Unlocking the potential of synthetic patients for accelerating hematologic research

Marta Cipriani

fondazione GIMEMA onlus
per la promozione e lo sviluppo della ricerca scientifica sulle malattie ematologiche. FRANCO MANDELLI

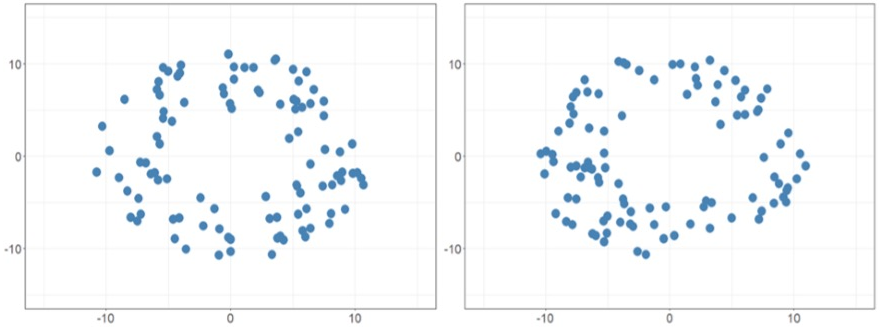
GIMEMA International Meeting

New Frontiers in Hematologic Research: Quality of Life and Artificial Intelligence

Disclosures of Marta Cipriani

Company name	Research support	Employee	Consultant	Stockholder	Speakers bureau	Advisory board	Other

What is synthetic data?

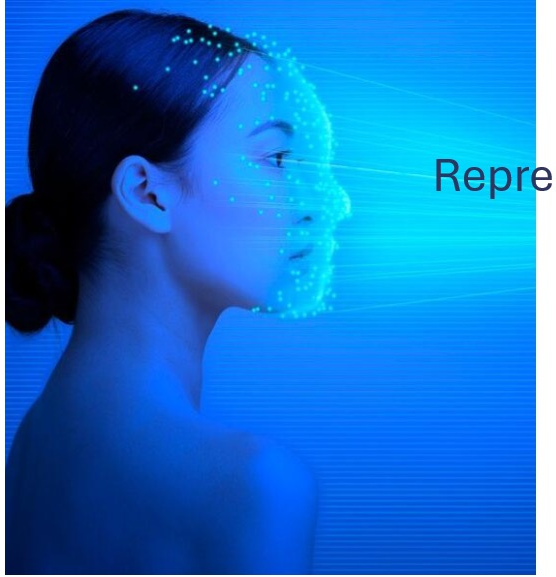


Original data

Synthetic data

Synthetic data is information that has been crafted by a **model** rather than collected or measured in the **real world**.

Realistic



Representative

Synthetic data

Anonymous

Generation of synthetic data

1

DEEP LEARNING METHODS

2

MACHINE LEARNING METHODS

3

PARAMETRIC METHODS

4

OTHERS…

Generation of synthetic data

1

DEEP LEARNING METHODS

2

MACHINE LEARNING METHODS

SYNTHPOP by Nowok et al. (2016)

3

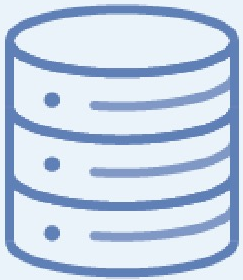
PARAMETRIC METHODS

FLESSY by Cipriani et al. (2025)

4

OTHERS…

How do we use synthetic data?

