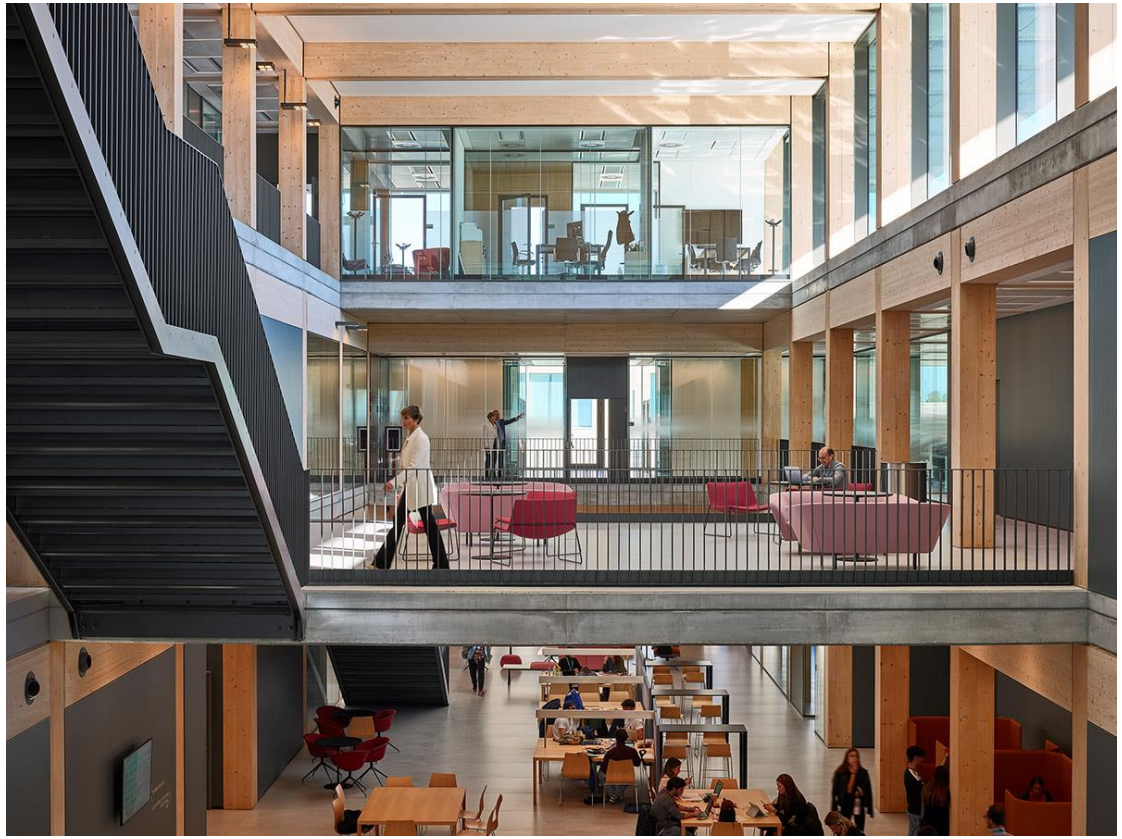
Applications of generative AI in Personalized Medicine

Matteo G Della Porta

Humanitas Research Hospital, Milan IT

HUMANITAS AI CENTER



TEAM:

- Saverio D'Amico, **Data Scientist**
- Elisabetta Sauta, **Bioinformatics Scientist**
- Gianluca Asti, **Computer Vision Specialist**
- Elena Zazzetti, **Data Scientist**
- Mattia Delleani, **Computer Vision Specialist**
- Marilena Bicchieri, **Project Manager**
- Victor Savevski, **Chief Innovation Officer Humanitas**

Artificial Intelligence (AI) for personalized medicine in Hematology

1- Machine Learning

2- Generative AI

HUMANITAS
RESEARCH HOSPITAL
AI CENTER

synthia

SYNTHEMA

IRCCS
HUMANITAS
RESEARCH HOSPITAL

EuroBloodNet

Realise RD

compRehensive mEthodological and operational Approach to cLinical trIalS in ultra-rarE Diseases

GEN MED 4ALL

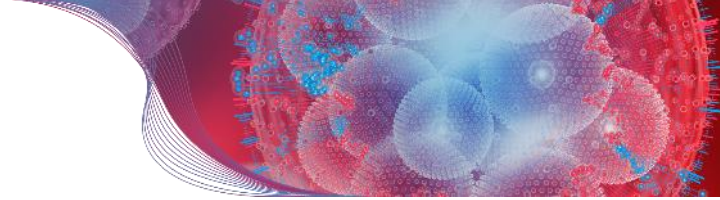
2021 WHO guidance on ethics and governance of AI for health

We have to address three important topics for a **right deployment of AI in hematology**:

- **Transparency of models.** We have to provide a good understanding of the models (interpretability and explainability)
- **Reliability of models.** The main vulnerabilities of AI models are related to lack of generalizability. Therefore, extensive, independent validation of generated AI-models is required.
- **Protection of data and data sharing**. Innovative technologies such as federated learning procedures for data collection and analysis (without moving sensitive medical data from their original locations) are required to facilitate clinical implementability of AI solutions

1. *The World Health Organization. 2021 WHO guidance on ethics and governance of artificial intelligence for health.* https://www.who.int/publications/i/item/9789240029200

Increased access to healthcare data is needed to accelerate the generation of clinical evidence in hematology



97%

of healtcare data remains unused

Source: Deloitte, Health Data, 2023

Main reasons:

- Privacy limitations (GDPR)
- Lack of data harmonization from different sources
- Data are not structured and are dispersed

Synthetic Data to accelerate research in Hematology

- **Synthetic data** are artificial data generated by an algorithm trained to learn all the essential characteristics of a real dataset. The new data are neither a copy nor a representation of the real data. Since they are not real data, they are not regulated by particular limitations so they can be easily accessed and shared.

