The role of AI tools to boost hematologic clinical research

M.G. Della Porta

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 MATTEO DELLA PORTA

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

Limitations and pitfalls of randomized clinical trials (RCT)

RCT may have major limitations is some clinical scenarios, including:

- Rare and ultrarare diseases (few patients available to be enrolled into a clinical trial)
- Patients with major unmet clinical needs in which best available therapy is not effective (ethical concerns to treat these patients with best available therapy)
- Patients with major unmet clinical needs in which most of the trials are failed (urgent need to accelerate the evaluation of the effectiveness of new treatment options)
- Diseases arising in the elderly people (limited opportunity to be enrolled into a conventional clinical trials)
- Diseases in which the landscape of standard treatment is rapidly changing and therefore is not easy to define an appropriate control arm (challenge in defining the study design)

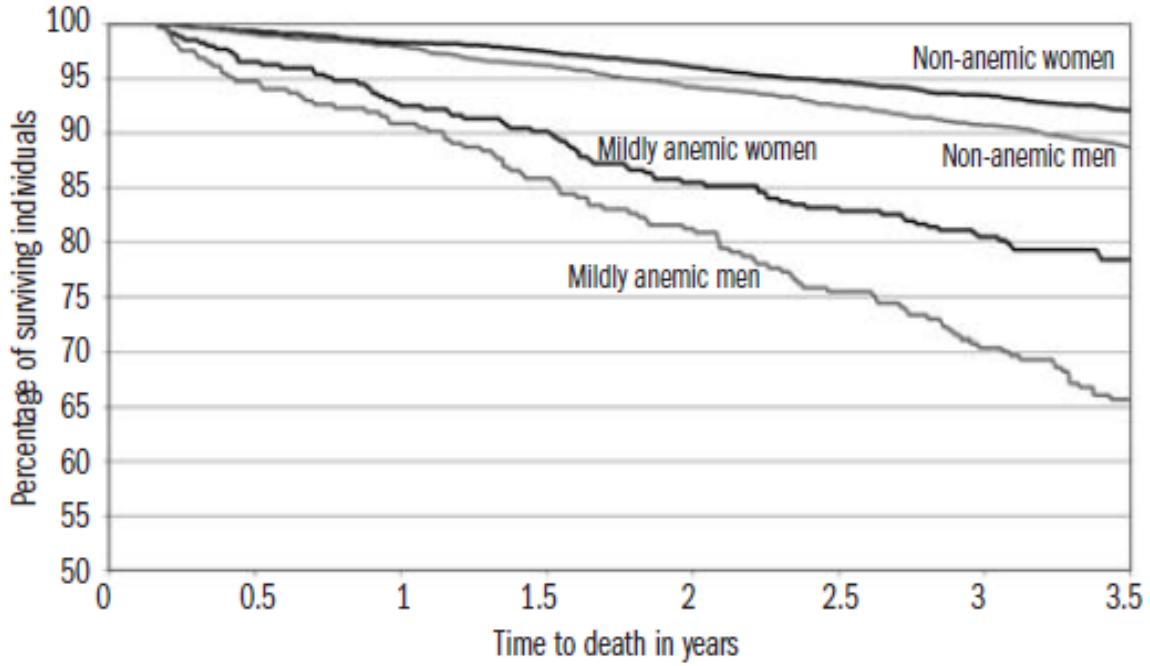
Serrano C et al. Nat Med. 2023 Nov;29(11):2689-2692
Hutson M. Nature 2024 Mar;627(8003):2-5.

Anemia in the elderly: the “Health & Anemia” study

Table 2. Prevalence of mild, moderate, and severe anemia in the elderly by age class.

Hemoglobin		Overall anemia Women: less than 12.0 g/dL Men: less than 13.0 g/dL		Mild anemia Women: 10.0 – 11.9 g/dL Men: 10.0 – 12.9 g/dL	
Age (years)[a]	*n*	Prevalence (%)	95% CI	Prevalence (%)	95% CI
65-69	2,040	6.3	5.3-7.5	5.4	4.5-6.6
70-74	2,165	8.9	7.7-10.3	7.8	6.7-9.1
75-79	1,933	13.4	11.8-15.2	11.3	9.9-12.9
80-84	1,348	18.9	16.7-21.4	15.7	13.6-17.9
85-89	779	26.6	23.1-30.4	22.1	18.9-25.6
90+	479	41.8	36.2-48.0	31.5	26.7-37.0
≥ 65	8,744	14.2	13.4-15.0	11.8	11.1-12.6
≥ 65	11,608	13.2[b]		11.1[b]	