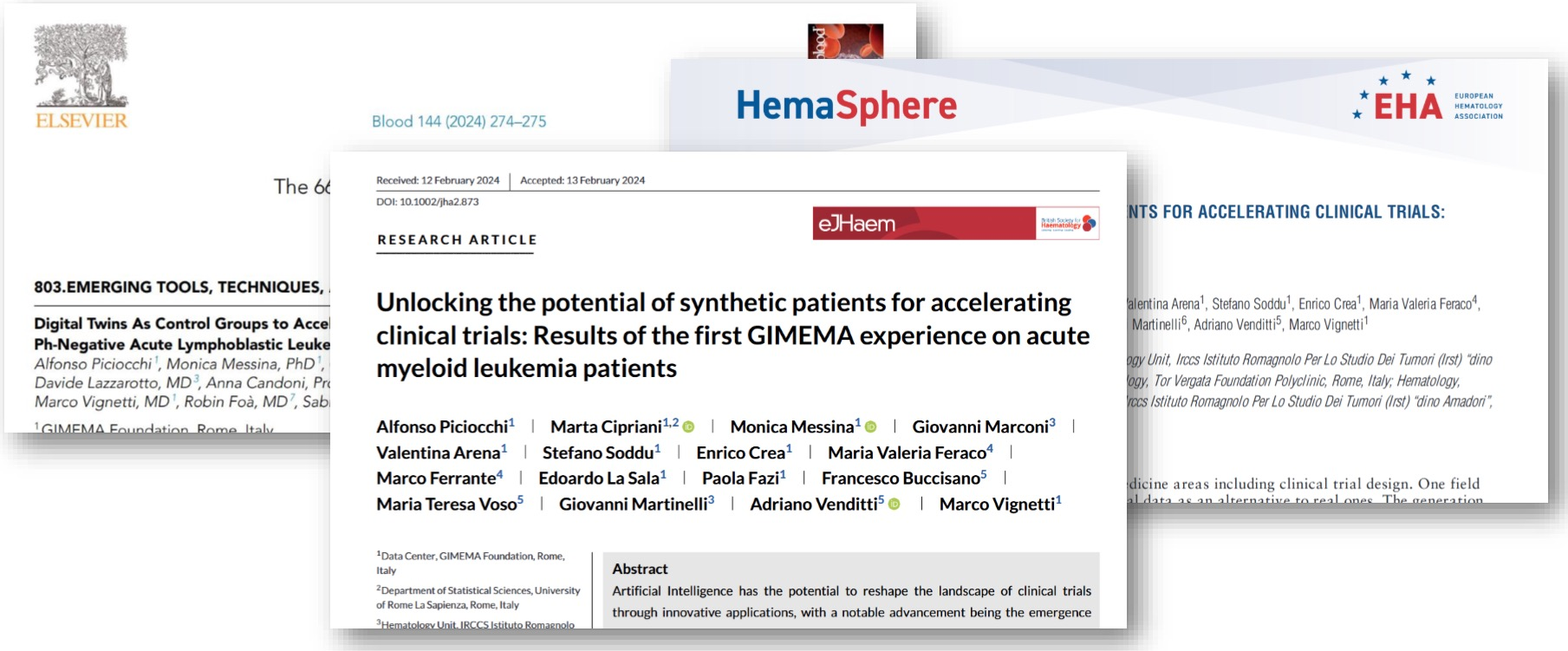


CREATE A FUNCTIONAL COPY OF THE REAL DATA

ACCELERATE TRANSITIONAL RESEARCH AND IMPROVE PROGNOSTIC MODEL

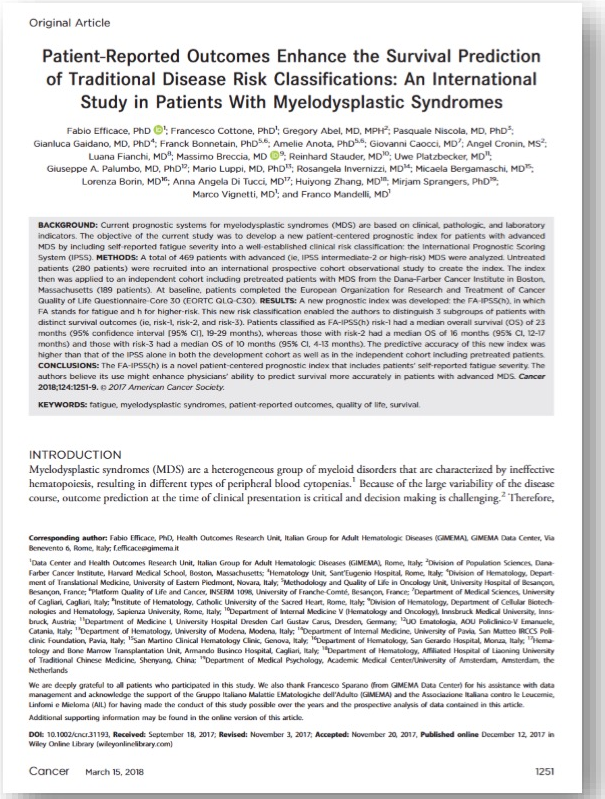
ACCELERATE CLINICAL RESEARCH AND DESIGN OF CLINICAL TRIALS

GIMEMA's experience so far...



Synthesizing Patient Reported Outcomes: is it possible?

PROMYS: MDS Study Overview



Original Article

Patient-Reported Outcomes Enhance the Survival Prediction of Traditional Disease Risk Classifications: An International Study in Patients With Myelodysplastic Syndromes

Fabio Efficace, PhD[1]; Francesco Cottone, PhD[1]; Gregory Abel, MD, MPH[2]; Pasquale Niscola, MD, PhD[3]; Gianluca Gaidano, MD, PhD[4]; Franck Bonnetain, PhD[5,6]; Amelie Anota, PhD[5,6]; Giovanni Caocci, MD[7]; Angel Cronin, MS[2]; Luana Fianchi, MD[8]; Massimo Breccia, MD[9]; Reinhard Stauder, MD[10]; Uwe Platzbecker, MD[11]; Giuseppe A. Palumbo, MD, PhD[12]; Mario Luppi, MD, PhD[13]; Rosangela Invernizzi, MD[14]; Micaela Bergamaschi, MD[15]; Lorenza Borin, MD[16]; Anna Angela Di Tucci, MD[17]; Huiyong Zhang, MD[18]; Mirjam Sprangers, PhD[19]; Marco Vignetti, MD[1]; and Franco Mandelli, MD[1]

BACKGROUND: Current prognostic systems for myelodysplastic syndromes (MDS) are based on clinical, pathologic, and laboratory indicators. The objective of the current study was to develop a new patient-centered prognostic index for patients with advanced MDS by including self-reported fatigue severity into a well-established clinical risk classification: the International Prognostic Scoring System (IPSS). **METHODS:** A total of 469 patients with advanced (ie, IPSS