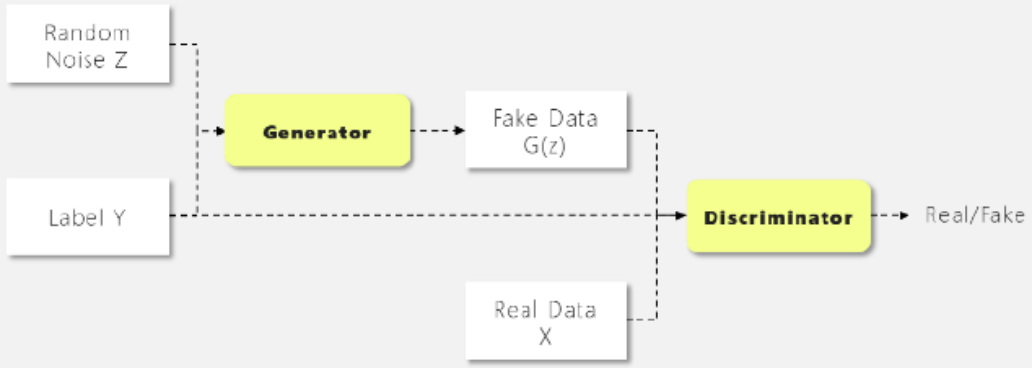


Figure 2 Conditional Generative Adversarial Networks architecture

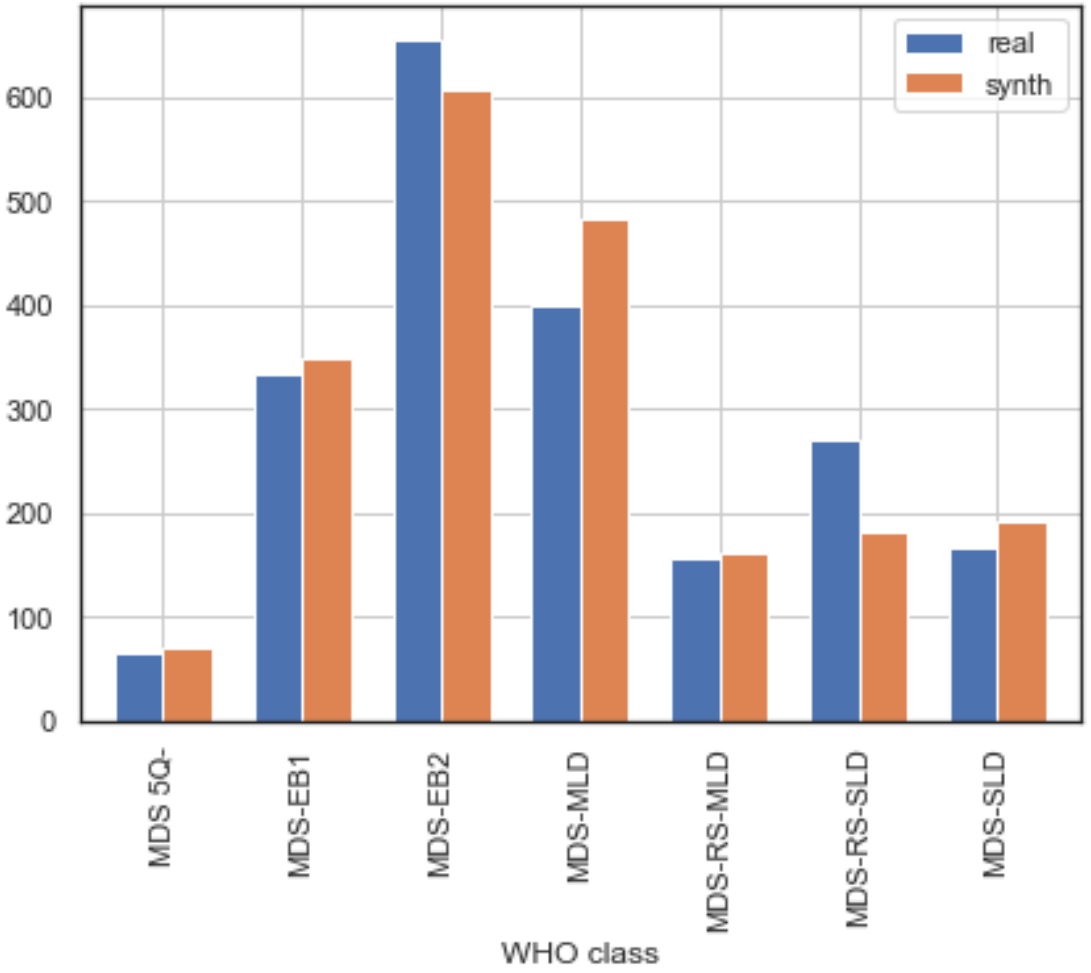
Properties and possible applications

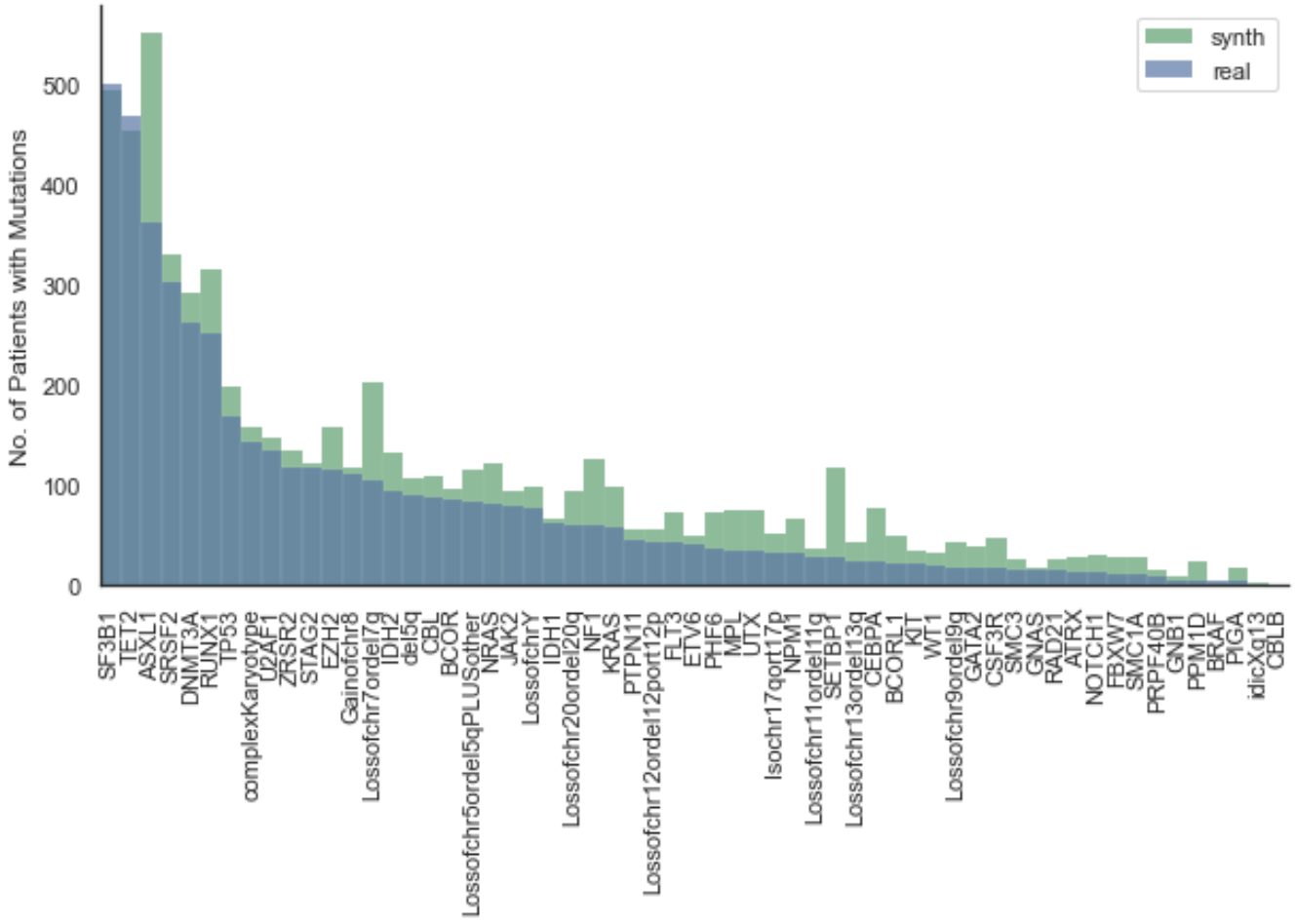
- Data sharing (GDPR)
- Classes balance and resolution of missing information
- Data augmentation
- Algorithms training and validation
- Generation of new evidence

Chen R et al. Nature Biomedical Engineering 2021;5:493–497 (2021)

HUMANITAS RESEARCH HOSPITAL AI CENTER

Synthetic vs. Real Data: comparison of clinical and molecular features in MDS



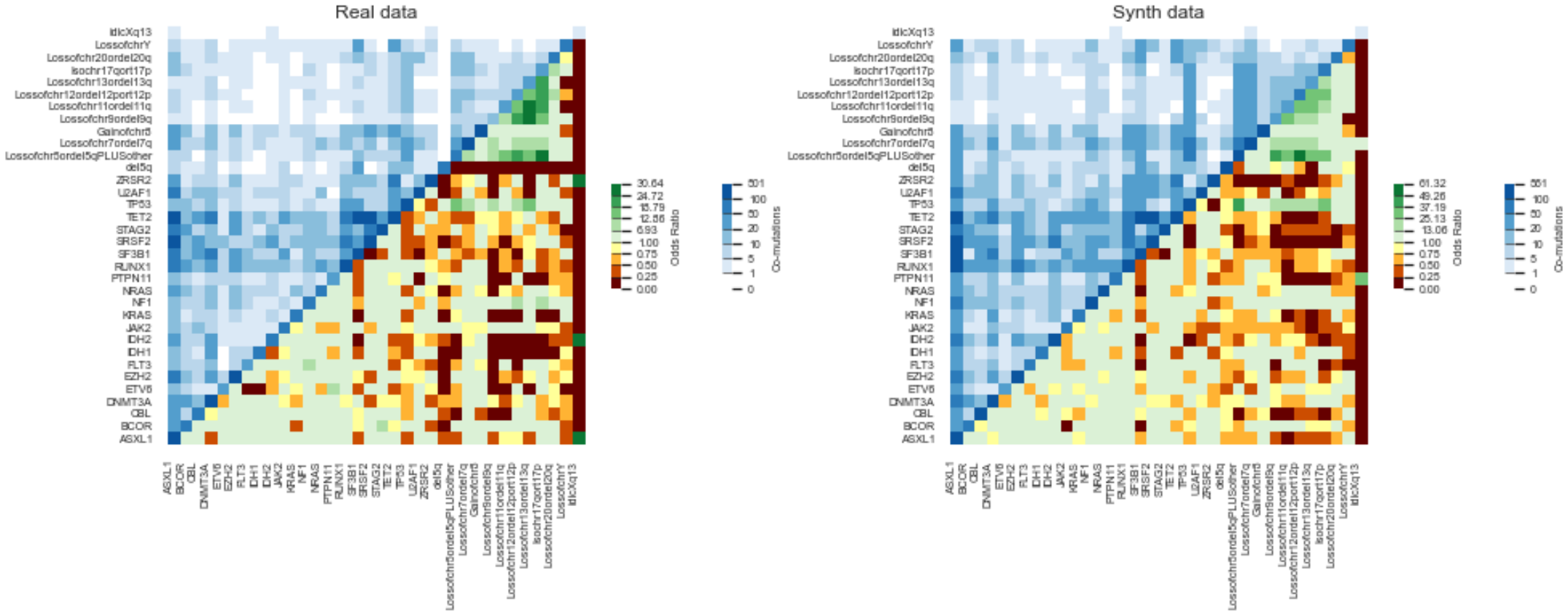


Jacobs F et al. Journal of Clin Oncol CCI 2023 Aug;7:e2300045
D'Amico S et al. Journal of Clin Oncol CCI 2023 Jun;7:e2300021
Bersanelli M et al. Journal of Clinical Oncol 2021; 11:1223-1233

