Patient and Caregiver Insights from Social Media into the Disease Burden of Myelodysplastic Syndrome With A Sub-Cohort View of High-risk Patients

Rohit Marwah, Subrat Mishra, Ben Gross, Sandra Couturiaux, Rico Calara,
Eduardo Jose Sabate Estrella, Cosmina Hogea

Submitted to: Journal of Medical Internet Research
on: August 15, 2024

Table of Contents

Original Manuscript........ 5

Patient and Caregiver Insights from Social Media into the Disease Burden of Myelodysplastic Syndrome With A Sub-Cohort View of High-risk Patients

Rohit Marwah[1*] MTECH; Subrat Mishra[1*] BTECH; Ben Gross[1*] PHD; Sandra Couturiaux[1*] MBA; Rico Calara[2*] phd; Eduardo Jose Sabate Estrella[2*] phd; Cosmina Hogea[2*] phd

[1]Definitive Healthcare Framingham US
[2]Gilead Sciences Inc Foster City US
[*]these authors contributed equally

Corresponding Author:
Rohit Marwah MTECH
Definitive Healthcare
492 Old Connecticut Path Suite 401
Framingham
US

Abstract

Background: Social media platforms offer valuable insights into the patient's experience, revealing organic conversations that reflect their immediate concerns and needs. Through active listening to lived experiences, we can identify unmet needs and discover real-world challenges patients and caregivers face.

Objective: This study aimed to develop a reusable framework to collect and analyze evolving social media data, capturing insights into the experiences of individuals with MDS and higher-risk myelodysplastic syndromes (HR-MDS) and their caregivers. The findings can inform the development of appropriate patient support interventions.

Methods: We conducted an extensive Google search to identify social posts of interest using validated URLs and keywords on English-language websites relevant to MDS. The search covered the period from 1/1/2008 to 12/31/2022. We utilized scraping algorithms to collect, clean, and standardize pertinent information. To classify the perspective of each experience as either that of a patient or caregiver, we employed classification algorithms. This involved contextualizing and summarizing all user posts, followed by decision tree tagging to assign them to the patient or caregiver category. Advanced algorithms were employed to analyze the semantic and temporal structure of the data. Patients or caregivers were categorized as HR-MDS based on contextual mentions of high-risk in their posts or specific factors aligned with NCCN guidelines (e.g., blast percentage, transplantation, use of high-intensity chemotherapy or hypomethylating agents, or disease progression). Each post was assigned major themes and sentiments using a supervised classification machine learning model. Additionally, we employed a semi-supervised machine learning approach for the identification of latent themes in the data corpus.

Results: The data collected comprised approximately 5.5 million words from 42,000 posts across 5,500 threads, involving about 4,000 users predominantly from the US, UK, and Canada. Out of the 1,249 users classified as HR-MDS, 588 (47%) were patients and 661 (53%) were caregivers. Among the HR-MDS users, the predominant sentiments included concern (78%), anxiety (60%), frustration (58%), fear (58%), and confusion (49%). Concern was the predominant sentiment expressed by caregivers (n=971, 59%), and anxiety by patients (n=752, 55%). Common concerns were specifically related to blood counts (n=677, 54%), burden of the disease (43%), QoL (36%), available treatment options and effectiveness (31%), and disease progression and prognosis (31%). Anxiety related to health and disease (48%), treatment (26%), and the diagnostic process (20%) were also common. The most common sentiments related to fear were the potential development of health complications and the manifestation of symptoms (19%) and the progression and exacerbation of MDS (19%). Additionally, confusion was pervasive among participants, with 295 (24%) individuals finding it challenging to comprehend the nuances of MDS and its diagnosis. A systematic analysis of the principal domains for which information is being sought about HR-MDS revealed frequent mention amongst users of acquiring information on therapeutic intervention (19%), and an interest in ongoing research associated with the disease (17%)

Conclusions: The application of sophisticated NLP techniques demonstrates promise in effectively identifying the emerging complex themes and sentiments experienced by HR-MDS users, thereby highlighting the unmet needs, barriers, and facilitators associated with the disease. Clinical Trial: NA

(JMIR Preprints 15/08/2024:65460)
DOI: https://doi.org/10.2196/preprints.65460

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

Marwah R[1], Mishra S[1], Gross B[1], Couturiaux S[1], Calara R[2], Sabate E[2], Hogea C[2]
[1]Definitive Healthcare, Framingham, MA, USA, [2]Gilead Sciences, Inc., Foster City, CA, USA

Patient and Caregiver Insights from Social Media into the Disease Burden of Myelodysplastic Syndrome With A Sub-Cohort View of High-risk Patients

Background: Social media platforms offer valuable insights into the patient's experience, revealing organic conversations that reflect their immediate concerns and needs. Through active listening to lived experiences, we can identify unmet needs and discover real-world challenges patients and caregivers face. This study aimed to develop a reusable framework to collect and analyze evolving social media data, capturing insights into the experiences of individuals with MDS and higher-risk myelodysplastic syndromes (HR-MDS) and their caregivers. The findings can inform the development of appropriate patient support interventions.

Methods: We conducted an extensive Google search to identify social posts of interest using validated URLs and keywords on English-language websites relevant to MDS. The search covered the period from 1/1/2008 to 12/31/2022. We utilized scraping algorithms to collect, clean, and standardize pertinent information. To classify the perspective of each experience as either that of a patient or caregiver, we employed classification algorithms. This involved contextualizing and summarizing all user posts, followed by decision tree tagging to assign them to the patient or caregiver category. Advanced algorithms were employed to analyze the semantic and temporal structure of the data. Patients or caregivers were categorized as HR-MDS based on contextual mentions of high-risk in their posts or specific factors aligned with NCCN guidelines (e.g., blast percentage, transplantation, use of high-intensity chemotherapy or hypomethylating agents, or disease progression). Each post was assigned major themes and sentiments using a supervised classification machine learning model. Additionally, we employed a semi-supervised machine learning approach for the identification of latent themes in the data corpus.

Results: The data collected comprised approximately 5.5 million words from 42,000 posts across 5,500 threads, involving about 4,000 users predominantly from the US, UK, and Canada. Out of the 1,249 users classified as HR-MDS, 588 (47%) were patients and 661 (53%) were caregivers. Among the HR-MDS users, the predominant sentiments included concern (78%), anxiety (60%), frustration (58%), fear (58%), and confusion (49%). Concern was the predominant sentiment expressed by caregivers (n=971, 59%), and anxiety by patients (n=752, 55%) (Figure 1). Common concerns were specifically related to blood counts (n=677, 54%), burden of the disease (43%), QoL (36%), available treatment options and effectiveness (31%), and disease progression and prognosis (31%). Anxiety related to health and disease (48%), treatment (26%), and the diagnostic process (20%) were also common. The most common sentiments related to fear were the potential development of health complications and the manifestation of symptoms (19%) and the progression and exacerbation of MDS (19%). Additionally, confusion was pervasive among participants, with 295 (24%) individuals finding it challenging to comprehend the nuances of MDS and its diagnosis. A systematic analysis of the principal domains for which information is being sought about HR-MDS revealed frequent mention amongst users of acquiring information on therapeutic intervention (19%), and an interest in ongoing research associated with the disease (17%) (Figure 2).

Conclusion: The application of sophisticated NLP techniques demonstrates promise in effectively identifying the emerging complex themes and sentiments experienced by HR-MDS users, thereby highlighting the unmet needs, barriers, and facilitators associated with the disease.

Keywords: Myelodysplastic Syndrome, Bone marrow failure syndromes, Artificial intelligence, Machine Learning, Natural Language Processing

Introduction

Myelodysplastic syndromes (MDS) are a diverse group of hematologic malignancies characterized by clonal hematopoiesis and one or more cytopenia. The clinical course of MDS and the annual risk of progression to acute myeloid leukemia (AML) is highly variable[1]. Guidelines by the National Comprehensive Cancer Network (NCCN) [2] and the European Society for Medical Oncology (ESMO) MDS guidelines [3] outline risk stratification at diagnosis, with older guidelines recommending the use of the International Prognostic Scoring System – Revised [4]. Increasingly, risk stratification systems that consider molecular markers are being used. The International Prognostic Scoring System – Molecular (IPSS-M) [5] is increasingly used in centers of excellence.