intermediate-2 or high-risk) MDS were analyzed. Untreated patients (280 patients) were recruited into an international prospective cohort observational study to create the index. The index then was applied to an independent cohort including pretreated patients with MDS from the Dana-Farber Cancer Institute in Boston, Massachusetts (189 patients). At baseline, patients completed the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30 (EORTC QLQ-C30). **RESULTS:** A new prognostic index was developed: the FA-IPSS(h), in which FA stands for fatigue and h for higher-risk. This new risk classification enabled the authors to distinguish 3 subgroups of patients with distinct survival outcomes (ie, risk-1, risk-2, and risk-3). Patients classified as FA-IPSS(h) risk-1 had a median overall survival (OS) of 23 months (95% confidence interval [95% CI], 19-29 months), whereas those with risk-2 had a median OS of 16 months (95% CI, 12-17 months) and those with risk-3 had a median OS of 10 months (95% CI, 4-13 months). The predictive accuracy of this new index was higher than that of the IPSS alone in both the development cohort as well as in the independent cohort including pretreated patients. **CONCLUSIONS:** The FA-IPSS(h) is a novel patient-centered prognostic index that includes patients' self-reported fatigue severity. The authors believe its use might enhance physicians' ability to predict survival more accurately in patients with advanced MDS. ***Cancer*** **2018;124:1251-9,** © *2017 American Cancer Society.*

KEYWORDS: fatigue, myelodysplastic syndromes, patient-reported outcomes, quality of life, survival.

