Machine Learning Approach to Understand Real-World Treatment in Patients With Higher-Risk Myelodysplastic Syndromes

Vandana Priya,[1] Vivek P Vaidya,[1] Smita Agrawal,[1] Neeraj Singh,[1] Kaveri Chatra,[1] Dhaval Parmar,[1] Raymond Yan,[1] Rahul K Das,[1] Mahnoush Haririfar,[1] Peter McMahon,[2] Mellissa Williamson,[2] Islam Sadek,[2] Cosmina Hogea[2]

[1]ConcertAI, Cambridge, MA; [2]Gilead Sciences, Inc., Foster City, CA

5100

Poster Session Online

American Society of Hematology
Helping hematologists conquer blood diseases worldwide

903. Health Services and Quality Improvement -Myeloid Malignancies: Poster III
Vivek Vaidya

MON-5100

ASH2023

Conclusions

We have built a family of patient data-driven machine learning (ML) models for patients with Higher-risk myelodysplastic syndrome (HR-MDS), based on data collected in routine clinical practice, towards predicting time-to and probability-of first line systemic therapy with reasonable accuracy.

Most patients diagnosed with HR-MDS in this study appeared to receive delayed or no systemic therapy in their treatment journey. To bridge gaps in care and facilitate timely treatment in eligible patients, data-driven predictive modeling that considers detailed patient-level factors is critical.

Further analysis on outcomes of untreated patients linked with their key clinical characteristics identified using these models may provide opportunities to improve clinical care by implementing strategies for treatments based on patient-level factors

Plain Language Summary

In this research, < half of the people with HR-MDS were over 73 years old. A majority were men, receiving care in community setting; more than half of the patients had lower-than-normal hemoglobin levels.

Less than half of all patients got active treatment. For half of those who did get treated, it took more than 4 months to start.

Things like changes in blood levels (white blood cells, red blood cells, and platelets) and other health measures (like glucose, performance status, weight, and systolic pressure) can affect whether a patient gets treatment or not.

Introduction

- HR-MDS is a severe and aggressive form of MDS that represents a group of rare and heterogeneous hematologic disorders that primarily affect the bone marrow and blood cells.
- Treatment typically involves a combination of supportive care and active treatment with targeted therapies, hypomethylating agents, chemotherapy, and hematopoietic stem cell transplant.
- In this study, we have use of ML models to identify clinical and demographic features that may impact how patients receive systemic therapies.

References
1. Chen, Tianqi, et al. *R package version 0.4-2* 1.4 (2015): 1-4. **2.** Harrell Jr, Frank E., et al. *Statistics in medicine* 3.2 (1984): 143-152. **3.** Akaike, H. (1998). In Selected papers of hirotugu akaike (pp. 199-213). New York, NY: Springer New York.

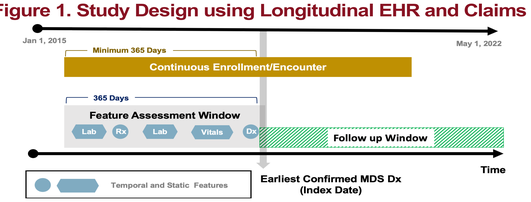
Acknowledgments: We extend our thanks to the patients, their families, and all participating investigators. This study was funded by Gilead Sciences, Inc.

Disclosures: **PM**, **MW**, **IS**, and **CH** are employees of Gilead Sciences, Inc. **VP**, **VPV**, **SA**, **NS**, **KC**, **DP**, **RY**, **RKD**, and **MH** are employees of ConcertAI, Inc. ConcertAI received funding from Gilead to conduct this study.

Correspondence: Cosmina Hogea, Cosmina.Hogea@gilead.com

Methods

Figure 1. Study Design using Longitudinal EHR and Claims



Modeling approach: eXtreme Gradient Boosting (XGBoost [1]) with Hazard Cox regression objective function.

Model Validation: 5-fold cross-validation cross-validation (data was split 5 ways; for each fold, a different 20% is left out for validation, while 80% of the data was used to train the model).

