

This study was supported by Blueprint Medicines Corporation, Cambridge, MA. Blueprint Medicines Corporation reviewed and provided feedback on the poster. The authors had full editorial control of the poster and provided their final approval of all content. D Herman has participated in an advisory board for Blueprint Medicines Corporation. Full disclosures for all authors are available upon request at medinfo@blueprintmedicines.com.

SCAN ME