INTRODUCTION

Myelodysplastic syndromes (MDS) are a heterogeneous group of myeloid disorders that are characterized by ineffective hematopoiesis, resulting in different types of peripheral blood cytopenias.[1] Because of the large variability of the disease course, outcome prediction at the time of clinical presentation is critical and decision making is challenging.[2] Therefore,

Corresponding author: Fabio Efficace, PhD, Health Outcomes Research Unit, Italian Group for Adult Hematologic Diseases (GIMEMA), GIMEMA Data Center, Via Benevento 6, Rome, Italy; f.efficace@gimema.it

[1]Data Center and Health Outcomes Research Unit, Italian Group for Adult Hematologic Diseases (GIMEMA), Rome, Italy; [2]Division of Population Sciences, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts; [3]Hematology Unit, Sant'Eugenio Hospital, Rome, Italy; [4]Division of Hematology, Department of Translational Medicine, University of Eastern Piedmont, Novara, Italy; [5]Methodology and Quality of Life in Oncology Unit, University Hospital of Besançon, Besançon, France; [6]Platform Quality of Life and Cancer, INSERM 1098, University of Franche-Comté, Besançon, France; [7]Department of Medical Sciences, University of Cagliari, Cagliari, Italy; [8]Institute of Hematology, Catholic University of the Sacred Heart, Rome, Italy; [9]Division of Hematology, Department of Cellular Biotechnologies and Hematology, Sapienza University, Rome, Italy; [10]Department of Internal Medicine V (Hematology and Oncology), Innsbruck Medical University, Innsbruck, Austria; [11]Department of Medicine I, University Hospital Dresden Carl Gustav Carus, Dresden, Germany; [12]UO Ematologia, AOU Policlinico-V Emanuele, Catania, Italy; [13]Department of Hematology, University of Modena, Modena, Italy; [14]Department of Internal Medicine, University of Pavia, San Matteo IRCCS Policlinic Foundation, Pavia, Italy; [15]San Martino Clinical Hematology Clinic, Genova, Italy; [16]Department of Hematology, San Gerardo Hospital, Monza, Italy; [17]Hematology and Bone Marrow Transplantation Unit, Armando Businco Hospital, Cagliari, Italy; [18]Department of Hematology, Affiliated Hospital of Liaoning University of Traditional Chinese Medicine, Shenyang, China; [19]Department of Medical Psychology, Academic Medical Center/University of Amsterdam, Amsterdam, the Netherlands

We are deeply grateful to all patients who participated in this study. We also thank Francesco Sparano (from GIMEMA Data Center) for his assistance with data management and acknowledge the support of the Gruppo Italiano Malattie EMatologiche dell'Adulto (GIMEMA) and the Associazione Italiana contro le Leucemie, Linfomi e Mieloma (AIL) for having made the conduct of this study possible over the years and the prospective analysis of data contained in this article.

Additional supporting information may be found in the online version of this article.

DOI: 10.1002/cncr.31193, **Received:** September 18, 2017; **Revised:** November 3, 2017; **Accepted:** November 20, 2017, **Published online** December 12, 2017 in Wiley Online Library (wileyonlinelibrary.com)

Background

Current systems for MDS are based on clinical and laboratory indicators.

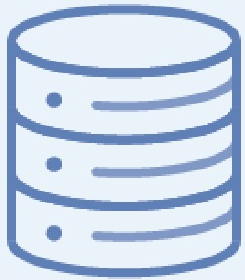
Methods

927 patients enrolled and a new index was created: the **FA-IPPS-R**.

Results

The FA-IPSS-R is **highly prognostic,** identifying three subgroups with different survival outcomes.

How do we use synthetic data?



CREATE A FUNCTIONAL COPY OF THE REAL DATA

ACCELERATE TRANSITIONAL RESEARCH AND IMPROVE PROGNOSTIC MODEL

ACCELERATE CLINICAL RESEARCH AND DESIGN OF CLINICAL TRIALS

PROMYS: Replicating Real-World Data

SYNTHETIC COHORT

Generated 5,000 synthetic patients preserving clinical, demographic, and **patient-reported outcomes** like fatigue.