CI: confidence interval; [a]age at blood test. [b]estimates (see text)



Only 7% of patients with a diagnosis of MDS are potentially eligible to be enrolled in a clinical trial

Tettamanti et al. Haematologica 2010;95:1849-1856; Rossi M et al. Blood 2021;138:2093-2105

The opportunity of EMA Qualification Registry initiative

The European Medicines Agency (EMA) established the Patient Registry Initiative to explore ways of supporting the use of patient registries in generating high-quality data for regulatory decision making and to enable a systematic approach to their use.

Use cases including external control data in marketing applications (hematology)

Product and Indication	Pivotal Data	Context for use of patient registries
Axicabtagene ciloleucel	Open label, single arm study (ZUMA 1 Phase II) with a primary endpoint of objective response rate (CR)	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	Open-label single arm study (C2201) with a primary endpoint of overall response rate	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	Single arm, open-label Phase I/II study with a primary endpoint of immune reconstitution and an ongoing open label RCT	The **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.

A RWD control arm for refractory metastatic colorectal cancer: the **"no placebo initiative"** *

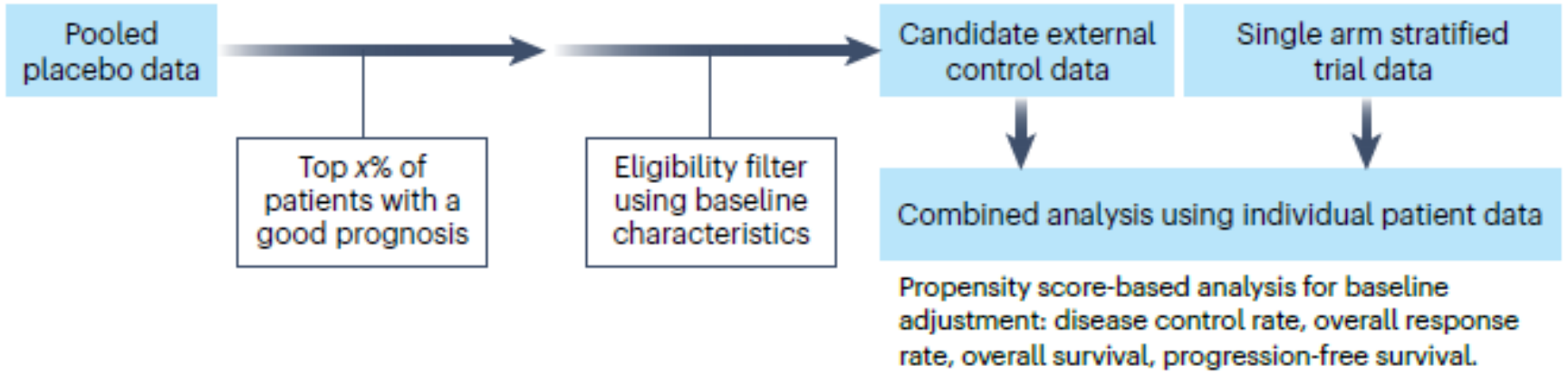


Fig. 1 | Three-step analysis for no-placebo initiative. First, participants enrolled in trials with placebo arms will be selected based on compatible patient demographics and key characteristics. These data will form the synthetic control arm. Second, patients in the top percentile for overall survival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

** In partnership with FDA, EMA and Pharmaceuticals and Medical Devices Agency of Japan*

Yoshino T, et al. Nat Med. 2023 Oct;29(10):2389-2390

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.