High-intensity treatment for patients with a high-risk or very high-risk stratification has been shown to prolong overall and leukemia-free survival [6] and [7]. These treatments include allogeneic hematopoietic cell transplantation (a-HCT) if eligible [8], intensive chemotherapy [9], or targeted therapies in a small subset of patients where targetable genetic features are present [10]. The decision to initiate any high-intensity treatment is based on the medical fitness and frailty of the patient [9]. While there are some objective inputs into the overall determination of fitness and frailty, for example, the ECOG score [11] and Charlson's comorbidity index [12], the final determination is subjective.

NCCN guidelines recommend shared decision-making for a treatment plan among providers, patient, and family members in high-risk patients based on patient goals. Key to shared decision-making is a patient well informed about their risk status, prognosis, treatment options, and treatment risks.

Social media platforms are an increasingly rich source of insight into the broader patient experience as they manage illness [13] and [14]. These unfiltered, unsolicited, and genuine dialogues demonstrate what is foremost on the mind of the patient or their caregiver at any given moment, and over time, provide the specific concerns and challenges throughout their illness.

Recent work has looked at general themes of the patient and caregiver experience in MDS, but none have focused on the specific experiences of high-risk MDS patients.

Two social media listening approaches have been recently reported for patients with MDS; [15] and [16]

Booth et al. 2019 studied the disease and treatment experiences of 347 patients with AML or MDS who were ineligible for intensive chemotherapy. The primary objective was to "identify the most important factors that drive patients' treatment decisions by identifying their priorities and reasons for pursuing certain types of treatments over others, including preferences in end-of-life situations." Qualitative data analysis was performed manually to categorize themes and determine if the sample size was sufficient for theme analysis. Major themes were segmented into five categories: Humanistic, Treatment Decision, Unmet needs, Life Milestones, and Economic Burden.

While this work did not specifically identify risk status and performed a qualitative analysis of user posts, the authors reached the following important conclusions: patients felt that there was a lack of treatment options, and that age alone should be a reason to not receive a particular treatment. The authors also found a clear theme around a general lack of information and knowledge regarding MDS/AML and a lack of information about treatment options, which resulted in the feeling that the patients and caregivers could not make well-informed decisions.

Frank et al. 2022 searched and mined over 20,000 comments from publicly available, country-specific online forums from 2011 to 2019. The primary objective was to identify and categorize unmet needs of patients with MDS and their caregivers in the USA, the UK, Spain, Canada, France, and China. This study used a proprietary natural language processing (NLP) algorithm developed by NetBase Quid to identify keywords and subsequently used an artificial intelligence (AI) engine to identify the associated semantic context. The AI algorithm analyzed posts to identify common topics discussed online, which were then grouped into themes, identifying patterns and commonalities across posts to establish a “network” of discussion topics and how related certain topics and themes are across users and posts.

Our work also found that patients in the USA, UK, and Canada expressed negativity at the time of diagnosis related “to experiences that highlighted educational and awareness gaps of their doctors and uncertainty of treatment outcomes due to individual disease variability. Negative patient experiences included general inexperience with MDS and HR-MDS of non-specialist physicians (inconsistency in diagnosis and treatment options presented compared with specialist physicians at MDS CoE), conflicting treatment recommendations, and unclear information about diagnosis and testing.”

Using disease-specific social media posts of MDS and HR-MDS patients and their caregivers, our study captured the factors, emotions, and milestone events considered most important concerning the disease journey and thereby uncovered structural evidence to help inform and characterize those perspectives through an analytical framework.

The primary objectives of our study were to:

1. Gather, analyze, and inform on the patient journey insights, gaps, and burden of an MDS and an HR-MDS patient and caregiver from an organic lens of social media, to
1. Construct a psychographic thematic representation of themes & sentiments and decipher specific themes and insights from posts that mention high risk either explicitly or by intrinsic methods that we deploy on posts.

Methods

Data Sources

This study is a descriptive exploratory analysis of publicly available posts related to MDS and HR-MDS patient and caregiver experiences in social media and blog sites described in Section 3.4 from August 2008 through November 2022. The start of the study period was selected because it corresponds to the label expansion of azacitidine to treat high-risk MDS patients. A preliminary analysis of cumulative posts over time was performed to confirm that there was an adequate number of posts during this time period to complete the study objectives.

Google searches was the initial screening tool utilized in an exploratory analysis to identify potential data sources for this study. Validation of sites to ensure they were legitimate is described in section 3.6. Search phrases carefully designed to capture holistically MDS patient and caregiver experiences were executed on Google Search results collected and analyzed to build a complete and vetted list of potential data sources. An exhaustive list of search phrases is provided below in Table 1. Website relevance was also considered to make sure information was collected from reliable sources and is detailed in section 3.8

Table 1 **List of keywords and key search terms**

caring for MDS
caring for myelodysplastic Syndrome
high risk MDS (story or experience)
high risk myelodysplastic syndrome (story or experience)
I survived MDS
I survived myelodysplastic Syndrome
living with MDS
living with myelodysplastic Syndrome
MDS advocacy group
MDS aplastic anemia
MDS blog
MDS bone marrow transplant (story or experience)
MDS caregiver advocacy group
MDS caregiver blog
MDS caregiver discussion
MDS caregiver experience
MDS caregiver forum
MDS caregiver journey
MDS caregiver story
MDS caregiver support group
MDS discussion
MDS experience
MDS forum
MDS journey
MDS patient advocacy group
MDS patient blog
MDS patient discussion
MDS patient experience
MDS patient forum
MDS patient journey
MDS patient story

MDS patient support group
MDS stem cell therapy (story or experience)
MDS story
MDS support group
MDS survivor advocacy group
MDS survivor blog
MDS survivor discussion
MDS survivor experience
MDS survivor forum
MDS survivor journey
MDS survivor story
MDS survivor support group
MDS treatment advocacy group
MDS treatment blog
MDS treatment discussion
MDS treatment experience
MDS treatment forum
MDS treatment journey
MDS treatment story
MDS treatment support group
myelodysplastic syndrome advocacy group
myelodysplastic syndrome aplastic anemia
MDS transfusion (story or experience)
myelodysplastic syndrome transfusion (story or experience)
myelodysplastic syndrome blog
myelodysplastic syndrome bone marrow transplant (story or experience)
myelodysplastic syndrome caregiver advocacy group
myelodysplastic syndrome caregiver blog
myelodysplastic syndrome caregiver discussion
myelodysplastic syndrome caregiver experience
myelodysplastic syndrome caregiver forum
myelodysplastic syndrome caregiver journey
myelodysplastic syndrome caregiver story
myelodysplastic syndrome caregiver support group

myelodysplastic syndrome discussion
myelodysplastic syndrome experience
myelodysplastic syndrome forum
myelodysplastic syndrome journey
myelodysplastic syndrome patient advocacy group
myelodysplastic syndrome patient blog
myelodysplastic syndrome patient discussion
myelodysplastic syndrome patient experience
myelodysplastic syndrome patient forum
myelodysplastic syndrome patient journey
myelodysplastic syndrome patient story
myelodysplastic syndrome patient support group
myelodysplastic syndrome stem cell therapy (story or experience)
myelodysplastic syndrome story
myelodysplastic syndrome support group
myelodysplastic syndrome survivor advocacy group
myelodysplastic syndrome survivor blog
myelodysplastic syndrome survivor discussion
myelodysplastic syndrome survivor experience
myelodysplastic syndrome survivor forum
myelodysplastic syndrome survivor journey
myelodysplastic syndrome survivor story
myelodysplastic syndrome survivor support group
myelodysplastic syndrome treatment advocacy group
myelodysplastic syndrome treatment blog
myelodysplastic syndrome treatment discussion
myelodysplastic syndrome treatment experience
myelodysplastic syndrome treatment forum
myelodysplastic syndrome treatment journey
myelodysplastic syndrome treatment story
myelodysplastic syndrome treatment support group
need help with MDS
need help with myelodysplastic syndrome
surviving MDS

surviving myelodysplastic Syndrome

Setting

In this study, we conducted searches using keywords associated with Myelodysplastic Syndromes (MDS) across all publically available sources of information where patients and caregivers were involved in digital conversations. The dataset for our analysis comprised social media posts sourced from publicly accessible websites. Criteria for inclusion encompassed posts featuring a user-identifiable username, distinguishing them as either MDS patients or caregivers/family members, with content directly related to MDS. Forums had to exhibit discussions pertinent to the primary study objectives and maintain activity with posts dating from 2022 onwards. Only English-language posts were assessed. We chose posts from websites where accessibility without registration and consent for third-party data analysis was mandatory. Posts meeting these criteria were considered retrievable for inclusion in our analysis. Once the relevant sites, forums, social channels, and discussions were identified, data extraction algorithms specific to each site were written in R or Python programming languages to extract posts from each site. The data extracted included the username, post-text, thread title, date of the post, URL (in cases where it is possible).