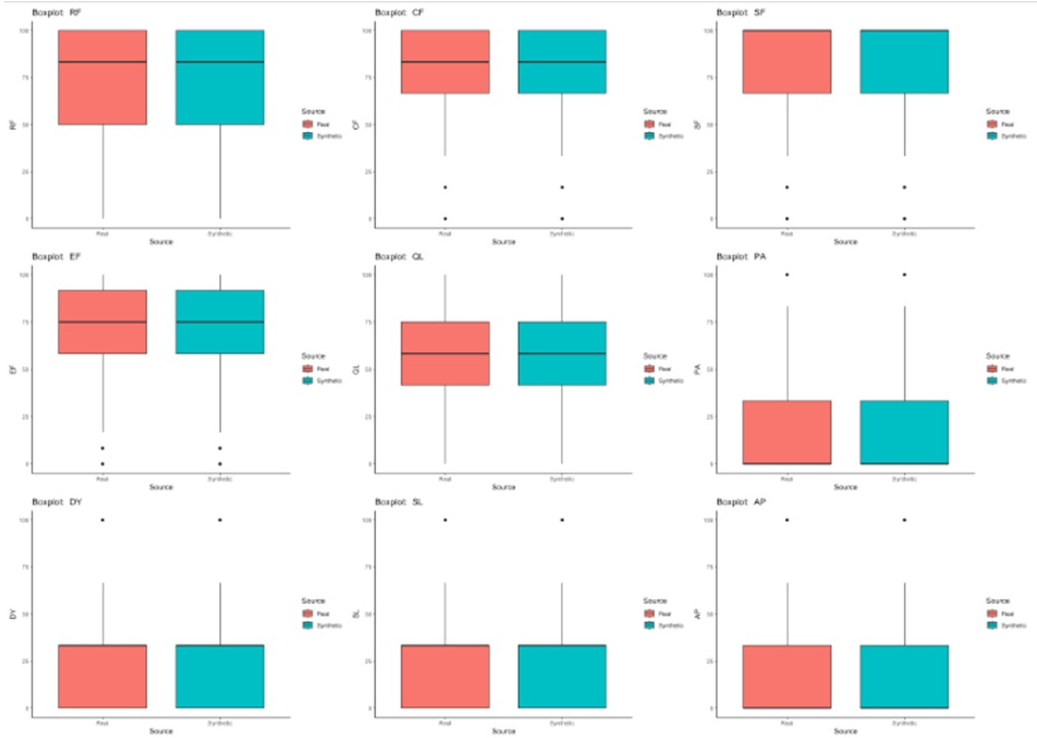


Figure. QLQ-C30 variables

PROMYS: Replicating Real-World Data

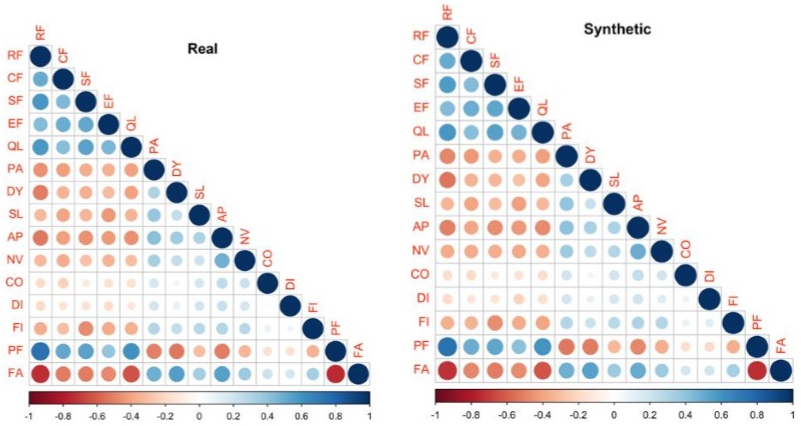


Figure. Pearson correlation matrix among PRO measures

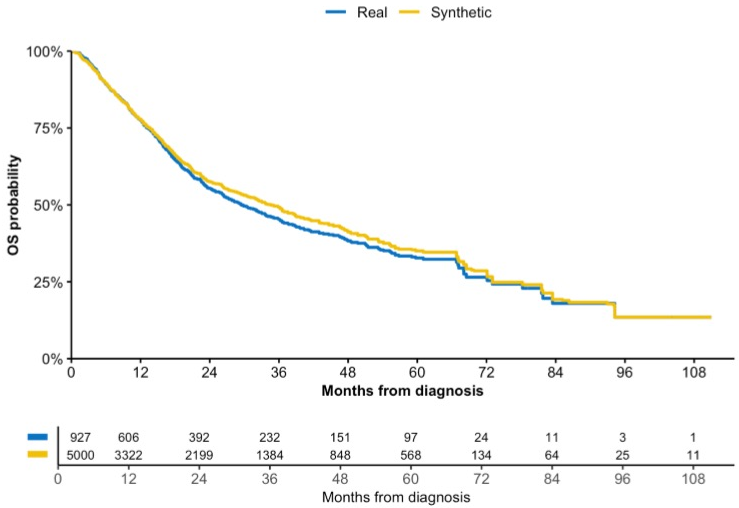
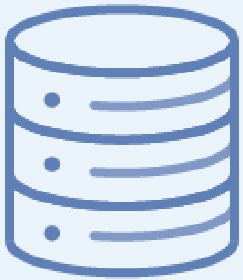


Figure. Overall survival

How do we use synthetic data?



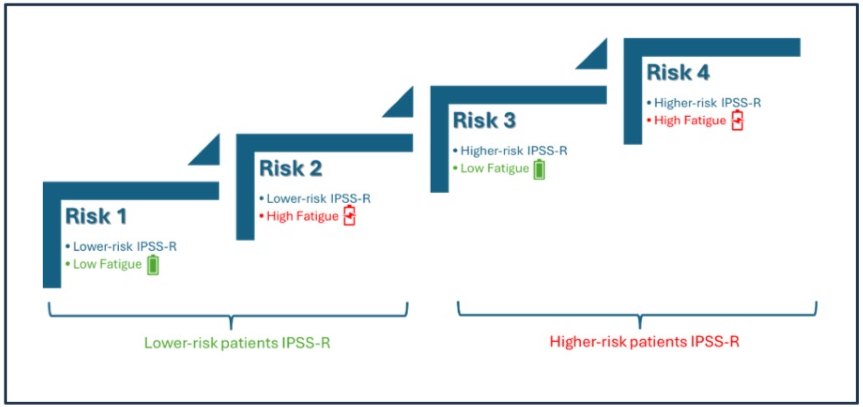
CREATE A FUNCTIONAL COPY OF THE REAL DATA

ACCELERATE TRANSITIONAL RESEARCH AND IMPROVE PROGNOSTIC MODEL

ACCELERATE CLINICAL RESEARCH AND DESIGN OF CLINICAL TRIALS

PROMYS: Hybrid validation of FA-IPSS-R

We introduce a novel synthetic data approach incorporating PRO data to enable **internal validation** of the FA-IPSS-R index, through a 5000-patient **synthetic cohort** and using Kaplan-Meier estimates, log-rank tests, and Cox models.