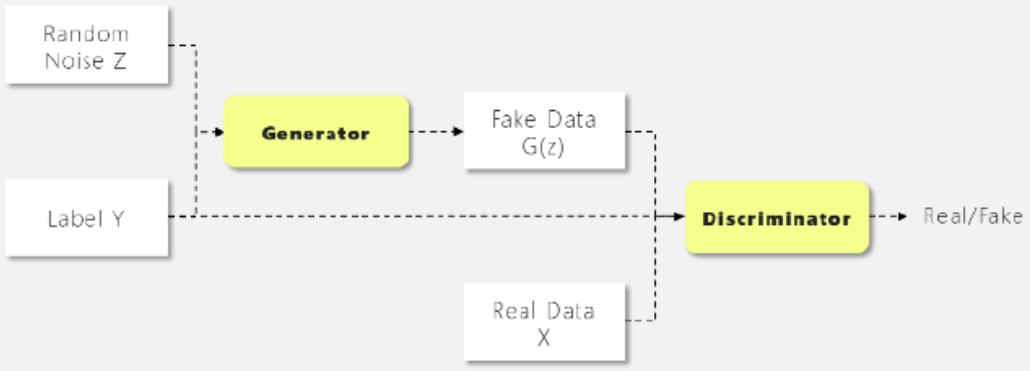


Figure 2 Conditional Generative Adversarial Networks architecture

Properties and possible applications

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

Chen R et al. Nature Biomedical Engineering 2021;5:493–497
Savage N et al. Nature 2023 doi: 10.1038/d41586-023-01445-8