HUMANITAS RESEARCH HOSPITAL AI CENTER

Synthetic vs. Real Data: comparison of clinical and molecular features in MDS

Pairwise associations among genes and cytogenetic abnormalities



Jacobs F et al. Journal of Clin Oncol CCI 2023 Aug;7:e2300045
D'Amico S et al. Journal of Clin Oncol CCI 2023 Jun;7:e2300021
D'Amico S, et al. Journal of Clin Oncol CCI 2024, in press

HUMANITAS RESEARCH HOSPITAL AI CENTER

Synthetic vs. Real Data: comparison of clinical and molecular features in MDS

COX models (overall survival)

Real data:

Global Concordance: 0.779; Std.err:0.013

Partial Concordance of risk components:

	Clinical	CNA	Demographics	Genetics
concordant	0.711	0.569	0.630	0.782
std(c-d)	0.013	0.011	0.013	0.013

Synthetic data:

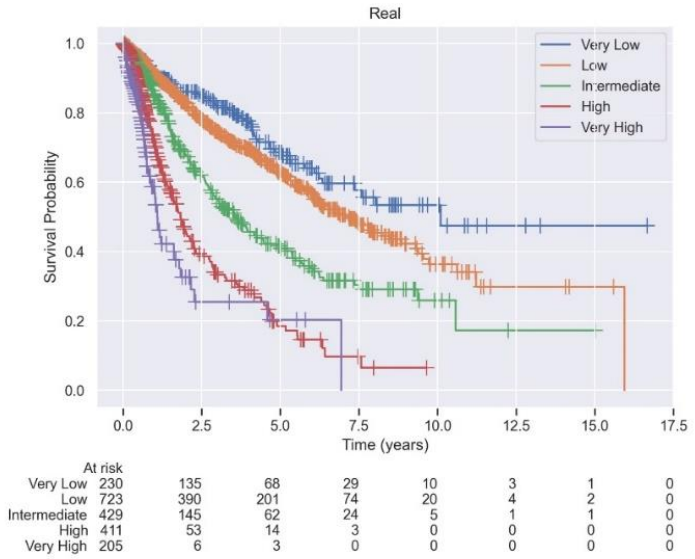
Global Concordance: 0.822; Std.err:0.013

Partial Concordance of risk components:

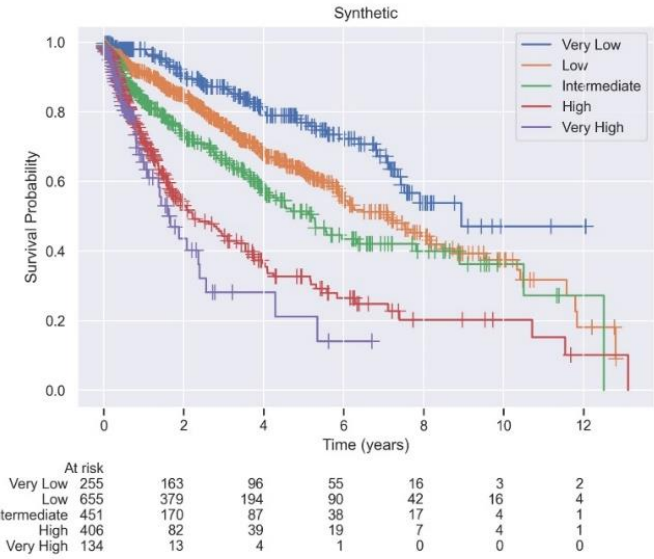
	Clinical	CNA	Demographics	Genetics
concordant	0.732	0.536	0.646	0.746
std(c-d)	0.013	0.011	0.013	0.013

Probability of OS stratified by IPSS-R

REAL DATA



SYNTHETIC DATA

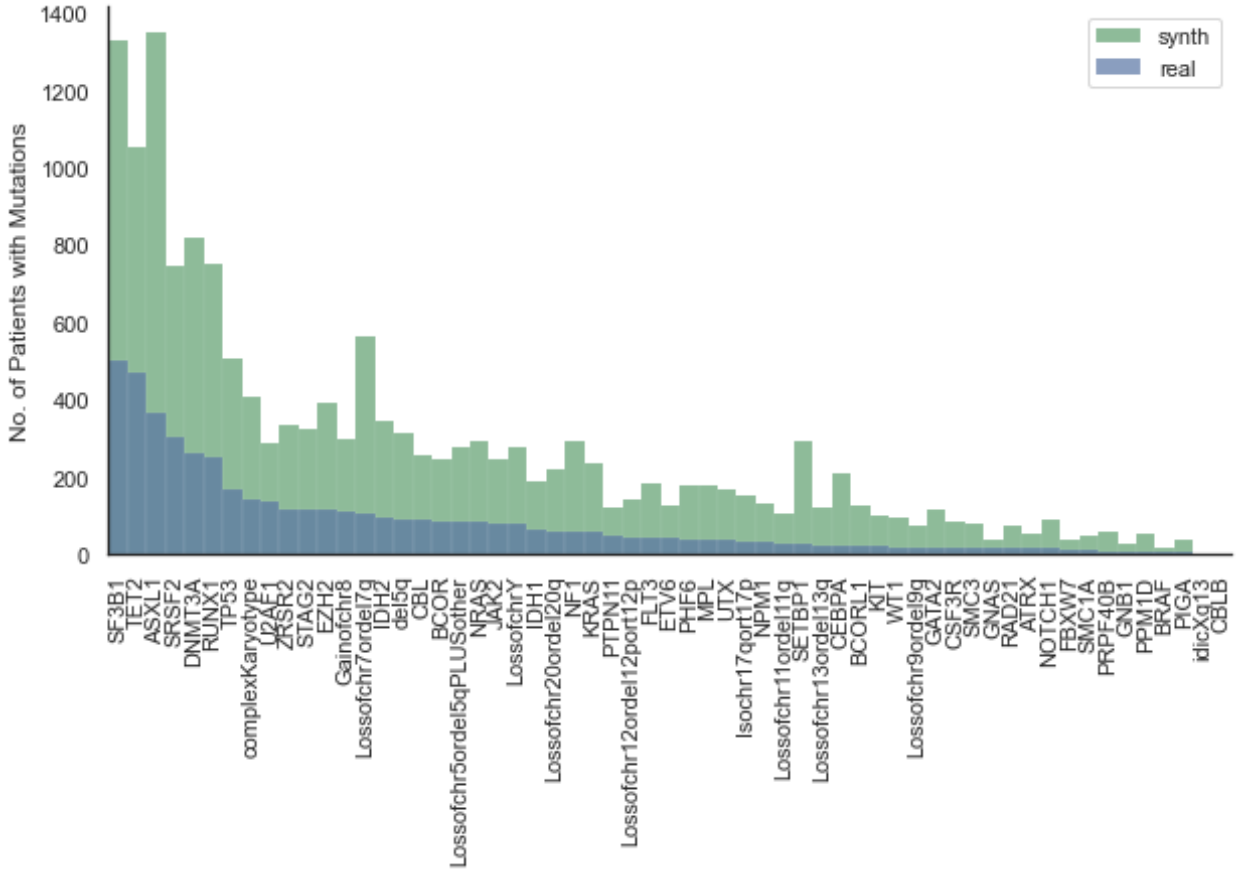


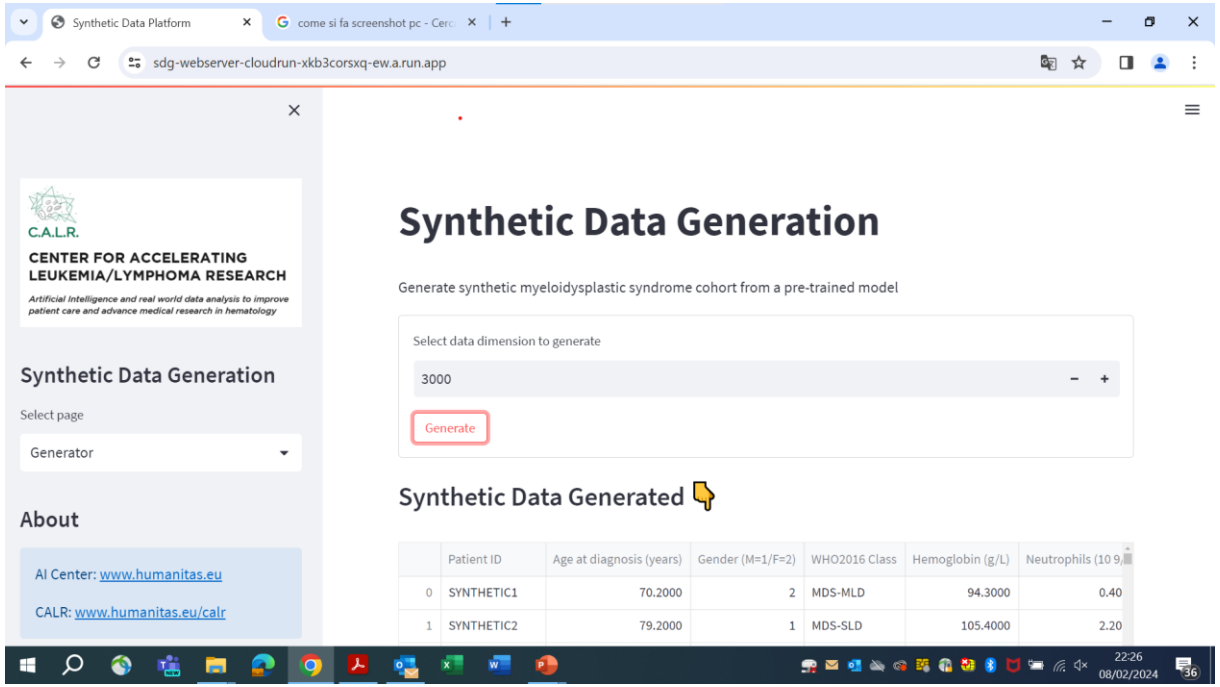
Jacobs F et al. Journal of Clin Oncol CCI 2023 Aug;7:e2300045
D'Amico S et al. Journal of Clin Oncol CCI 2023 Jun;7:e2300021
D'Amico S, et al. Journal of Clin Oncol CCI 2024, in press

