# Machine learning application to find lower-risk myelodysplastic syndrome patients from real-world data

Colden Johanson, Hu T. Huang, Danny Idryo, Ronda Broome, Matthew Rioth, Rayna Matsuno

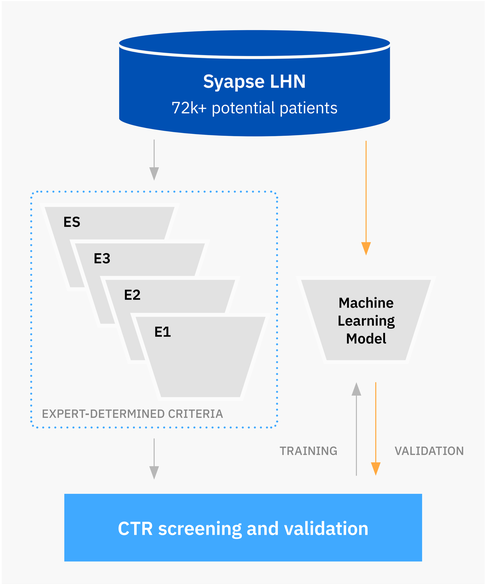
2022 ASCO Annual Meeting
*June 3–7, 2022*
*Abstract Number: #1555*

## BACKGROUND

- Finding patients with myelodysplastic syndrome (MDS) is a known challenge
- Current claims-based approaches have not been validated against an expert reference standard
- Manual screening by certified tumor registrars (CTRs) is expensive and time consuming, so efficient target criteria is essential

## METHODS

- Manually targeted patients based on demographics, diagnoses, medications, labs, and procedures
- Screened a sample of 1,549 patients from the Syapse Learning Health Network™ (LHN) and identified eligible patients
- Trained machine learning (ML) models on the screened patient data and selected the optimal model for patient identification




| Strategies | Description |
|---|---|
| E1 | Two MDS diagnosis* dates ≥ 90 days apart + evidence of ESA treatment§† |
| E2 | MDS Registry evidence by ICD-O-3 Histology‡ |
| E3 | Patients with suspected MDS based on manual review of clinician notes |
| E4 | Other small sample attempts |

* MDS ICD-10 codes: C94.6 and D46
‡ MDS ICD-O-3 Histology: 9980, 9981, 9982, 9983, 9984, 9985, 9986, 9987, 9988, 9989, 9993
§ ESA treatment: darbepoetin alfa or epoetin alfa
† The first code fell within the study window of 2016-01-01 to 2019-06-30

## Machine learning models may improve the efficiency of identifying subpopulations of patients with myelodysplastic syndrome

> The ML model improved the rate of ESA-treated MDS patient identification, even after the expertly-determined population was depleted

> The ML model identified additional patients that were missed by the expert-determined inclusion criteria

CORRESPONDENCE
**Colden Johanson**
cole.johanson@syapse.com

Copies of this presentation obtained through Quick Response (QR) Code are for personal use only and may not be reproduced without permission from ASCO® or the author of this presentation.

## RESULTS

- Patient selection strategies based on targeted, expertly-determined inclusion criteria were effective, but limited

**FIGURE 1** Patient matching rate comparison among different strategies

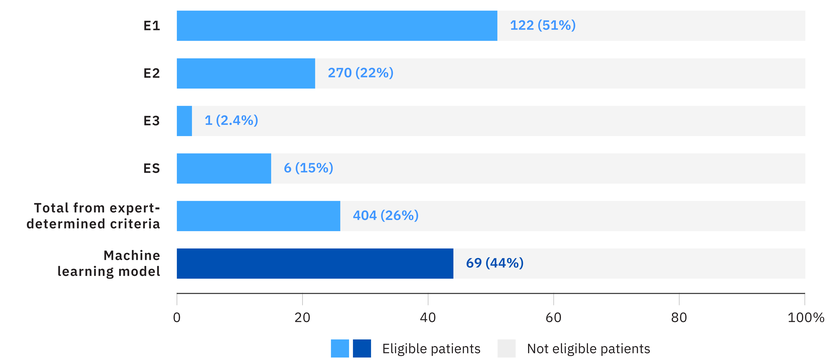
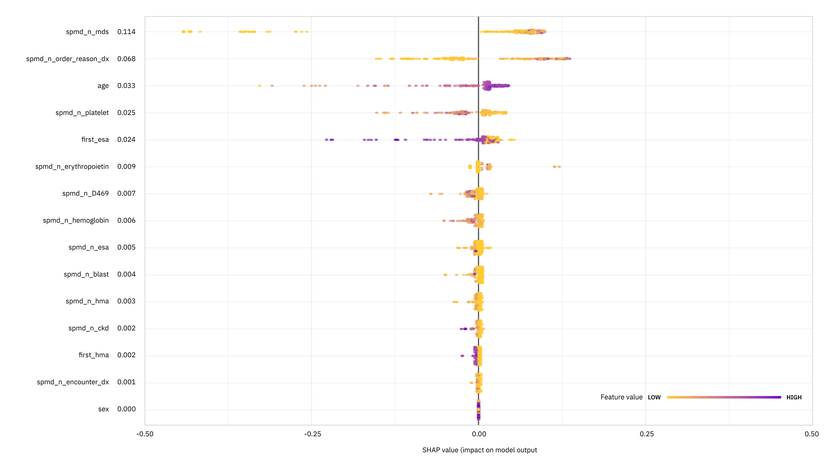



**FIGURE 2** Variable importance for the patient identification




- The machine learning model successfully identified additional patients with improved efficiency