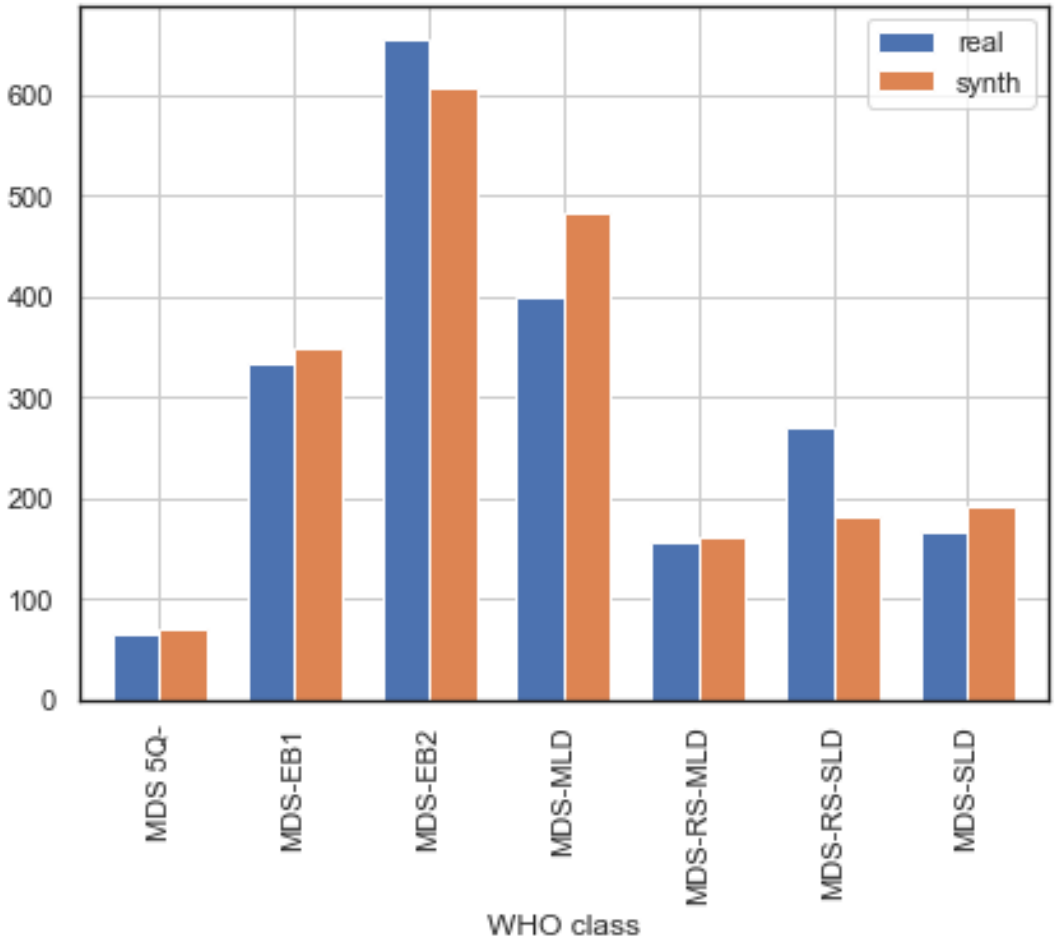
Synthetic vs. Anonymized data

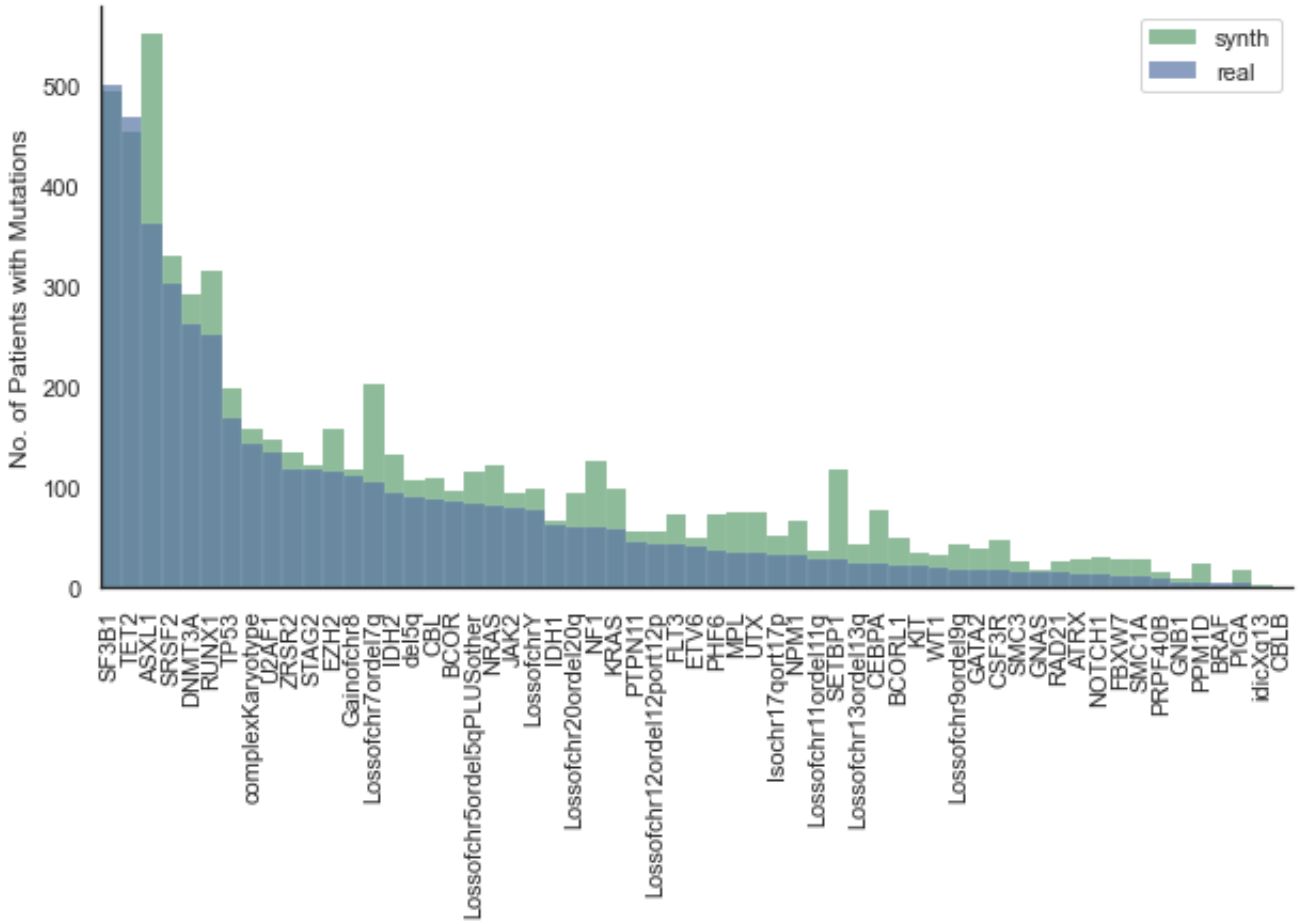
Advantages of Synthetic data:

- GDPR compliant by definition
- No loss of infomation associated to generation process
- No missing data
- Resolution of classes inbalance
- Data augmentation possible