HUMANITAS RESEARCH HOSPITAL AI CENTER

Synthetic vs. Real Data: comparison of clinical and molecular features in MDS

Data augmentation: from 2043 to 5000 patients





Jacobs F et al. Journal of Clin Oncol CCI 2023 Aug;7:e2300045
D'Amico S et al. Journal of Clin Oncol CCI 2023 Jun;7:e2300021
D'Amico S, et al. Journal of Clin Oncol CCI 2024, in press

HUMANITAS RESEARCH HOSPITAL AI CENTER

Performance of Synthetic Data

DEMOGRAPHIC, CLINICAL AND SURVIVAL DATA

92.1 %

SYNTHETIC CLINICAL FITNESS

Evaluated with distribution plot, Principal Component Analysis and correlation matricies.

GENOMIC DATA

90.2 %

SYNTHETIC GENOMIC FITNESS

Evaluated with mutation frequencies and pairwise association.

ALL DATA

70.6 %

PRIVACY PRESERVABILITY

Evaluated considering the possibility of tracing real data from synthetic ones.

Jacobs F et al. Journal of Clin Oncol CCI 2023 Aug;7:e2300045
D'Amico S et al. Journal of Clin Oncol CCI 2023 Jun;7:e2300021
D'Amico S, et al. Journal of Clin Oncol CCI 2024, in press

HUMANITAS RESEARCH HOSPITAL AI CENTER

Generation of Synthetic Data to accelerate clinical research in Hematology

Comparing endpoints of clinical trials using **real** and **synthetic** control arms. Real-world efficacy and safety of luspatercept in adult patients with transfusion-dependent anemia due to very low-, low and intermediate-risk myelodysplastic syndrome (MDS) with ring sideroblasts, who had an unsatisfactory response to or are ineligible for erythropoietin-based therapy: a multicenter study by Fondazione Italiana Sindromi Mielodisplastiche (FISIM)

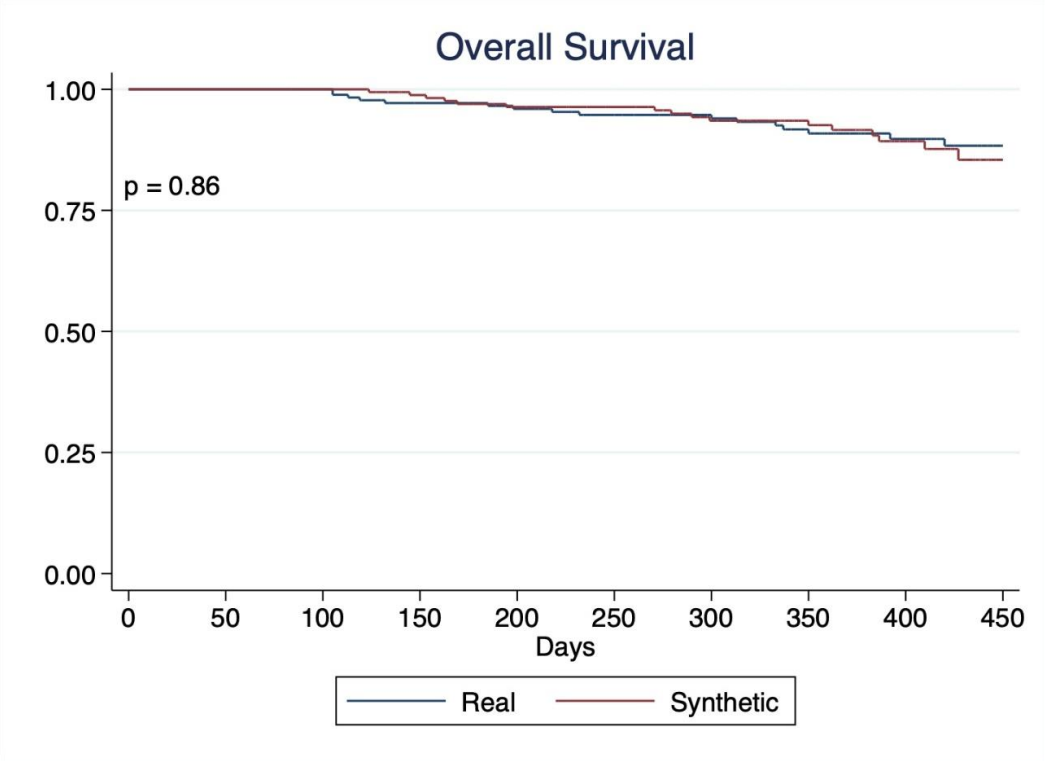
Primary endpoints

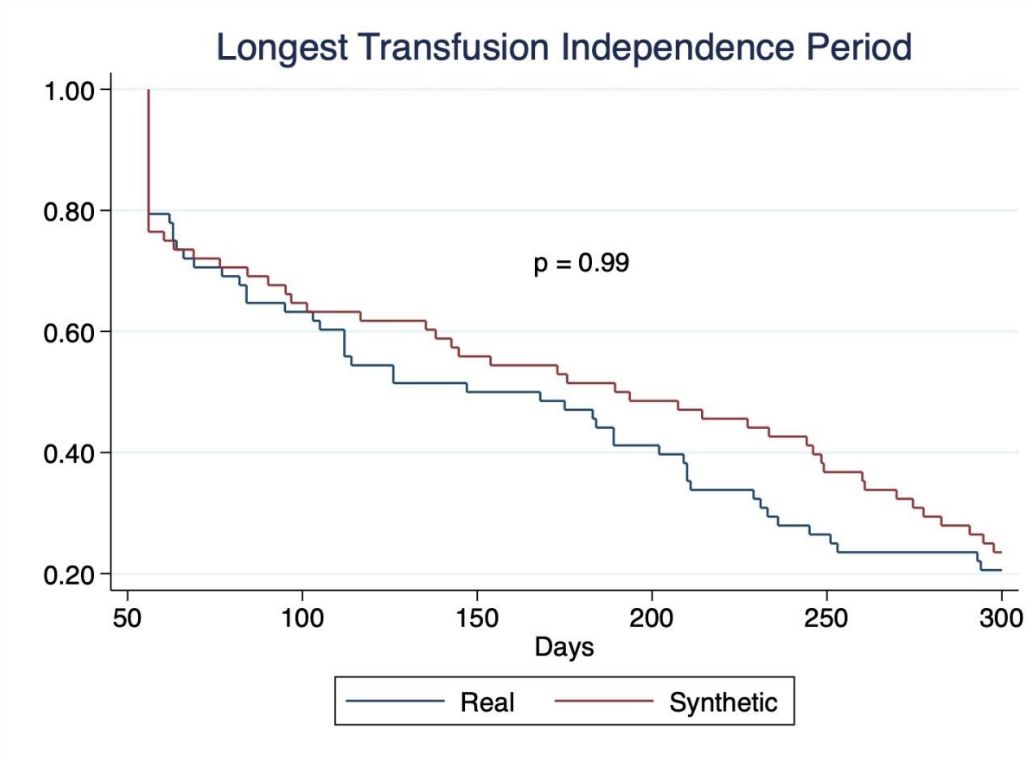
	Real data	Synthetic data	Pvalue
RBC-TI>=8 weeks 1-24	56 (31,5)	56 (31,5)	1.0
Longest Transfusion Independence Period (weeks), median (range)	195 (56-490)	191 (56-490)	0.34
RBC-TI>=8 weeks 1-48	68 (38,2)	61 (34,3)	0.50
RBC-TI>=12 weeks 1-24	36 (20,2)	41 (23,0)	0.60
RBC-TI>=12 weeks 1-48	51 (28,7)	46 (25,8)	0.63
Reduction>= 4 RBC	62 (34,8)	63 (35,4)	1.0
Reduction>=50%	77 (43,3)	72 (40,4)	0.66
AML Evolution	4 (2,2)	6 (3,4)	0.75
Discontinued patients	74 (41,6)	82 (46,1)	0.64

Jacobs F et al. Journal of Clin Oncol CCI 2023 Aug;7:e2300045
D'Amico S et al. Journal of Clin Oncol CCI 2023 Jun;7:e2300021
D'Amico S, et al. Journal of Clin Oncol CCI 2024, in press