At the time of extraction, all data was de-identified to remove any patient identifiers including but not limited to, email addresses, web links to personal social media accounts, websites, or blogs, and telephone numbers. We then applied additional checks to exclude posts with duplicates and non-Unicode Transformation Format-8 characters. We also applied autocorrection on misspellings using the Hunspell package in R.

Criteria for Anonymization

The proposed anonymization algorithm was designed to safeguard sensitive information within the dataset. It operated as follows: Firstly, if any patient or caregiver names, whether first, last, or both, were identified, they were promptly removed and replaced with the generic label "NAME" to ensure anonymity. Similarly, provider names, in any combination of first, last, or both, were substituted with the standardized term "PROVIDER NAME" to maintain confidentiality. To preserve longitudinal data integrity, usernames were tokenized, thus ensuring that posts from individual users remained coherent over time. Institutional names such as NCI and Dana Farber were retained within the dataset. Additionally, any therapeutic mentioned by its brand name as listed in Table 1 was converted to its corresponding International Nonproprietary Name (INN). Finally, manufacturer names extracted from the FDA's National Drug Code Directory were expunged and replaced with the generic designation "PRODUCT MANUFACTURER" to protect proprietary information and maintain anonymity throughout the dataset.Posts containing any mention of, or reference to, an investigational drug or specific clinical trial were not scraped into the data corpus (see Exclusion Criteria, below). As an example, the phrase, “my oncologist recommended I look at clinical trials” was included, but a post containing “I decided to start a clinical trial my doctor recommended a trial with a new drug called sabatolimab” was not scraped.

The methodology to anonymize names was implemented using the anonymizer package in R that identified given names, surnames, and titles (e.g., Dr. Smith).

The methodology to exclude references to clinical trials and investigational agents was programmed into the web scraper and validated using posts on forums for liquid tumors other than AML or MDS. Table 1 shows a subset of possible brand names to demonstrate the methodology. All brand names were mapped to the corresponding INN using the national drug code index.

Table 2. **Brand names to INN**

Brand	**INN**
Venclexta	venetoclax
Tibsovo	ivosidenib
Xospata	gilteritinib
Velcade	bortezomib
Daurismo	glasdegib
Idhifa	enasidenib
Mylotarg	gemtuzumab ozogamicin
Clolar	clofarabine

Data Exclusion

Social media posts were considered eligible for inclusion in this study if they met the following criteria: originating from one of the specified forums, containing one of the listed MDS search terms, and being authored by patients or caregivers of patients with Myelodysplastic Syndrome (MDS) who discussed diagnosis or caregiving for individuals diagnosed with MDS, or the progression of MDS to Acute Myeloid Leukemia (AML) or caregiving for individuals experiencing such progression.

Posts meeting any of the subsequent criteria were excluded from the study and subsequently removed from the data corpus: posts discussing MDS in a general context without evidence of the author being either a patient or a caregiver, posts explicitly mentioning participation in a clinical trial or discussing a clinical trial, news and research articles related to MDS lacking patient experiences, and posts related to drug approvals or those primarily focused on content generation without containing the experiences of patients or caregivers.

Data Management

The data corpus was compiled into an Amazon Web Services database according to the inclusion and exclusion criteria outlined earlier. To ensure the integrity of the dataset, multiple sampling checks were conducted by selecting a subset of posts, which underwent manual evaluation to identify and remove any potentially identifying variables. In addition to manual evaluation, extensive automated web scraping techniques were employed, particularly in cases where the primary data source consisted of text data. For data derived from audio or video sources, content was transcribed into text format, with a subset validated manually. Metadata associated with each data source and post was captured wherever available, including information on discussion threads such as title, number of posts, and views; details about users like tokenized user handle, join date, and location (country and state, if applicable); and post-specific information such as timestamp and replies. All data collected from primary sources were stored electronically as a text corpus along with their respective metadata for further analysis and processing.

Study Size

An exploratory feasibility analysis of the data sources and data collection criteria described above identified three main forums (mds-foundation.org/forum, forums.marrowforums.org, and mdspatientsupport.org.uk/forum) that MDS patients and caregivers widely use. These forums were the most critical data sources due to the high volume of posts and user engagement over a significant longitudinal timeframe.

Table 3 **Volume of posts found during the exploratory analysis of data sources.**

Forum	Posts	Threads	Users	Duration
mds-foundation org	6,834	1,426	1,379	2008 August to

marrow forums	32,009	3,722	1,847	2022 December
Mdspatientsupport UK	2,896	392	199	
Total	**41,739**[*]	**5,540**	**3,425**	

[*] *Over 60,000 posts were identified, of which ~41,000 were within this study period (2008, August – 2022, December). The start date was selected based on VIDAZA label expansion for High-Risk MDS in August 2008.*

Each data source was manually vetted before its inclusion in the study to ensure that the sites were established and legitimate.

The methodology used to assess the quality of sites was based on the site domain. For example, forums, YouTube accounts, and blogs with a verified domain from a cancer center, academic institution, or well-known patient or disease support organizations would be considered trusted sites.

Data sources that meet these criteria are listed below:

Table 4 Details of data sources that meet the study inclusion criteria found during the exploratory analysis.

blog.dana-farber	Putting College – and Field Hockey – on Hold for a Bone ...
blog.dana-farber	Making the Best of Things In The Hospital - Blog
blog.dana-farber	Feeling Lucky in an Unlucky Situation | Dana-Farber Cancer ...
blog.dana-farber	Battling Cancer: Restructuring and Enjoying Your New Life
blog.dana-farber	Stem Cell Transplant Gives MDS Patient a Second Chance at ...
bonemarrowtransplantexperience.wordpress	Bone Marrow Transplant: My Experiences with MDS and AML
cancer.osu	Patient Story: Mark Althouse | OSUCCC – James
childrenscancer	Joseph's Story: Surviving Myelodysplastic Syndrome (MDS)
csn.cancer	myelodysplastic - Cancer Survivors Network
emmafightsmds.wordpress	Emma's fight against MDS – My battle against the blood ...
forum.bloodcancer	Anybody here have Myelodysplastic Disorder?
mdandersontlc.libguides	Patient Stories - Leukemia and Myelodysplastic Syndromes ...

mdspatientsupport	MDS Patient Stories
myelomabeacon	Treatment-related myelodysplastic syndromes (MDS)
parkway.chop	Secondary Myelodysplastic Syndrome (MDS): Zamiyha's Story
patientworthy	Living with MDS: Ryan's Story Part 1 - Patient Worthy
patientworthy	Living with MDS: Ryan's Story Part 2 - Patient Worthy
rarediseases	Voices of Rare Cancer: Brian's Story - NORD (National ...
aamds	MY 12-YEAR BATTLE WITH MDS - By Jane Biehl PhD
aamds	"I'm Like You" Patient Stories | Aplastic Anemia & MDS International …
aamds	Stories of Hope | Aplastic Anemia & MDS International ...
anthonynolan	Emma's story - Living with MDS as a young woman
bloodresearch.or.kr	An interesting story of a clone - BLOOD RESEARCH
cancerresearchuk	Myelodysplastic syndrome | Cancer Chat
curetoday	In the Mist of My Fear: Experiences With MDS - CURE Magazine
curetoday	The Roller Coaster of MDS - Cure Today
elcaminohealth	Maria's Story: Myelodysplastic Syndrome (MDS)
leukaemia	Stories - Leukaemia Foundation
leukaemia.au	Diagnosed with MDS, AML, then MDS again
mdanderson	Why a myelodysplastic syndrome patient didn't settle
mdanderson	How a myelodysplastic syndrome survivor found strength
mdanderson	How art helped my daughter cope with MDS treatment
mdanderson	Spouses face back-to-back hairy cell leukemia and ...
mdanderson	Life as a myelodysplastic syndrome survivor: New beginnings
mdanderson	How a clinical trial gave me my life back after MDS and AML
mdanderson	Myelodysplastic syndrome survivor: A targeted therapy clinical ...
mdanderson	Myelodysplastic syndrome researcher: Leaps of faith led to my ...
mdanderson	Myelodysplastic syndromes clinical trial participant: I don't ...