Model Population: 821 patients aged ≥18 years with HR-MDS diagnosed between January 2015 and April 2022, using ConcertAI's RWD360™ database linked with open claims data. RWD360 consists of structured records from US-based oncology electronic health record systems.

Model Design: Demographic and clinical data up to 1 year prior to the date of HR-MDS diagnosis were used to develop a family of XGBoost-based ML models to investigate drivers of treatment. We created four models: a time-to-therapy model that predicts patient-level hazard ratios and four classification models that predict likelihood of therapy in the first 3, 6, and 12 months, respectively, following HR-MDS diagnosis (therapy yes/no models).

Feature Design: To better capture the temporal characteristics of a patient's disease state evolution, clinical features were organized into time windows alongside the variations within them. We used recursive feature elimination to bring the total features used in the model to 50 to enhance the interpretability and compactness.

Model performance: Harrell's concordance index [1] (C-index) and Akaike information criterion [3] (AIC) for the time-to-therapy model and area under curve (AUC) for the therapy yes/no models.

Results

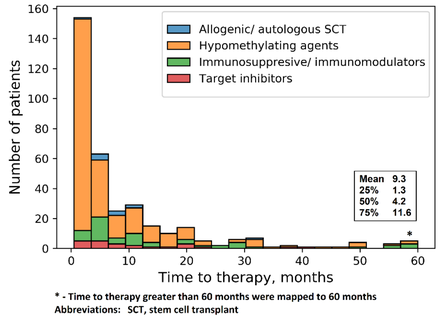
Table 1. Characteristics of HR-MDS patient cohort (N=821)

	Category	Numbers	%
# of Patients in HR-MDS cohort		821	100
Age	>=60	718	87.45
	<60	103	12.55
Race	White	601	73.2
	Non-white	204	24.84
	Unknown	16	1.94
Ethnicity	Not Hispanic or Latino	584	71.13
	Hispanic or Latino	221	26.91
	Unknown	16	1.94
Gender	Male	490	59.68
	Female	315	38.36
	Unknown	16	1.94
# of Patients where systemic drugs not taken		472	57.49
# of Patients where systemic drugs taken		349	42.51
	Systemic Drug Classification		
	Target Inhibitors	23	6.59
	Allogenic/ Autologous SCT	11	3.15
	Immunosuppresive/ Immunomodulators	55	15.76
	Hypomethylating Agents	259	74.21

Results (continued)

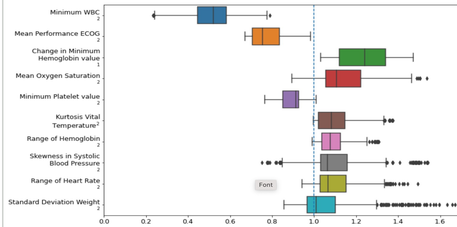
- The median age in this patient cohort was 73 years (range = 21–88 years); approximately 60% were male, and 66% patients received care in community setting.
- Median baseline hemoglobin concentration of 9.4 g/dL (range = 4.4–18.9).
- Median baseline Charlson Comorbidity Index of 1 (range = 0–10).
- Median baseline polypharmacy of 10 (based on patients who had at least 1 concomitant medication (49%)).
- Only 42.5% patients had evidence of systemic active therapy, and of those, the median time to therapy initiation was 4.2 months.

Figure 2. Distribution of Time to Systemic Therapy in HR-MDS Cohort Stratified by Therapy Type



* - Time to therapy greater than 60 months were mapped to 60 months
Abbreviations: SCT, stem cell transplant

