

Listening and Learning from the Voices of Patients with Myelodysplastic Syndromes and Their Caregivers: A Comprehensive Social Media Analysis Using Machine Learning and Natural Language Processing-Based Algorithms

Rohit Marwah,¹ Subrat Mishra,¹ Anirban Jana, PhD,¹ Benjamin Gross, PhD,¹ Sandy Couturiaux,¹ Rico Calara, PhD,² Eduardo J Sabate, MD,² Emily Goldmann PhD,³ C. Hoge, PhD²

¹Definitive Healthcare, LLC, Framingham, MA; ²Medical Affairs, Gilead Sciences, Inc., Foster City, CA; ³Department of Epidemiology, School of Global Public Health, New York University, New York, NY



Key Findings

- This study utilizes advanced algorithms and Natural Language Processing (NLP) techniques (Figure 4, Figure 5) to systematically analyze vast digital footprints of Myelodysplastic Syndromes (MDS) patients and caregivers, revealing that super-users (~1%) drive engagement across forums (Figure 1), with most users originating from the US followed by UK.
- A notable trend is the consistent engagement of MDS patients and caregivers across various forums; however, there's a recent shift from public to private forums, with the "MDS Patient Support" forum registering the highest user activity (Figure 2, Figure 3).
- The derived insights can be harnessed to develop dynamic knowledge bases, patient story hubs, and enhanced patient support forums, facilitating tailored assistance, guidance, and redirection to patients in need.

Conclusions

- The study analyzed 42,000 posts from 4,000 users, mainly from the US, UK, and Canada, revealing 49% of the users as patients and 42% as caregivers with user engagement ranging from 2 to 1,000+ days.
- MDS patients and caregivers consistently engage on forums with an evolving shift from public to private platforms. A small group of highly active super-users drives this engagement.
- Dominant emotions were worry (17%), frustration (16%), and hope (13%), driven by diagnosis severity, treatment concerns, and optimism related to treatment outcomes. Topics touched on disease management, treatment access, healthcare communication, quality of life, and a knowledge gap.

Introduction

- Social media platforms present a unique opportunity to understand the patient lived experience as they manage illness.
- Uncovering and mapping themes from such unsolicited spontaneous patient- or caregiver-driven dialogues can highlight unmet needs, concerns and challenges throughout a disease course, from diagnosis through treatment.

Objective

- This study aims to develop an agile reusable framework to collect and analyze evolving social media data to systematically capture insights into patients and caregivers' lived experiences related to myelodysplastic syndromes (MDS). Discerning unmet needs of individuals affected by MDS and their caregivers can help inform development of appropriate patient support tools.

Methods

- We identified and scraped English-language websites discussing MDS from 2008 to 2022 using a web-crawler and rigorous validation processes, resulting in normalized patient and caregiver narratives.
- The extracted textual information was converted into numeric vectors, preserving semantic essence, and was processed using Machine Learning techniques like topic modeling and unsupervised clustering.
- Algorithmically derived themes were reviewed and labeled by subject matter experts to ensure contextually relevant interpretations.

Fig 1. User activity: A small group of super-users (~1%) drive engagement across forums

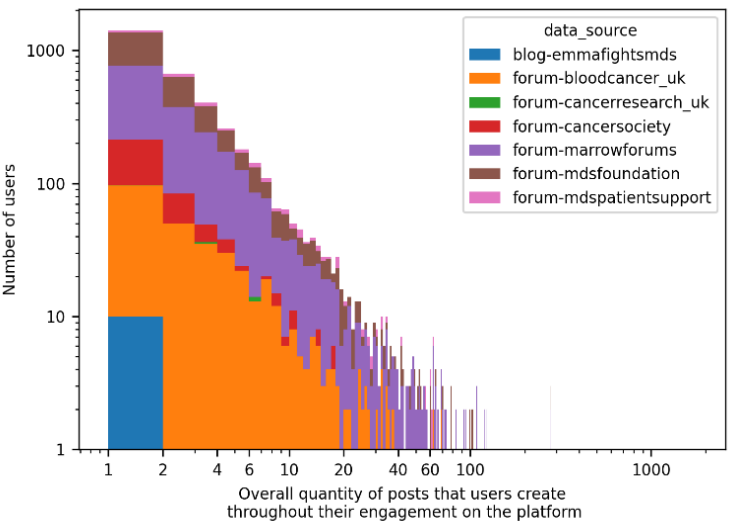


Fig 2. Post Frequency: MDS patients and caregivers consistently engage on forums, with a recent shift from public to private platforms observed