mdanderson	Myelodysplastic syndrome survivor: A stem cell transplant put ...
mdanderson	Myelodysplastic syndrome survivor: Why I joined a clinical trial
mdanderson	Myelodysplastic syndrome: What you need to know
mdanderson	My new normal: Life after myelodysplastic syndrome
mds-foundation	Patient Stories - MDS Foundation
mskcc	Nancie's Story | Memorial Sloan Kettering Cancer Center
youandmds	George's story: How did you find out you had MDS?
youandmds	Patient Video - George's story: How did you find out ...
youandmds	Patient Video - George's story: What has been the biggest ...
youandmds	Patient Video - David's story - You and MDS
youandmds	Abby's story: What advice do you have for other patients on ...
YouTube	Becky's story: What is your MDS gene mutation ... - YouTube
YouTube	Barry's story: What has been the biggest challenge ... - YouTube
YouTube	Patient Success Story | Bone Marrow Transplant | Dr. Rajib De
YouTube	Myelodysplastic Syndrome (MDS) patient discusses her ...
YouTube	A Patient's Story: Myelodysplastic Syndrome - YouTube
YouTube	David's story: What advice do you have for other patients on ...
YouTube	Donna's story: How has your gene mutation profile ... - YouTube
YouTube	Holly's story: What advice do you have for other MDS patients ...
YouTube	Story Corner: Memorie Munson (Caregiver, MDS) - YouTube
YouTube	Aplastic Anemia Patient Story - Mychaela Lovelace - YouTube
YouTube	Story Corner: Mario Rivera and Alison Hines - YouTube
YouTube	David's story: How did you find out you had MDS? - YouTube

YouTube	249: Azra Raza | Myelodysplastic Syndromes And ... - YouTube
YouTube	Max, Sophia and Tom Greb's Story - YouTube
YouTube	George's story: What advice do you have for other patients on ...
YouTube	Patient Story: Akara from Cambodia - YouTube
YouTube	Story Corner: Lisa Zieske (Aplastic Anemia, MDS, PNH)
YouTube	Barry's story: How was your MDS-related anemia diagnosed?
YouTube	Luke & Molly's Story - YouTube
YouTube	Story Corner: Joe Ellenberger (PNH, transplant) - YouTube
YouTube	National MDS Day: Jad Harris' Story - YouTube
YouTube	Story Corner: Adrienne Torrey and Joe Coffidis (MDS)
YouTube	Myelodysplastic syndrome (MDS) survivor Holly ... - YouTube
YouTube	Inspirational Story - Bone Marrow Transplant | Survivor Interview
YouTube	Sharing My Voice: Joan's Story - YouTube
YouTube	The Caregiver’s Journey: Caring for the Caregiver
YouTube	The Caregiver’s Journey: Exploring Emotions
YouTube	The Caregiver’s Journey: Finding Support

Data Analysis

A web crawler-based search was performed using the Google search engine to identify websites that had relevant information related to MDS and HR-MDS. After the validation of source URLs, custom scraping algorithms were created to collect all relevant information from these websites/links.

The corpus of text comprising MDS and HR-MDS patient & caregiver experiences described in section 3.4 was cleaned and normalized in preparation for data analysis. The text was converted into numeric vectors, which preserve the semantics and serve as the primary inputs for various algorithms that were used to analyze the data.

Machine Learning algorithms, including topic modeling and unsupervised clustering, were used to detect latent themes in the data automatically. Each latent theme and cluster uncovered by the algorithms was then manually evaluated for the semantic cohesion of members.

Natural Language Processing (NLP) was employed for data analysis by firstly, topic modeling to identify latent themes, enabling the clustering of documents based on these themes. A Machine Learning classifier was then applied to map posts to specific themes, structuring them in a format resembling a patient journey. In the next step, theme segmentation was conducted to categorize and attribute attributes to each theme. This process involved three primary approaches: a top-down approach that defined attributes based on overall disease knowledge, including symptoms, disease progression, quality of life (QoL), affect, sentiment, etc.; a bottom-up approach that utilized a word cloud and other aggregates from the data corpus to guide segmentation; and a hybrid approach incorporating elements of both top-down and bottom-up methodologies for a more nuanced categorization of comments.

HR-MDS Classification

Using natural language processing and contextualization we identified the mentions of **various keywords, and keyword combinations to deem a post and the patient and caregiver to the higher-risk MDS cohort. A post was considered positive for HR-MDS classification if the post mentions the use of** hypomethylating agents, progression to Acute Myeloid Leukemia, stem-cell transplantation, Allo-HCT/Haplo-, Azacitidine, Decitabine, Venetoclax, any chemotherapy related to MDS, donor lymphocyte infusion, failure of recombinant erythropoietin to improve anemia, multiple transfusions, stem cell transplantation, blast counts equal to or greater than 20%, and intensive chemotherapy

Sample Size and Saturation

Usually, theme generation is directly proportional to several posts included however, at a point, the new significant themes stop getting generated, and is called the saturation point of the analysis. Saturation was defined as when no new significant themes emerged from the analysis.

User Classification

Classification algorithms automatically classify each experience's perspective into that of a patient or caregiver. The automated classification results were reviewed systematically for inaccuracies, and appropriate corrective measures were applied.

Classification was a critical step in the analysis, to correctly identify from the post whether the user is a patient, caregiver, or physician, and was very crucial to tap into the correct themes for each group. Certain key phrases like “I have MDS”, “I was diagnosed with MDS”, “I am living with MDS”, “I suffer from MDS” etc. were used to determine the type of user from the posts scraped. This was a semi-autonomous step and was only able to classify the sample partially, some manual interventions were required to classify every post.

Analysis

QDA (qualitative data analysis) was performed for each objective. Posts were first qualitatively analyzed using the R Qualitative Data Analysis Package. Theme identification as defined in the methodology section was analyzed against the data corpus for each method until the saturation point was reached. The results of theme analysis for each QDA algorithm were analyzed for correct theme identification, number of themes identified, and posts required to reach the saturation point for each theme. Once the best algorithm was identified, the occurrence of each theme was quantitatively summarized using descriptive statistics. The corpus of text comprising MDS and HR-MDS patient & caregiver experiences described in section 2.4 was cleaned and normalized in preparation for data analysis. The text was then converted into numeric vectors, which captured the semantics and served as the primary inputs for various algorithms used to analyze the data.

Following the QDA, a classification algorithm was then employed to automatically classify each experience's perspective into that of a patient or caregiver. The automated classification results were reviewed systematically for inaccuracies, and appropriate corrective measures were applied.

We also evaluated the effect of "prolific users" that post frequently or in-depth posts in each forum or across forums. We normalized this outlier impact using statistical methods and analyzed any actions or sentiments other users might take based on these influencers, and how these themes propagate across users.

Machine Learning algorithms, including topic modeling and unsupervised clustering, were used to detect latent themes in the data automatically. Each latent theme and cluster were then evaluated systematically for the semantic cohesion of topic/cluster members.