Figure 3. Features for 'Time To Therapy' Model



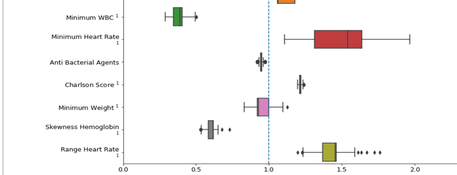
Predictors	Mean Hazard Ratio (25%, 75%)
Minimum WBC [2]	**0.522 (0.566, 0.582)**
ECOG [2]	**0.754 (0.836, 0.836)**
Change in Minimum Hemoglobin value [1]	**1.241 (1.337, 1.34)**
Mean Oxygen Saturation [2]	1.105 (1.189, 1.221)
Minimum Platelet value [2]	0.912 (0.923, 0.926)
Kurtosis Vital Temperature [2]	1.080 (1.14, 1.146)
Range of Hemoglobin [2]	1.077 (1.109, 1.126)
Skewness in Systolic Blood Pressure [2]	1.062 (1.124, 1.157)
Range of Heart Rate [2]	1.065 (1.132, 1.153)
Standard Deviation Weight [2]	1.011 (1.067, 1.1)

[1] From 6-12 months before to 30 days of MDS Dx
[2] Within 30 Days of MDS Dx

Hazard Ratio below 1: Feature is Protective
Hazard Ratio above 1: Feature is Risk Factor

Low Minimum WBCs, ECOG and Platelets key predictors for decreased time to initiate therapy. **Skewness features** suggest that **asymmetries or sudden shifts** within 30 days of Dx influence time to therapy

Figure 4. Features for 'Therapy within 3 Months of DX' Model



Predictors	Median Hazard Ratio (25%, 75%)
Slope Glucose Serum Plasma [2]	**1.322 (1.281, 1.323)**
Standard Deviation Weight [1]	1.064 (1.06, 1.179)
Minimum WBC [1]	**0.388 (0.346, 0.406)**
Minimum Heart Rate [1]	**1.541 (1.31, 1.638)**
Anti Bacterial Agents [1]	0.95 (0.942, 0.951)
Charlson Score	**1.21 (1.21, 1.22**
Minimum Weight [1]	0.926 (0.918, 1.0)
Skewness Hemoglobin [1]	**0.616 (0.585, 0.619)**
Range Heart Rate [1]	**1.455 (1.37, 1.461)**

[1] Within 30 Days of MDS Dx
[2] From 6-12 months before to 30 days of MDS Dx

Hazard Ratio below 1: Feature is Protective
Hazard Ratio above 1: Feature is Risk Factor

Low WBCs, falling Glucose Serum Plasma, and a high minimum Heart Rate key predictors for increased chance of therapy within 3 months of Dx

Table 2. Model performance and top features

Model feature analysis indicates fluctuation in blood parameters (WBC, hemoglobin, and platelets) as well as non-blood measures (glucose, ECOG, weight, and systolic pressure) can forecast how long it will take for an HR-MDS patient to receive therapy.

Model Outcome	Performance	Top 3 Features and impact on model prediction
Time to Active Therapy	C-Index = 0.68 AIC = 3314.2	**Time to therapy prediction shortens when**: 1. Minimum value of WBC count within 30 days of DX is Low 2. Mean of ECOG values observed within 30 days of DX is Low 3. Rate of fall in Hemoglobin during the 180 to 30 days before Dx is High
Active Therapy received within 3 months of Dx	AUC = 0.72	**Likelihood of therapy rises when**: 1. Rate of fall in measurements of blood glucose during the 180 to 30 days before Dx is Low 2. Standard deviation of weight within 30 days of Dx is High 3. Minimum WBC count within 30 days of Dx is Low
Active Therapy received within 6 months of Dx	AUC = 0.7	**Likelihood of therapy rises when**: 1. Minimum WBC count within 30 days of Dx is Low 2. Rate of fall in measurements of blood Glucose during the 180 to 30 days before Dx is High 3. Mean of systolic blood pressure within 30 days of Dx is High
Active Therapy received within 12 months of Dx	AUC = 0.72	**Likelihood of therapy rises when**: 1. Minimum WBC count within 30 days of Dx is Low 2. Mean Platelet count within 30 days of Dx is Low 3. Mean of ECOG values observed within 30 days of Dx is Low

AIC, Akaike information criterion; AUC, area under the curve; C-index, Harrell's concordance index; Dx, diagnosis; ECOG, Eastern Cooperative Oncology Group; WBC, white blood cell.