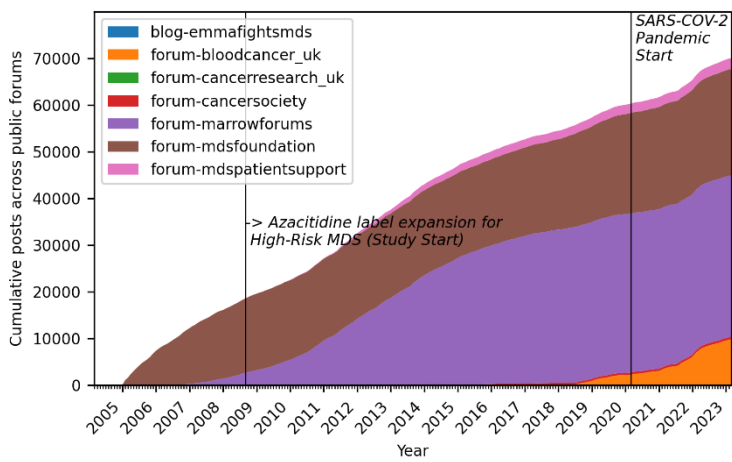
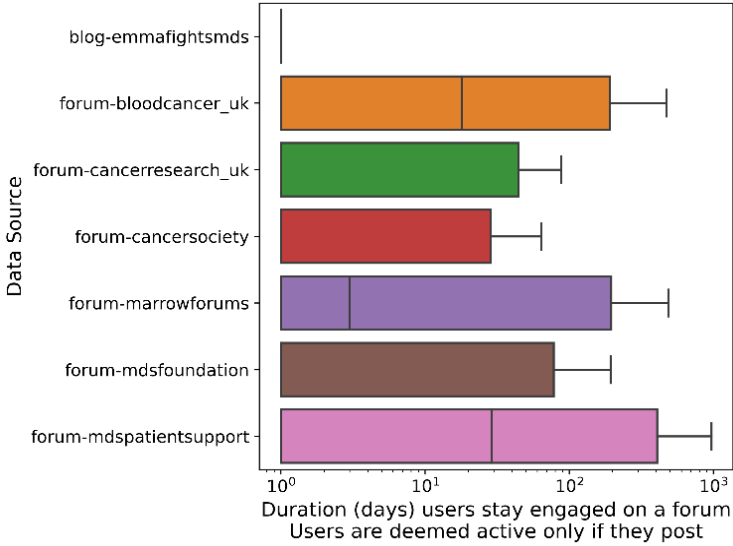


Fig 3. Activity hubs: User engagement, as measured by the median number of days that users were active (posting) on the forum



The collected text totaled 5,500,864 words, roughly five times the length of the seven-book Harry Potter series.

In this study of approximately 4,000 users, 49% were patients and 42% were caregivers. Geographic data was available for 44% of these users, with the majority from the US, followed by the UK.

Results

Fig 4. Natural Language Processing model classifications examples: Patient and caregiver emotions are well captured