Primary Analysis

Advanced Natural Language Processing (NLP) techniques combined with Machine Learning methods were deployed to analyze the data corpus comprehensively. Additionally, data visualization techniques, such as word clouds, heatmaps, histograms, and network graphs, were utilized to enhance the analysis process. NLP played a crucial role in discerning whether posts were authored by patients or caregivers, determining the sentiment of the authors, and categorizing each post into one or more broad themes:

1. Clinical: Posts addressing patient conditions, diagnoses, disease progression, and monitoring.

2. Diet and Lifestyle: Posts focusing on dietary adjustments and day-to-day challenges associated with living with MDS.

3. Education and Logistics: Posts centered around understanding the disease and available care options.

4. Emotional: Posts providing or seeking emotional support among patients or between patients and caregivers.

5. Physical: Posts concerning the management of disease symptoms and treatment-related side effects.

6. Transplants: Posts discussing the benefits, risks, and personal experiences with transplants.

7. Treatments: Posts related to inquiries or concerns about treatments, including treatment cycles, regimens, decisions, transfusions, and other palliative treatments. This category is further divided into Disease Burden for patients and Human Burden for caregivers.

Moreover, topic modeling and clustering algorithms were employed to uncover additional broad themes solely from the data, leading to the segmentation of posts into more detailed sub-themes.

Posts mapped to various themes and sub-themes underwent further analysis to reveal insights regarding the disease journey, treatment or care gaps, unmet needs, disease knowledge, and the stage or phase of the disease journey. Each broad theme and sub-theme resulting from the analyses were categorized as actionable or non-actionable and reported accordingly.

Determination of the author:

The author of each post could either be the patient himself/herself, a caregiver, or a third actor who is neither of the two. Each post, therefore, was assigned one or more of the following labels:

- Patient
- Caregiver

The process of categorizing users into two distinct groups - Patients and Caregivers - based on their posts presented challenges due to the extensive amount of text associated with each user and the prevalence of irrelevant or ambiguous information within the posts. To address this complexity, an innovative classification approach was implemented wherein initially, a classification model was deployed to assign each post with either a 'Patient' or 'Caregiver' label. Subsequently, a user was classified as either a Patient or a Caregiver if at least 80% of their posts received the same label from the classification model. In instances where no clear majority (80% or higher) was evident, the posts were systematically analyzed in chronological order, progressing from the earliest to the most recent. This analytical process continued until the model confidently assigned a label ('Patient' or 'Caregiver') to the user.

The data corpus utilized in this study consisted of a substantial volume of posts, necessitating an automated mechanism to assign each post one or more labels as defined previously. This process unfolded in two stages, outlined below:

Initially, a numerical vector representing each post was generated utilizing one or more of the following methodologies:

Employing a transformer-based Large Language Model such as Bidirectional Encoder Representations from Transformers (BERT) or a similar alternative model to parse the post and generate a semantic numerical vector representation.

Utilizing a word frequency-based numerical representation of the post that did not capture the semantic nuances of the post.

Subsequently, the numerical vector representation of each post was inputted into a multi-class classification model trained using a representative subset of the corpus. This model then automatically assigned a predetermined label to each post, accompanied by a confidence score for each label. Various algorithms were employed for the multi-class classification task, including Linear Classifiers (Logistic, Bayesian), Support Vector Machines, Decision Tree-based classifiers, and Ensemble methods (AdaBoost, LightGBM, XGBoost).

The selection of the algorithm was informed by random sample analyses and the categorization of posts by three human curators, each reviewing at least 5% of all posts independently. Following individual reviews, the curators convened to reach a consensus on the categorization. The model exhibiting the highest sensitivity and specificity in comparison to the final curation was ultimately chosen.

Theme Assignment:

Each post was mapped to several themes in a theme hierarchy. These themes were organized in an inverted hierarchical tree composed of five high-level “root” themes with more detailed themes mapped as branches for one or more roots.

The major themes in the hierarchy were:

1. Disease Burden
2. Treatment Decision
3. Unmet needs

4. Life Milestones
5. Logistic Burden

A combination of two approaches were utilized for the theme assignment, described below:

Top-down analysis: Using this technique, each post was mapped to one of the pre-determined major themes documented above. This is essentially a supervised classification machine learning problem and therefore the solution that was deployed here is like the one described in section 3.7.4.1

Bottom-up analysis: A complementary approach to the top-down analysis, this technique was used to uncover hidden or latent themes in the corpus. Each post was algorithmically assigned to one of several latent buckets. Each latent bucket was analyzed by a subject matter expert. To perform this analysis, each post was transformed into representation vectors using the procedure described in part 1 of section 3.7.4.1. Following that, the numerical vector representations of the posts were utilized as input to a range of unsupervised machine learning algorithms, which are elaborated upon below to unveil latent themes. We used Latent Dirichlet Analysis and Latent Semantic Analysis for topic modeling. We also used unsupervised clustering algorithms including K-means clustering, Spectral clustering, Agglomerative Clustering, DBSCAN, and OPTICS.

Each latent theme discovered during the bottom-up approach was either mapped to one of the major themes as a sub-theme or used to expand the list of major themes.

Each major/minor theme in the hierarchy of themes that is an outcome of this multi-dimensional approach was subsequently categorized and reported as actionable or non-actionable.

At the heart of the primary objectives of this study is identifying themes that are important to MDS and HR-MDS patients and caregivers. Themes were identified by applying topic modeling and unsupervised clustering to posts and blog articles in the cleaned and normalized data corpus. Themes were scored by their prevalence, i.e., the number of unique discussions in the data corpus that mention a theme. Once scored, themes were categorized as minor or major by applying a threshold to the prevalence score. Themes were also categorized as actionable or non-actionable. An actionable theme is one that identifies a gap or need that can be addressed, whereas a non-actionable theme would typically be emotional. Furthermore, a network of themes was constructed by connecting two themes if they are co-mentioned in the same discussion, the strength of the connection being the number of such co-mentions.

Secondary Analyses

Every post captured during the process of data collection was collected with a date stamp. Each date stamp was converted to the month and year of the post and was used to determine the pre-covid activity or post-covid activity. A matched pre-covid period included posts from January 2018 through February 2020. All posts from March 2020 were included in the post-covid period. Most of the analysis that was performed on the complete data corpus was also conducted during these two time periods to determine changes in activity, behaviors, and perception of patients and their caregivers.

One of the secondary objectives was to determine themes and insights from a very specific group of patients who either in their post mention high-risk or are deemed high-risk due to the presence of certain high-risk treatments or procedures. These treatments and procedures were determined using the NCCN guidelines for high-risk MDS population and were used as surrogates to tag a post as originating from high-risk versus low-risk patients..

To address the secondary objective of elucidating the experiences of high-risk MDS patients and their caregivers, we first used NLP to automatically label each discussion in the data corpus as "low", "intermediate," high" and "very high" or unknown risk. Quality Control and Pharmacovigilance

Each data source underwent a manual evaluation to ensure the quality and relevance of the data sources considered for the study. The diverse formats of the data sources required distinct criteria tailored to each type of resource. For each data source, at least one or more of the following parameters were utilized for evaluation:

1. The reputation of the domain hosting the resource.
2. For discussion forums, considerations included the number of users and the longitudinal duration of user engagement.

For audio/video content, factors such as the number of views and comments were considered whenever available.