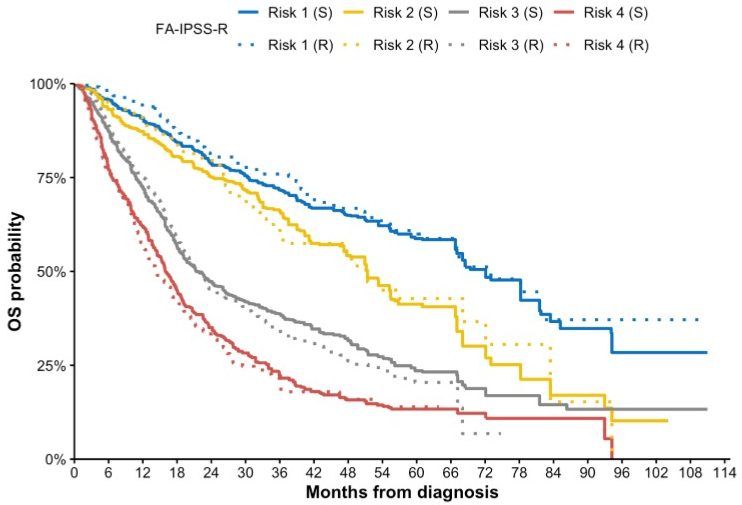
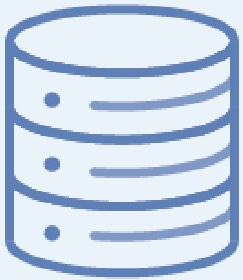


Figure. Overall survival stratified by FA-IPSS-R groups

Abbreviations: FA-IPSS-R= Fatigue Revised International Prognostic Scoring System; IPSS-R= Revised International Prognostic Scoring System. For IPSS-R cutoff is 3.5; for fatigue cutoff is 35 for low-risk patients and 45 for high-risk patients.

How do we use synthetic data?



CREATE A FUNCTIONAL COPY OF THE REAL DATA

ACCELERATE TRANSITIONAL RESEARCH AND IMPROVE PROGNOSTIC MODEL

ACCELERATE CLINICAL RESEARCH AND DESIGN OF CLINICAL TRIALS

Virtual Cohort Study – Luspatercept in Elderly MDS

BACKGROUND (MEDALIST trial)

- Proven efficacy (↓ transfusions, ↑ Hb)
- Gap: limited real-world PRO/HRQoL data (fatigue)

STUDY DESIGN

- **Exposed**: MDS pts on luspatercept
- **Synthetic control:** PROMYS cohort (N=5,000, matched)

OBJECTIVES

- **Primary:** Fatigue (PRO)
- **Secondary:** HRQoL, transfusion independence, Hb, response duration, safety

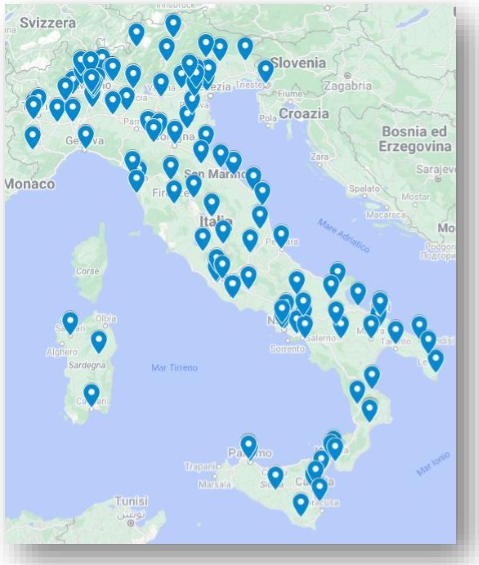
Pros and Cons

- ANONYMIZATION
- RESEARCH REPRODUCIBILITY AND DATA ACCESSIBILITY
- DATA AUGMENTATION
- RESOURCE OPTIMIZATION

- REGULATORY HURDLES
- REPRESENTATIVENESS OF THE SAMPLE

fondazione GIMEMA onlus
per la promozione e lo sviluppo della ricerca scientifica sulle malattie ematologiche. FRANCO MANDELLI

Thank you for your attention!



Alfonso Piciocchi
Monica Messina
Giovanni Marsili
Valentina Arena
Stefano Soddu
Maria Valeria Feraco
Marco Ferrante
Edoardo la Sala
Serena Puglia
Fabio Efficace
Paola Fazi
Marco Vignetti

Sabina Chiaretti
Davide Lazzarotto
Anna Candoni
Alessandro Rambaldi
Davide Lazzarotto
Robin Foà
Giovanni Marconi
Francesco Buccisano
Maria Teresa Voso
Giovanni Martinelli
Massimo Breccia
Adriano Venditti