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

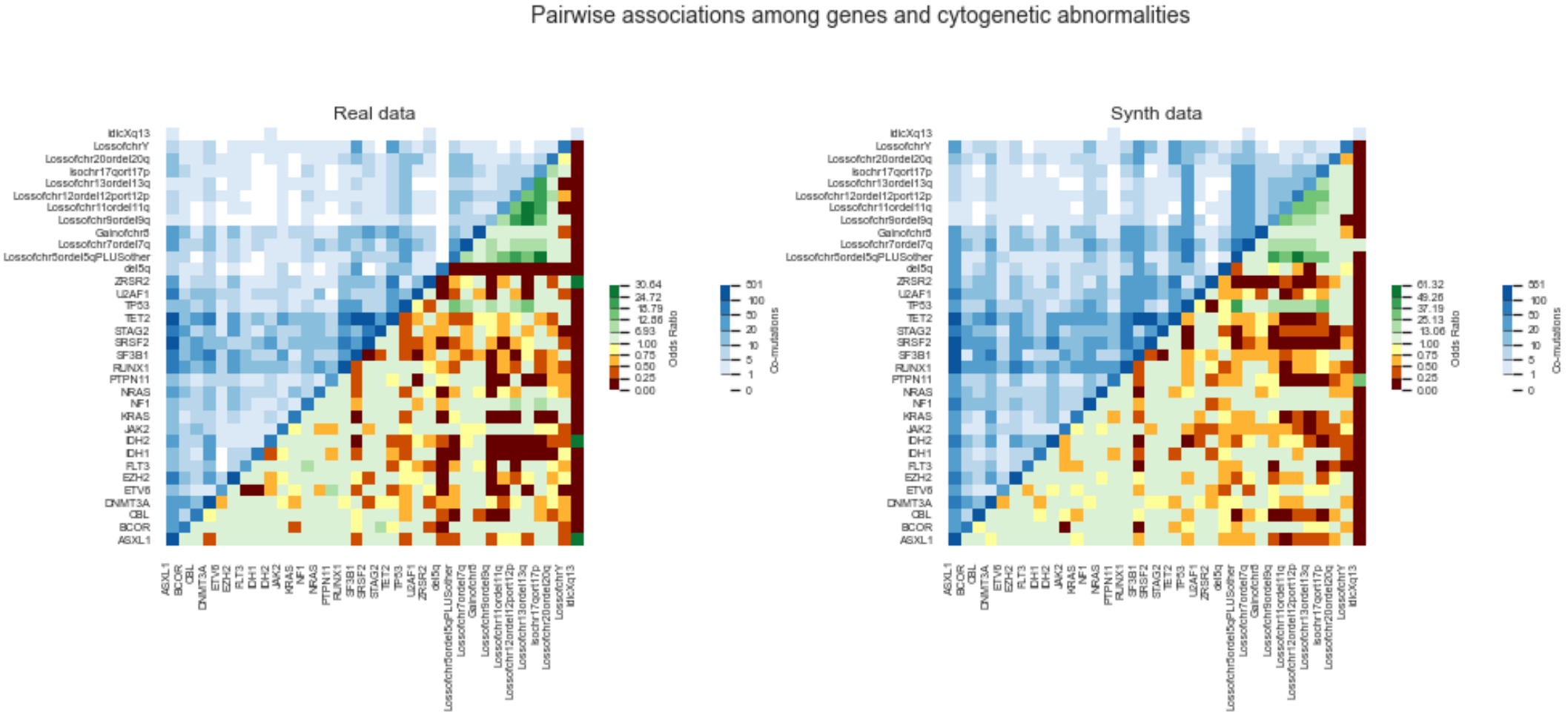
Synthetic vs. Real Patients: comparison of clinical and molecular features in MDS





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

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



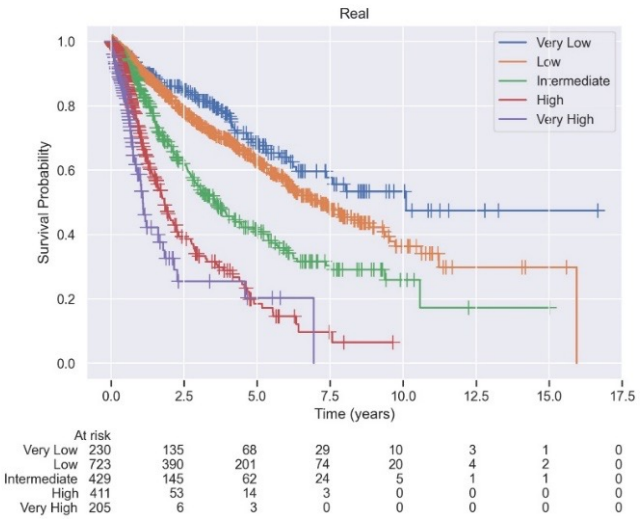
Jacobs F et al. Journal of Clin Oncol CCI 2023 Aug;7:e2300045
D'Amico S et al. Journal of Clin Oncol CCI 2023 Jun;7:e2300021
D'Amico S, et al. Journal of Clin Oncol CCI 2024, in press