Results

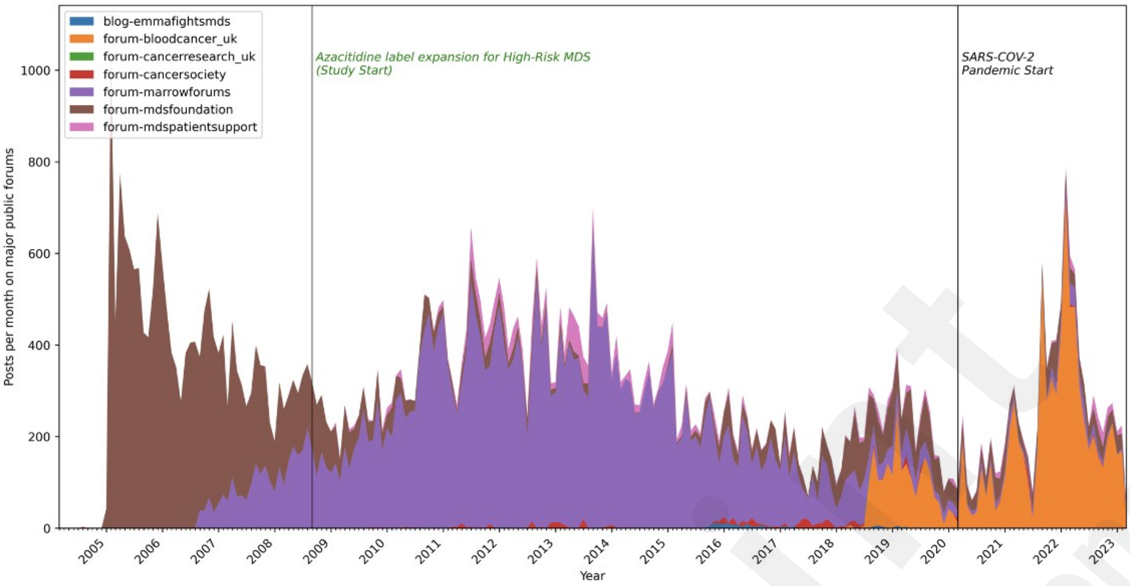
A substantial dataset was amassed, comprising roughly 5.5 million words extracted from 42,000 posts distributed across 5,500 threads, involving approximately 4,000 users. Our algorithms identified 49% of these users as patients, 42% as caregivers, and 9% as others, primarily hailing from the United States, United Kingdom, and Canada. Among the identified users, 1,249 were classified as High-risk MDS users, with 588 (47%) being patients and 661 (53%) caregivers.

Approximately 30 distinct sites were scrutinized for pertinent textual and video content, revealing three prominent forum sites that dominated the discourse: mds-foundation.org/forums, forums.marrowforums.org, and mdspatientsupport.org.uk.

Analysis of the data suggests that a small fraction of super-users (~1%) wield significant influence over the engagement levels of other users. Notably, the MDS patient support forum exhibited the highest level of user engagement, quantified by the duration of user activity (posting) on the forum, which ranged from 2 days to over 1000 days, with a median of 10 days.

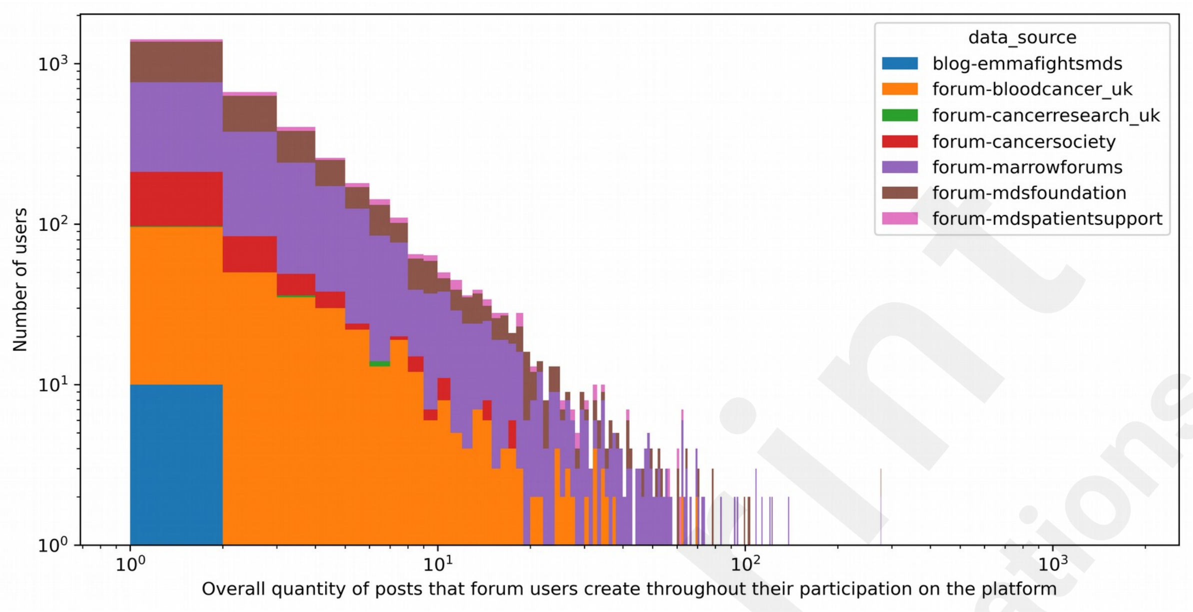
Our findings underscore the consistent presence of MDS patients and caregivers across various forums. However, recent observations indicate a notable shift in activity levels from publicly accessible and analyzable forums to more private platforms, as illustrated in the post frequency chart in Figure 1.

Figure 1 **Post frequency of conversation from various sources mapped on a yearly scale from 2005 to 2023.**



It was noted that most online forums have low user retention rates, with most users disengaging within two days. However, Marrow Forums and the MDS Foundation show relatively higher retention rates compared to others. Some superusers are the perpetrators of these conversations, some reactors react to these posts, and there are passive consumers who consume this information without active participation. Figure 2

Figure 2 **Analysis of data reveals that that a small group of super-users or influencers significantly impact the engagement of other users. The "emmafightsmds" forum exhibited a skewed pattern due to the behavior of these super-users, while other forums displayed a more dispersed engagement.**



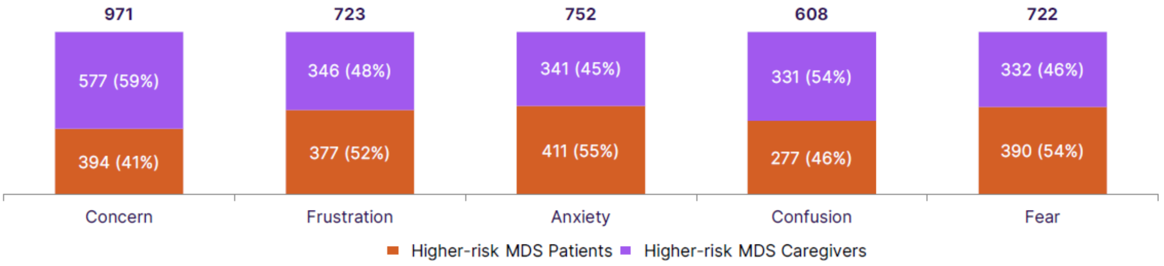
Emotional Classification

In our comprehensive analysis of the posts, we discerned seven major emotional categories. A significant portion of the posts (37%) exhibited a notably neutral emotional tone. Following closely, sadness (22%) and joy (20%) emerged as the second and third most prevalent emotions. Surprisingly, 13% of the posts conveyed surprise, while the remaining 8% depicted emotions revolving around fear, disgust, and anger.

Delving deeper, we endeavored to categorize these primary emotions into finer distinctions. It transpired that 17% of the posts resonated with a sense of worry, while 13% exuded hopefulness. Summarizing our emotional analysis, posts attributed to patients underscored the emotional weight accompanying the onset of the disease. Patients grappled with fear and harbored anxiety regarding their impending treatment. Numerous posts depicted a common theme: both patients and their caregivers felt ill-prepared for the diagnosis, grappling with confusion stemming from their limited understanding of the disease and available treatment options.

Within the subset of high-risk MDS patients identified, prevalent sentiments included concern (78%), anxiety (60%), frustration (58%), fear (58%), and confusion (49%). Notably, concern predominated among caregivers (n=971, 59%), while anxiety was more pronounced among patients (n=752, 55%) (see Figure 3).