HUMANITAS RESEARCH HOSPITAL AI CENTER

Generation of Synthetic Data to accelerate clinical research in Hematology





Comparing endpoints of clinical trials using **real** and **synthetic** control arms. Real-world efficacy and safety of luspatercept in adult patients with transfusion-dependent anemia due to very low-, low and intermediate-risk myelodysplastic syndrome (MDS) with ring sideroblasts, who had an unsatisfactory response to or are ineligible for erythropoietin-based therapy: a multicenter study by Fondazione Italiana Sindromi Mielodisplastiche (FISIM)

HUMANITAS RESEARCH HOSPITAL AI CENTER

Synthetic Data in clinical trials: Status of regulation and next plans

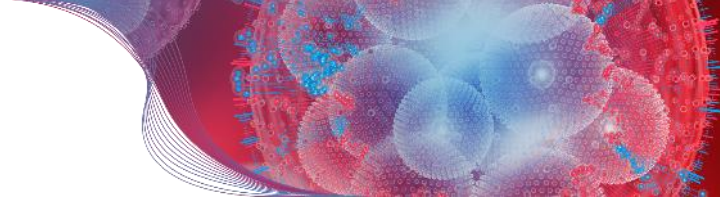
Product and Indication	Pivotal Data	Context for use of patient registries
Axicabtagene ciloleucel (Yescarta™)	Open label, single arm study (ZUMA 1 Phase II) with a primary endpoint of objective response rate defined as complete remission (CR) or partial remission (PR).	A retrospective patient level pooled analysis of two Phase III RCTs and two observational studies (SCHOLAR 1) was developed as a companion study to contextualise the results of ZUMA 1.
Tisagenlecleucel (Kymriah™)	Open-label single arm study (C2201) with a primary endpoint of overall response rate defined as the proportion of patients with CR or PR.[2]	Further long term follow up of efficacy will be captured via a prospective observational study in patients with refractory/relapsed DLBCL based on data from registry with efficacy outcomes similar to study C2201.
Genetically modified allogeneic T cells (Zalmoxis™)	Single arm, open-label Phase I/II study with a primary endpoint of immune reconstitution and an ongoing open label randomised controlled Phase III trial.[3]	The European Bone Marrow Transplantation (EBMT) patient registry was used to compile an appropriate control group selected on same criteria as the control arm of the on-going Phase III trial and a specific set of matching parameters.

HUMANITAS RESEARCH HOSPITAL AI CENTER

The opportunity of EMA qualification registry initiative

- The EMA qualification registry initiative represents a unique opportunity to improve the availability of high-quality data for regulatory purposes.
- There is an increasing access to EMA qualification procedure by registry platforms, and relevant clinical networks/registries for hematological diseases have already or are interested to join the procedure (including the European Society for Blood and Marrow Transplantation (EMBT) registry, https://www.ebmt.org/, EuroBloodNET, the EU reference network for rare hematological diseases, www.eurobloodnet.eu, and Harmony Alliance www.harmony-alliance.eu)
- WEBSITE: https://www.ema.europa.eu/en/human-regulatory-overview/post-authorisation/patient-registries

Synthetic Data in clinical trials: Status of regulation and next plans



- **Alectinib** obtained conditional EU approval as a treatment for lung cancer in 2017, with acceptance of a synthetic control arm of 67 patients as a trial, thus accelerating drug's availability in the EU by 18 months

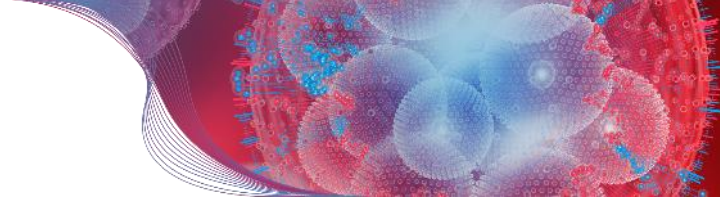
- **Avelumab** for the treatment of Merkel cell carcinoma, approved in 2018, used data from electronic medical records in a synthetic control arm

- Accelerated approval was obtained for **Blinatumumab** for the treatment of leukemia by the FDA in 2014 and the EMA in 2015, using a comparator arm of historical data from 694 patients based on 2,000 patient records for the phase 2 study

- FDA and the scientific community are forming an alliance for the **no-placebo initiative** to use external comparison arms to study new therapies for gastrointestinal stromal tumor and other rare cancers, facilitating drug trials and regulatory approvals.

- **Synthema Synthia and Realise-D,** as a European consortia are building a Synthetic Validation Framework (SVF) for privacy, utility and clinical evaluation of synthetic data in healthcare.

HUMANITAS RESEARCH HOSPITAL AI CENTER



Correspondence

https://doi.org/10.1038/s41591-023-02488-0

A synthetic control arm for refractory metastatic colorectal cancer: the no placebo initiative

nature medicine

Volume 29 | October 2023 | 2389–2390 | **2389**

HUMANITAS RESEARCH HOSPITAL AI CENTER

Synthetic control arms for clinical trials

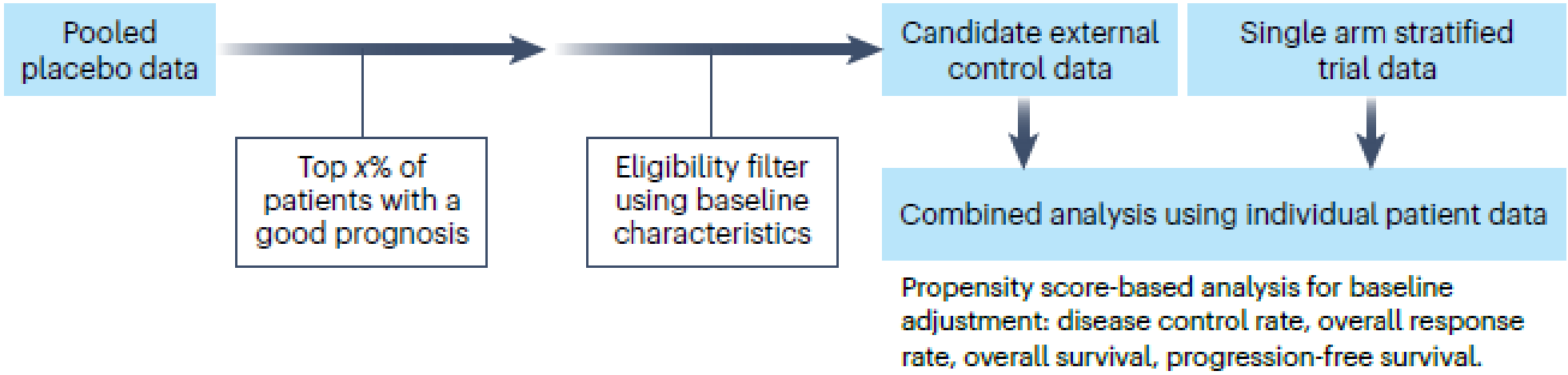


Fig. 1 | Three-step analysis for no-placebo initiative. First, participants enrolled in trials with placebo arms will be selected based on compatible patient demographics and key characteristics. These data will form the synthetic control arm. Second, patients in the top percentile for overall survival will be extracted from the synthetic control arm. Third, the synthetic control arm will be compared with patients in the trial, using propensity scored-based analysis.

HUMANITAS RESEARCH HOSPITAL AI CENTER

SYNTHEMA

White Paper on synthetic data for scientific medical research – active discussion with EMA