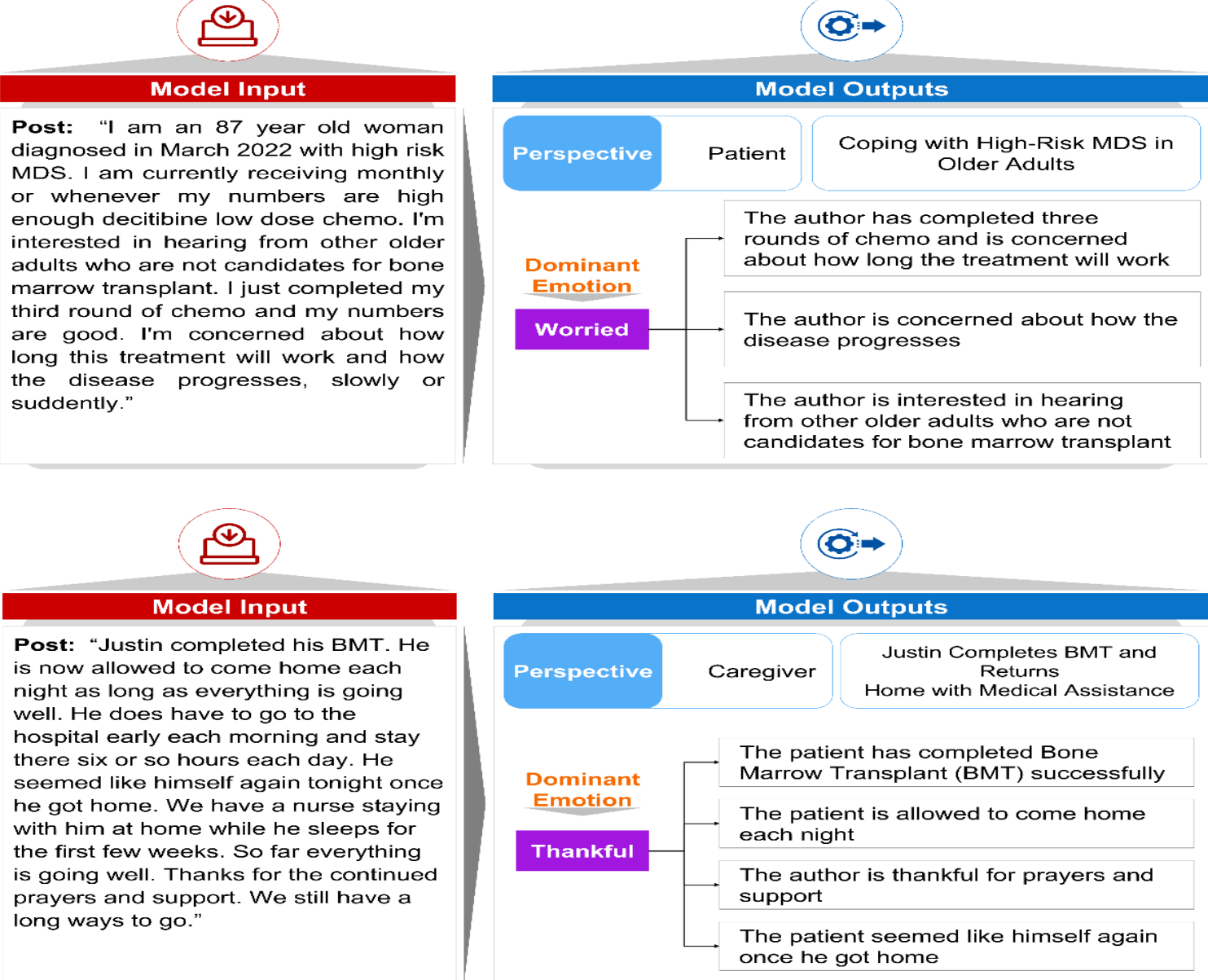


Fig 5. Emotion Classification: Spectrum of emotions is complex across the ~50,000 posts analyzed; the dominant emotions expressed appear primarily related to worry (17%), gratitude (16%), and hope (13%)



References: Frank, Pauline P et al. *Adv Ther.* 2022;1-15. Booth, Alison et al. *J Med Internet Res.* 2019;21(11):e14285. Watson J. *Seminars in Oncology Nursing.* 2018;34(2):126-131. Monti, Corrado et al. *Sci Rep.* 2022;12:17920.

Acknowledgments: This study was funded by Gilead Sciences, Inc., Foster City, CA, USA.

Disclosures: RM, SM, AJ, BJ, SC, are employed by Definitive Healthcare, LLC. CH, RC and ES are employed by Gilead Sciences, Inc. EG reports no conflicts of interest.

Correspondence: Sandy Couturiaux (scouturiaux@definitivehc.com)