Figure 3 **Predominant sentiments by High-risk MDS patients and caregivers.**



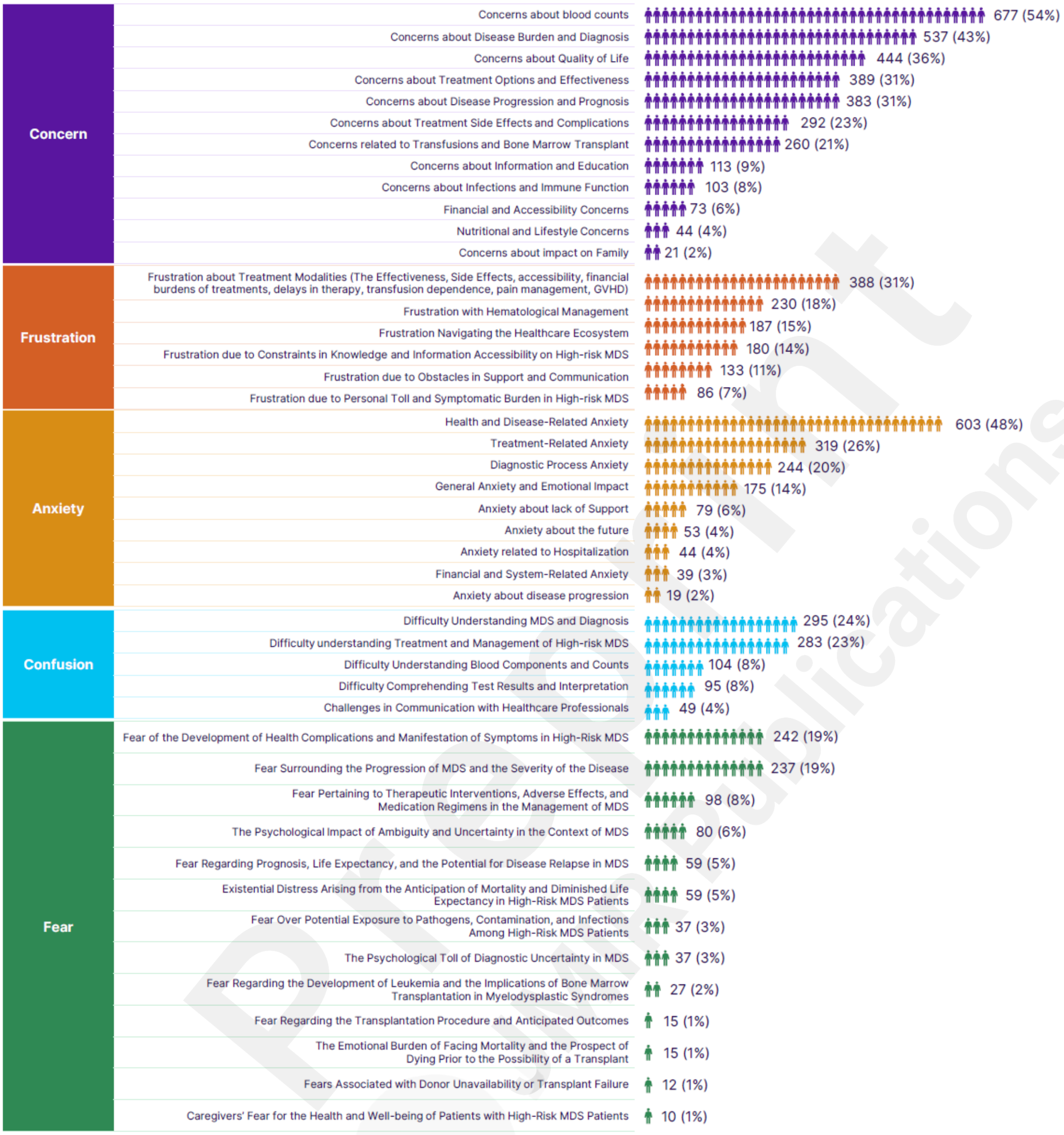
To take our understanding deeper, we created an organic emotional theme map for each top theme we observed during our data exploration.

Within the major emotional bucket related to "Concern", we found that there were some very specific themes related to concerns related to blood counts (n=677, 54%), the burden of the disease (43%), QoL (36%), available treatment options and effectiveness (31%), and disease progression and prognosis (31%). Anxiety related to health and disease (48%), treatment (26%), and the diagnostic process (20%) were also common (Figure 4).

Similarly, within the "Frustration" bucket the top themes observed were related to the frustration of patients and caregivers related to treatment modalities (31%), hematological management (18%), navigating through the healthcare system (15%), constraint in knowledge and information accessibility on MDS (14%) (Figure 4).

The most common sentiments related to fear were the potential development of health complications and the manifestation of symptoms (19%) and the progression and exacerbation of MDS (19%). Additionally, confusion was pervasive among participants, with 295 (24%) individuals finding it challenging to comprehend the nuances of MDS and its diagnosis. A systematic analysis of the principal domains for which information is being sought about HR-MDS revealed frequent mention amongst users of acquiring information on therapeutic intervention (19%), and an interest in ongoing research associated with the disease (17%) (Figure 4).

Figure 4 Organic view of sub-sentiments expressed by High-risk MDS patients and caregivers for each major sentiment buckets.



We conducted a more detailed analysis and discovered that within the thematic heading "frustration related to treatment modalities," the predominant topics of conversation included discussions on side effects and efficacy, delayed or postponed therapeutic interventions, dependence on transfusions, inadequate pain management, and the economic burden associated with treatment, among others. Within the thematic heading “Navigating the healthcare system’ we observed conversations around discontentment with the physicians and their communication, frustration due to perceived dismissiveness, and lack of information provided by physicians. Of all the posts, 14% of the posts had topics of conversation related to “Frustration to constraint in knowledge” and the most common conversation within this thematic bucket were around lack of understanding of the MDS as a disease, heterogeneity of the disease, and ambiguity of the information from physicians. We also observed through these posts that like any other disease, patients, and caregivers, go through severe psychological distress, they suffer due to unknown implications of the available treatment options, complications that come with the disease and side effects, lack of empathy in social circles, as well as insufficient support and communication from physicians (Figure 3).

Figure 3 **Sub theme level analysis for the Thematic Category of Frustration**

Treatment Modalities: Frustrations concerning the effectiveness, side effects, accessibility, and financial burdens of treatments, as well as delays in therapy, transfusion dependence, pain management, and complications such as Graft-Versus-Host Disease.

Hematological Management: Frustrations faced in managing blood and platelet transfusions, including difficulties in improving platelet counts, identifying compatible donors, handling complications, and addressing logistical issues.

Navigating the Healthcare Ecosystem: frustration encompassing discontentment with professionals' communication and approach, perceived dismissiveness and insufficient information from physicians, systemic challenges, and health insurance coverage issues.

Constraints in Knowledge and Information Accessibility on MDS: Frustrations concerning Myelodysplastic Syndromes (MDS), including the scarcity of research, complexity, limited resources, diagnostic challenges, difficulties in identifying causes of clinical manifestations, and ambiguity in information.

Obstacles in Support and Communication: Deficiency of resources and support systems, inadequacies in communication and information, lack of empathy from social circles, extended waiting times for transplants, and insufficient support and communication from healthcare professionals.

Personal Toll and Symptomatic Burden in High-Risk MDS: Frustrations due to the ineffectiveness of treatments, persistence of symptoms, personal constraints, and ophthalmological complications.

Discussion

In recent years, social media platforms have emerged as valuable repositories of insights into the diverse patient experiences while navigating illness, as noted by Watson et al. (2022) and Monti et al. (2022). The unfiltered and spontaneous dialogues observed on these platforms offer genuine reflections of patients' and caregivers' immediate concerns and evolving challenges throughout their illness trajectories. By attentively listening to these unprompted exchanges, researchers can discern not only existing unmet needs but also previously unrecognized support requirements for patients and their caregivers. Our study meticulously examines the experiences of patients and caregivers coping with MDS, delving beyond overarching themes to explore the nuances of their conversations and uncover the underlying reasons prompting their engagement with digital platforms.

Recent investigations have explored general themes surrounding the experiences of patients and caregivers dealing with MDS. Frank et al. (2022) and Booth et al. (2019) have employed social media listening techniques to gain insights into the MDS patient community. Booth et al. (2019) focused on the disease and treatment encounters of 347 patients with AML or MDS ineligible for intensive chemotherapy, though the study's limited sample size offers only directional insights from a constrained user pool. In contrast, our research presents a more comprehensive directional profile based on data from over 4,000 users, offering deeper insights into patient and caregiver perspectives.

While Booth et al. (2019) conducted a qualitative analysis of user posts without specifically identifying risk statuses, our study delineates two distinct patient archetypes and presents comparative results. Consistent with our findings, Booth et al. (2019) highlighted significant concerns, including patients' perceptions of limited treatment options and age-based treatment eligibility criteria. They also underscored a prevailing lack of information and knowledge regarding MDS/AML treatment options, fostering uncertainty and decision-making challenges among patients and caregivers. Similarly, our study underscores the anxiety and frustration stemming from the inadequate awareness surrounding the disease among patients and caregivers. Our findings emphasize the critical need for improved physician education and disease awareness, which can positively impact treatment engagement and management behaviors, thereby alleviating emotional burdens for patients and caregivers.