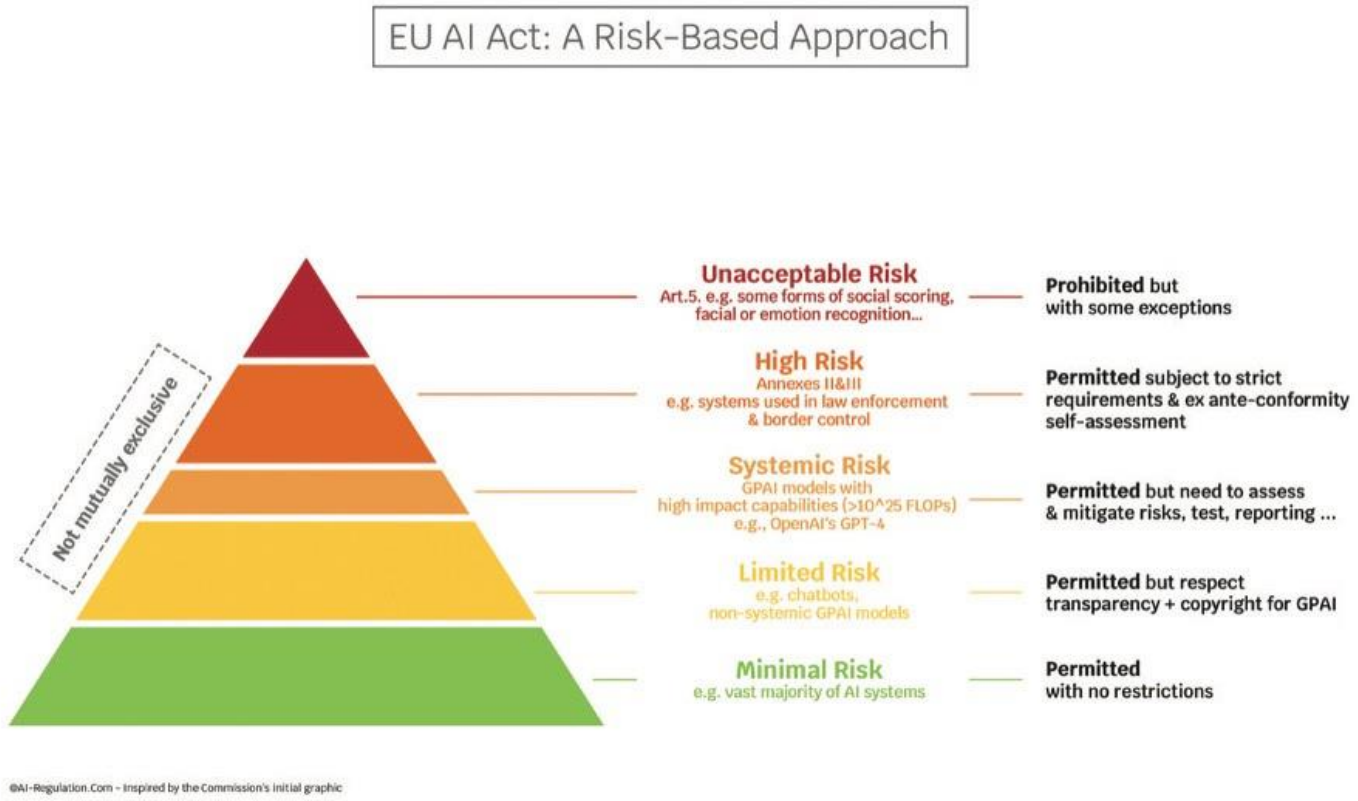
HARMONY

The AI Act

Synthetic data in the Ai Act

Data Governance obligations for High Risks Systems:
Art. 10(5) lett. a: "(...) the bias detection and correction cannot be effectively fulfilled by processing other data, including synthetic or anonymised data;"

Further Processing of Personal Data for Developing Certain AI Systems in the Public Interest in the AI Regulatory Sandbox:
Art. 59(1) lett. b: " the data processed are necessary for complying with one or more of the requirements referred to in Chapter III, Section 2 where those requirements cannot be effectively fulfilled by processing anonymised, synthetic or other non-personal data;"



Synthetic Data for Clinical Decision Support Systems (CDSS)

Clinical Decision Support Systems are computer-based tools that assist clinicians in making decisions by providing them with real-time, patient-specific information and recommendations

Benefits:

- Increase patient safety (potential drug interactions, dosing errors, etc.)
- Improve quality and efficiency of care
- Update clinicians about latest guidelines and research
- Reduce costs and optimize expenses

Vasey B et al. Nat Med 28:924-933, 2022

HUMANITAS RESEARCH HOSPITAL AI CENTER

Clinical and genomic-based Decision Support System to define the optimal timing of allogeneic hematopoietic stem cell transplantation in MDS

- To develop and validate a Decision Support System (DSS) to define the optimal timing of HSCT in MDS patients based on clinical and genomic information as provided by IPSS-M

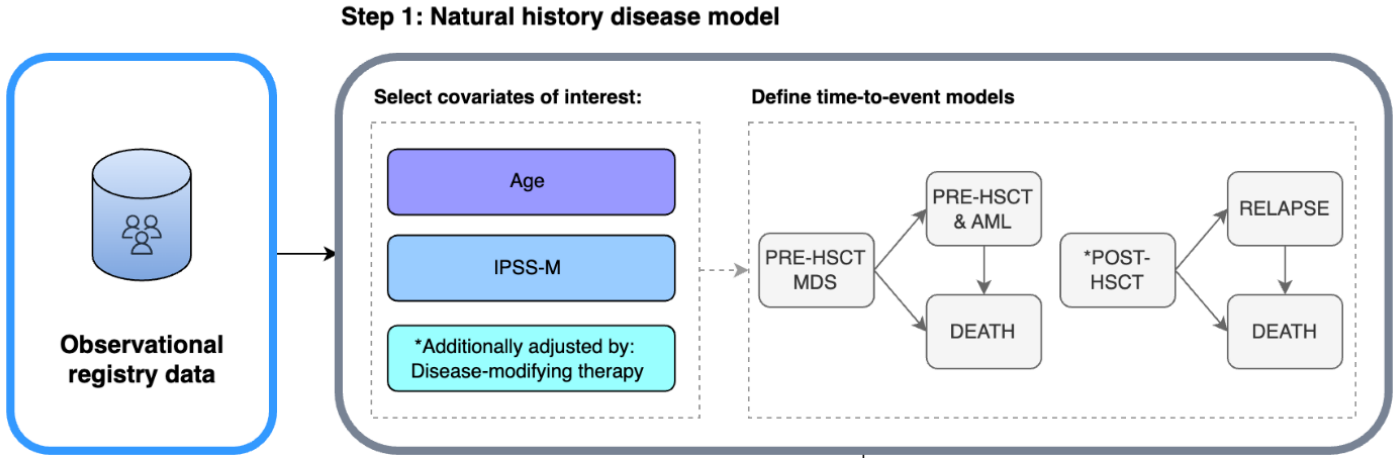
- To compare the outcome of transplantation policies based on IPSS-M vs original IPSS-R and to measure the proportion of patients in which the optimal timing for HSCT would change by introducing molecular information in the decision process

Sauta E at al. Journal of Clin Oncol. 2023;41:2827-2842
Gregorio C et al. Journal of Clin Oncol CCI, 2024, May;8:e2300205.
Tentori C et al. Journal of Clin Oncol. 2024, May 9:JCO2302175.

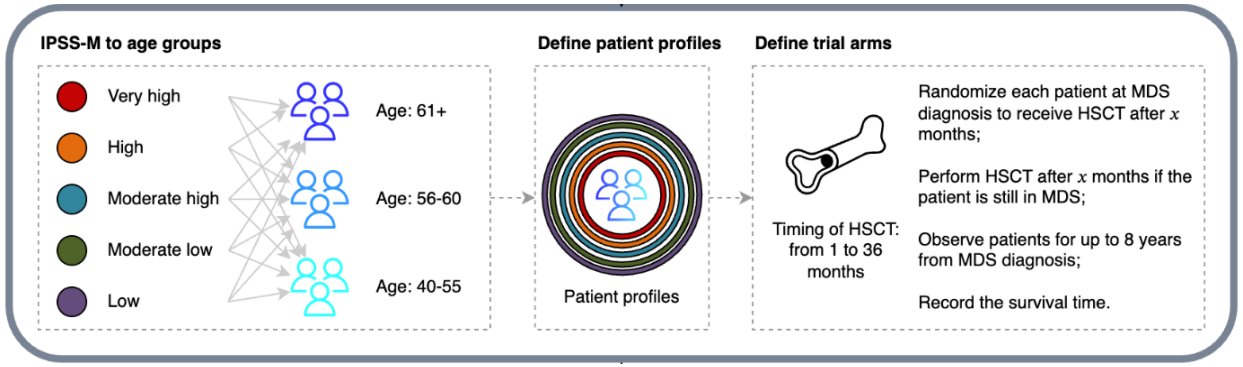
HUMANITAS RESEARCH HOSPITAL AI CENTER

Development of the Decision Support System

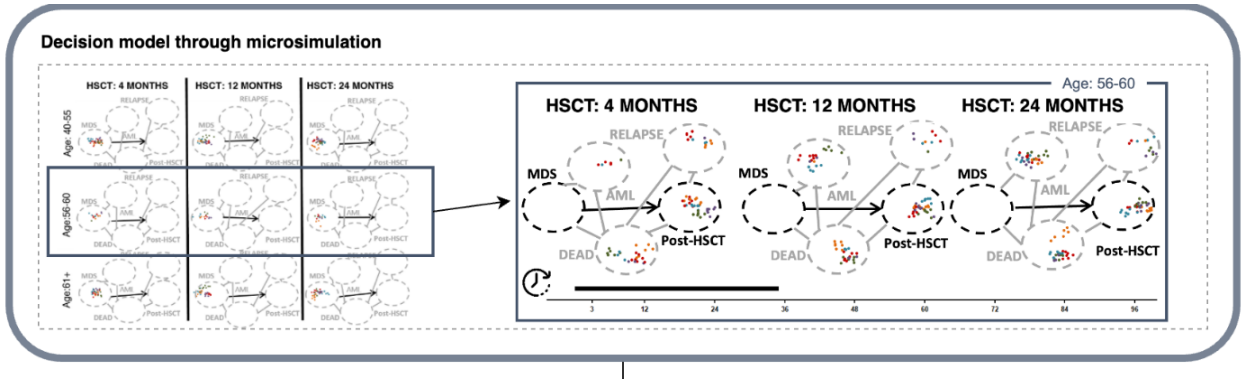
STEP 1 – Model of the disease natural history



STEP 2 Simulation of the target trial



STEP 3 Scenario analysis - microsimulation

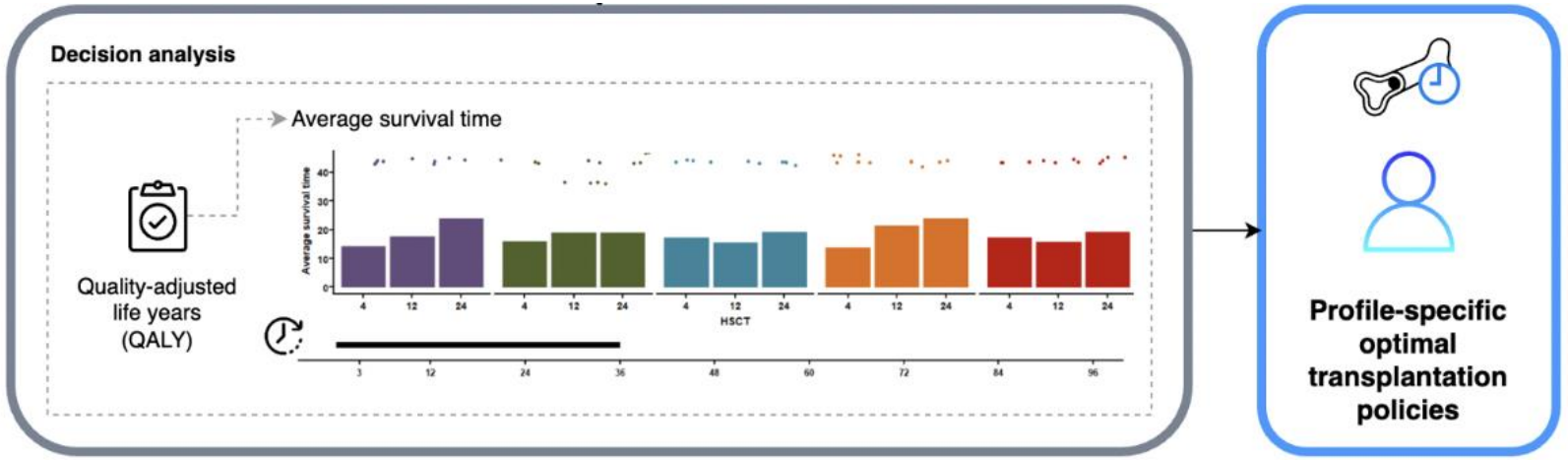


Gregorio C et al. Journal of Clin Oncol CCI, 2024, May;8:e2300205.
Tentori C et al. Journal of Clin Oncol. 2024, May 9:JCO2302175.

HUMANITAS RESEARCH HOSPITAL AI CENTER

Development of the Decision Support System

STEP 4 Optimal timing of HSCT based on the patient's profile



- Results were used to estimate the average survival time over a time horizon of 8 years for each combination of covariates, known as Restricted Mean Survival Time (RMST).
- RMST represents the expected time a patient spends in the model before reaching the "death" state.
- The optimal transplantation policy conditioned on the covariates of interest was defined as the 95% confidence interval for the timing that maximized the average survival time
- Average survival time was also estimated accounting for quality of life (QoL), using quality-adjusted life years (QALY).

Gregorio C et al. Journal of Clin Oncol CCI, 2024, May;8:e2300205.
Tentori C et al. Journal of Clin Oncol. 2024, May 9:JCO2302175.

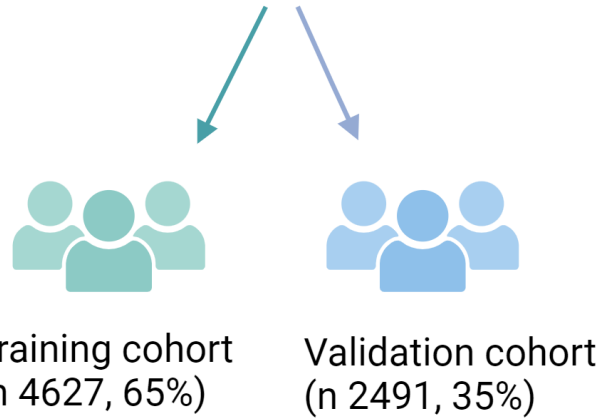
HUMANITAS RESEARCH HOSPITAL AI CENTER

Study Population

GENOMED4ALL SYNTHEMA

ebn ERN | EuroBloodNet MDS

7118 patients from 26 institutions



Inclusion criteria:

- age ≥18 years
- a diagnosis of primary MDS according to WHO 2016 criteria
- available information on IPSS-M related variables

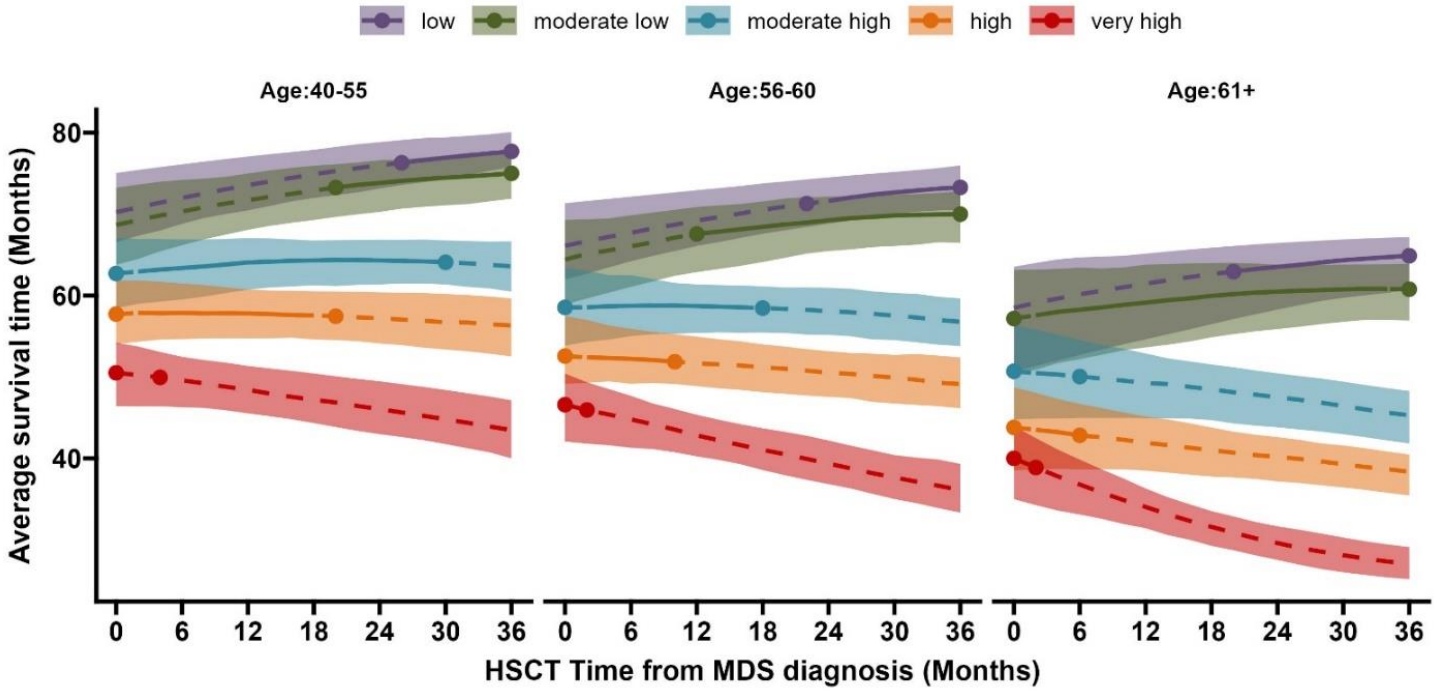
Exclusion criteria:

- patients affected with therapy-related MDS, acute myeloid leukemia (AML) from MDS
- incomplete information on IPSS-M variables

Gregorio C et al. Journal of Clin Oncol CCI, 2024, May;8:e2300205.
Tentori C et al. Journal of Clin Oncol. 2024, May 9:JCO2302175.