Frank et al. (2022) conducted an extensive analysis of over 20,000 comments from country-specific online forums spanning 2011 to 2019. Utilizing proprietary natural language processing algorithms, they categorized unmet needs of MDS patients and caregivers across various countries. Their study identified common discussion topics and themes, revealing patterns and interconnections across posts. Like our research, Frank et al. (2022) noted expressions of negativity among patients in the USA, UK, and Canada at the time of diagnosis, attributed to perceived educational and awareness gaps among healthcare providers and uncertainties regarding treatment outcomes due to disease variability. Negative patient experiences included encounters with non-specialist physicians lacking sufficient expertise in MDS diagnosis and treatment options, as well as conflicting treatment recommendations and ambiguous diagnostic information.

In summary, our study contributes to the growing body of literature elucidating the multifaceted experiences of patients and caregivers grappling with MDS. By leveraging social media platforms and advanced analytical techniques, we offer valuable insights into the nuanced challenges faced by this community, underscoring the importance of enhanced physician education and disease awareness to optimize patient care and support.

Limitations

This study used data collected from various electronic sources available publicly on the internet. While the analysis of this data can lead to powerful and actionable insights, this study also acknowledges the limitations arising from such data usage.

Lack of representativeness: The MDS patient/caregiver population using the internet for research is limited to those who choose to engage in social media. Therefore, this study implicitly excluded any MDS patients or caregivers who do not have access to the internet or decided not to engage with social media platforms we were collecting.

Geographical bias: The objective of this study was to capture MDS patient/caregiver experience in the English language. This very choice limits the scope of this study to the regions of the world that use the English language.

Active engagement bias: An unknown proportion of the users of online resources are likely to be passive consumers. The experiences that this segment of the MDS patient/caregiver population may not necessarily be reflected in the data used for this study.

Lack of Heterogeneity: Healthcare systems vary widely across even English-speaking regions, and so do the experiences of the MDS patient/caregivers. This study does not claim to capture this heterogeneity across various geo-political entities accurately.

Missing data: It is anticipated that not all patients or caregivers posted all variables listed in the variable table.

Variable Longitudinality: There were substantial variation in frequency of posts. In those cases, themes and sentiment could only be determined at snapshot(s) along the patient's journey.

Comparison with Prior Work

Conclusions

MDS patients and caregivers predominantly experience concern, frustration, anxiety, confusion, and fear. Concern stems from apprehensions about blood counts, disease burden, and quality of life. Frustration arises due to perceived inadequacies and complexities in treatment and management, while anxiety is linked to health deterioration and treatment outcomes. Confusion is attributed to the difficulty in comprehending the nature of MDS and its management. Uncovering and mapping underlying themes with sentiments along the MDS patient journey can inform areas of need for patient-centered care and the development of patient-focused solutions with the potential to improve the patient experience. Ongoing exploratory research includes focus on higher-risk MDS patients and specific related issues. Findings from this study can inform areas of need for HR-MDS patient-centered care and development of patient-focused solutions with the potential to improve the patient experience.

Acknowledgements

Funding for this research was provided by Gilead Sciences Inc.
Authors' Contributions: All authors provided substantial contributions to the conception and design of work, and interpretation of results and were involved in the review and approval of this manuscript for publication.

Conflicts of Interest

“None”.

Abbreviations

AE: adverse event
AML: acute myeloid leukemia
QoL: quality of life
MDS: myelodysplastic syndrome
HR-MDS: High-risk Myelodysplastic syndrome
ESMO: European Society for Medical Oncology
IPSS-M: International Prognostic Scoring System – Molecular
a-HCT: allogeneic hematopoietic cell transplantation
NCCN: National Comprehensive Cancer Network
ECOG: Eastern Cooperative Oncology Group
AI: Artificial Intelligence
NLP: Natural Language Processing
ML: Machine Learning

References

1. Cazzola M. Introduction to a review series: the 2016 revision of the WHO classification of tumors of hematopoietic and lymphoid tissues. Blood. 2016;127(20):2361-2364. doi:10.1182/blood-2016-03-657379
2. Greenberg PL, Stone RM, Al-Kali A, et al. NCCN Guidelines® Insights: Myelodysplastic Syndromes, Version 3.2022: Featured Updates to the NCCN Guidelines. Journal of the National Comprehensive Cancer Network. 2022;20(2):106-117. doi:10.6004/jnccn.2022.0009
3. Fenaux P, Haase D, Santini V, Sanz GF, Platzbecker U, Mey U. Myelodysplastic syndromes: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†☆. Annals of Oncology. 2021;32(2):142-156. doi:10.1016/j.annonc.2020.11.002
4. Schanz J, Tüchler H, Solé F, et al. New Comprehensive Cytogenetic Scoring System for Primary Myelodysplastic Syndromes (MDS) and Oligoblastic Acute Myeloid Leukemia After MDS Derived From an International Database Merge. J Clin Oncol. 2012;30(8):820-829. doi:10.1200/JCO.2011.35.6394
5. Bernard E, Tuechler H, Greenberg PL, et al. Molecular International Prognostic Scoring System for Myelodysplastic Syndromes. NEJM Evidence. 2022;1(7):EVIDoa2200008. doi:10.1056/EVIDoa2200008
6. Estey E, Thall P, Beran M, Kantarjian H, Pierce S, Keating M. Effect of Diagnosis (Refractory Anemia With Excess Blasts, Refractory Anemia With Excess Blasts in Transformation, or Acute Myeloid Leukemia [AML]) on Outcome of AML-Type Chemotherapy. Blood. 1997;90(8):2969-2977. doi:10.1182/blood.V90.8.2969
7. Hofmann WK, Heil G, Zander C, et al. Intensive chemotherapy with idarubicin, cytarabine, etoposide, and G-CSF priming in patients with advanced myelodysplastic syndrome and high-risk acute myeloid leukemia. Ann Hematol. 2004;83(8):498-503. doi:10.1007/s00277-004-0889-0
8. Alessandrino EP, Della Porta MG, Pascutto C, Bacigalupo A, Rambaldi A. Should Cytoreductive Treatment Be Performed Before Transplantation in Patients With High-Risk Myelodysplastic Syndrome? JCO. 2013;31(21):2761-2762. doi:10.1200/JCO.2012.48.0525
9. Garcia-Manero G, Chien KS, Montalban-Bravo G. Myelodysplastic syndromes: 2021 update

on diagnosis, risk stratification and management. American Journal of Hematology. 2020;95(11):1399-1420. doi:10.1002/ajh.25950

10. Hellström-Lindberg E, Tobiasson M, Greenberg P. Myelodysplastic syndromes: moving towards personalized management. Haematologica. 2020;105(7):1765-1779. doi:10.3324/haematol.2020.248955
11. Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol. 1982;5(6):649-655.
12. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8
13. Watson J. Social Media Use in Cancer Care. Seminars in Oncology Nursing. 2018;34(2):126-131. doi:10.1016/j.soncn.2018.03.003
14. Monti C, Aiello LM, De Francisci Morales G, Bonchi F. The language of opinion change on social media under the lens of communicative action. Sci Rep. 2022;12:17920. doi:10.1038/s41598-022-21720-4
15. Frank PP, Lu MXE, Sasse EC. Educational and Emotional Needs of Patients with Myelodysplastic Syndromes: An AI Analysis of Multi-Country Social Media. Adv Ther. Published online September 22, 2022:1-15. doi:10.1007/s12325-022-02277-0
16. Booth A, Bell T, Halhol S, et al. Using Social Media to Uncover Treatment Experiences and Decisions in Patients With Acute Myeloid Leukemia or Myelodysplastic Syndrome Who Are Ineligible for Intensive Chemotherapy: Patient-Centric Qualitative Data Analysis. J Med Internet Res. 2019;21(11):e14285. doi:10.2196/14285