HUMANITAS RESEARCH HOSPITAL AI CENTER

IPSS-M based transplantation policy

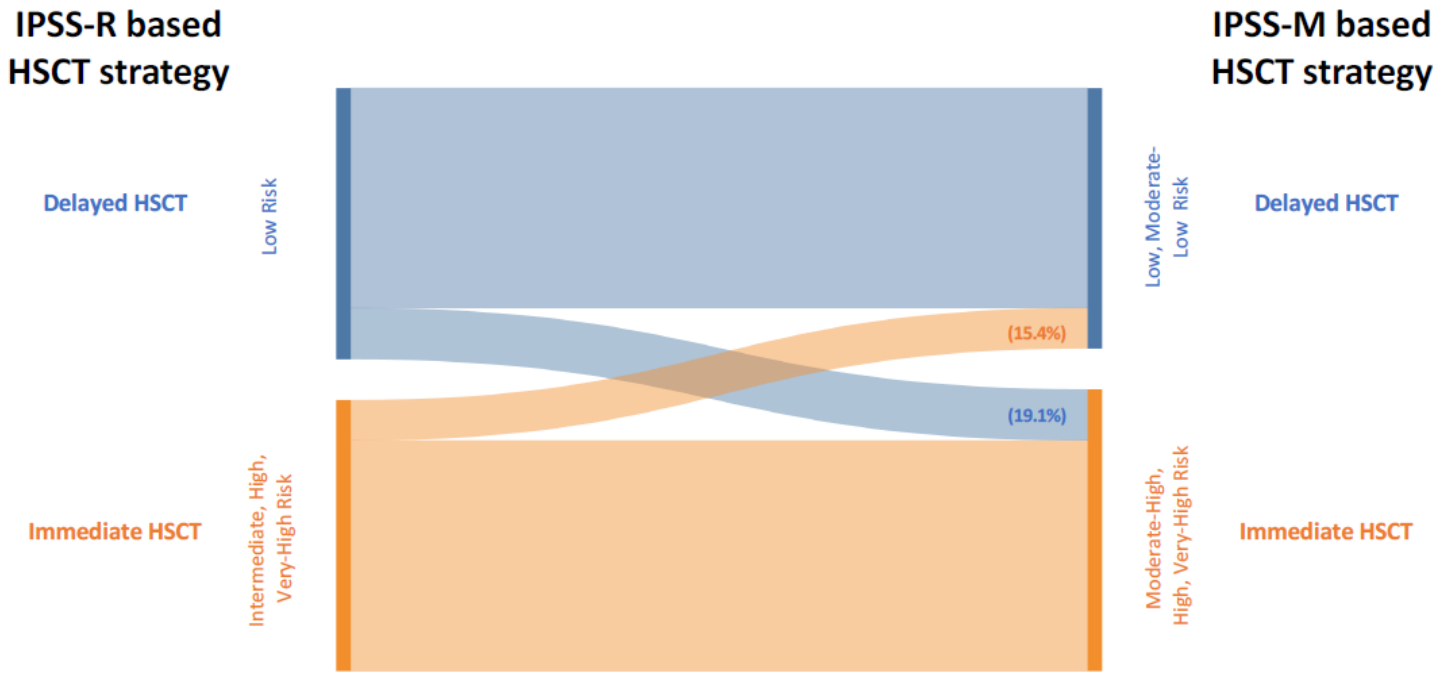


- Under an IPSS-M based policy, in the patients with either low- and moderate-low risk benefited from a delayed transplantation policy, while in those belonging to moderate-high, high- and very-high risk categories immediate transplantation was associated with a prolonged RMST

Gregorio C et al. Journal of Clin Oncol CCI, 2024, May;8:e2300205.
Tentori C et al. Journal of Clin Oncol. 2024, May 9:JCO2302175.

HUMANITAS RESEARCH HOSPITAL AI CENTER

Comparison of IPSS-R vs IPSS-M transplantation policy

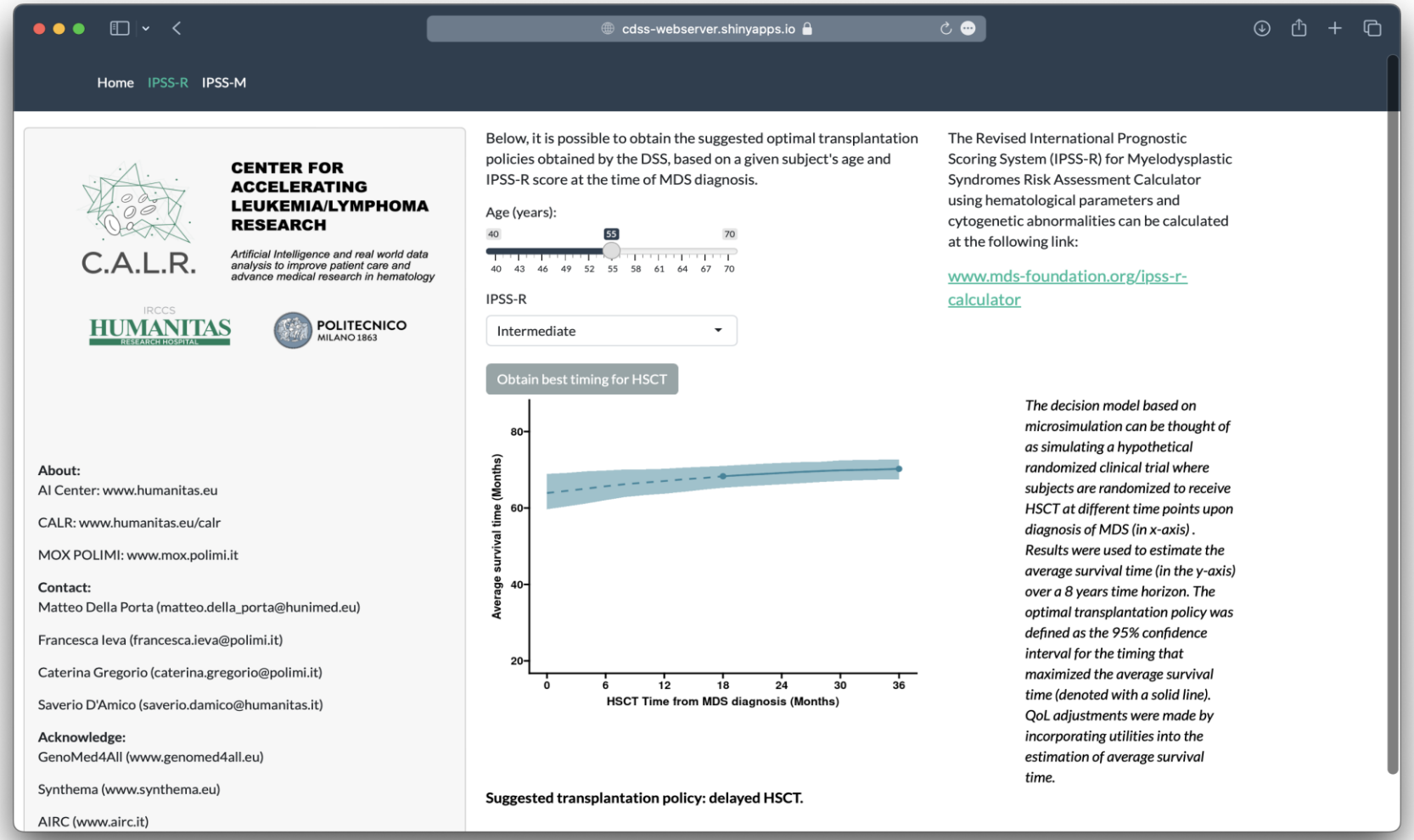


- Modelling decision analysis on IPSS-M vs. original IPSS-R in this population changed transplantation policy in a significant proportion of patients (17%)
- The comparison of life expectancy for the optimal transplantation policies obtained using different scoring systems (IPSS-R/IPSS-M) resulted in a significant gain of RMST under an IPSS-M based policy across all age groups (P=0.001)

Gregorio C et al. Journal of Clin Oncol CCI, 2024, May;8:e2300205.
Tentori C et al. Journal of Clin Oncol. 2024, May 9:JCO2302175.

HUMANITAS RESEARCH HOSPITAL AI CENTER

Clinical Decision Support System for Transplantation in MDS - WEB TOOL



Gregorio C et al. Journal of Clin Oncol CCI, 2024, May;8:e2300205.
Tentori C et al. Journal of Clin Oncol. 2024, May 9:JCO2302175.

HUMANITAS RESEARCH HOSPITAL AI CENTER

Summary

- Synthetic Data are artificial data generated by an algorithm (generative AI) trained to learn all the essential characteristics of a real dataset.
- Properties of Synthetic Data include
  - ✓ Data sharing (GDPR)
  - ✓ Classes balance and resolution of missing information
  - ✓ Data augmentation
  - ✓ Generation of new evidence
- Synthetic Data can accelerate clinical research in hematology (Clinical Trials)
- Synthetic Data can contribute to generate clinical evidence in specific scenarios in which there is a lack of randomized trials (Clinical Decision Support Systems)
- High quality of data input and clinical validation of generated data are required to assess the quality of Synthetic Data

HUMANITAS RESEARCH HOSPITAL AI CENTER

GEN MED 4ALL

Horizon 2020
European Union Funding
for Research & Innovation

SYNTHEMA

AIRC
Metastatic disease: the key unmet need in oncology
Mynerva

IRCCS
HUMANITAS
RESEARCH HOSPITAL

C.A.L.R.
CENTER FOR ACCELERATING LEUKEMIA/LYMPHOMA RESEARCH

M Robin
N Gagelmann
C Gurnari
S Ball
JC Caballero Berrocal
M Meggendorfer
LP Zhao
JP Bewersdorf
D Sallman
F Sole
G Garcia-Manero
U Germing

S Kordasti
V Santini
G Sanz
U Platzbecker
M Diez-Campelo
JP Maciejewski
L Ades
P Fenaux
T Haferlach
AM Zeidan
R Komrokji
F Ieva
G Castellani

S D'Amico
L Lanino
G Maggioni
A Campagna
C Tentori
E Sauta
M Ubezio
G Todisco
G Asti
M Delleani
E Zazzetti
C Gregorio
A Russo
V Savevski