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

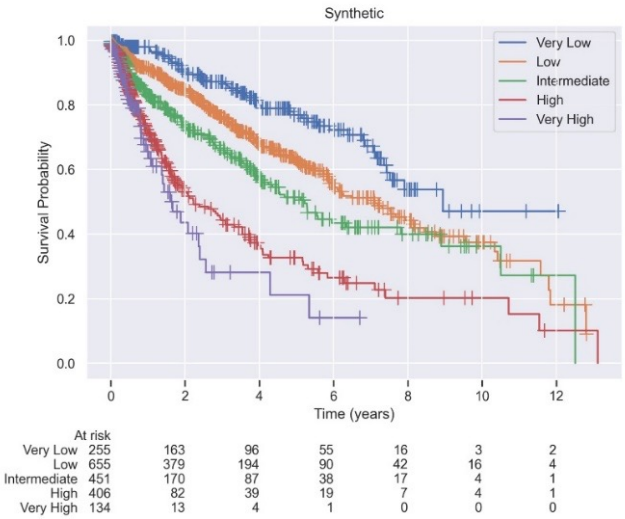
Probability of OS stratified by IPSS-R

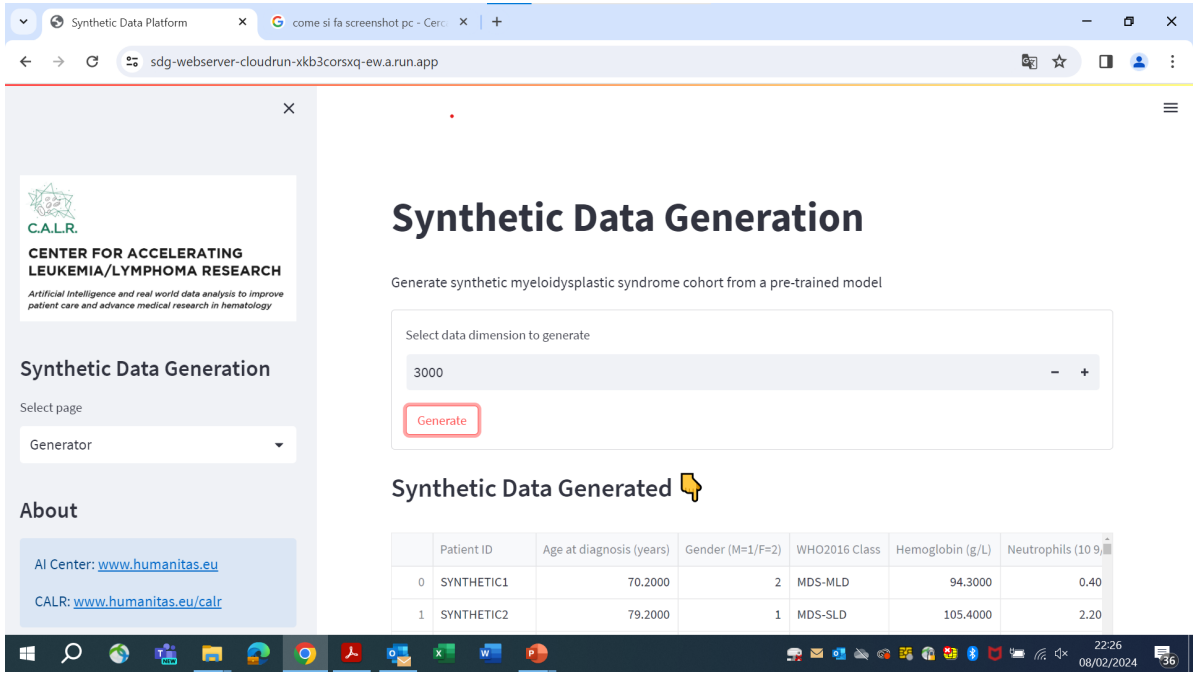
Data augmentation:
from 2043 to 10,000 patients

REAL PATIENTS



SYNTHETIC 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
Tentori CA et al Journal of Clin Oncol. 2024 May 9:JCO2302175

Performance of Synthetic Data

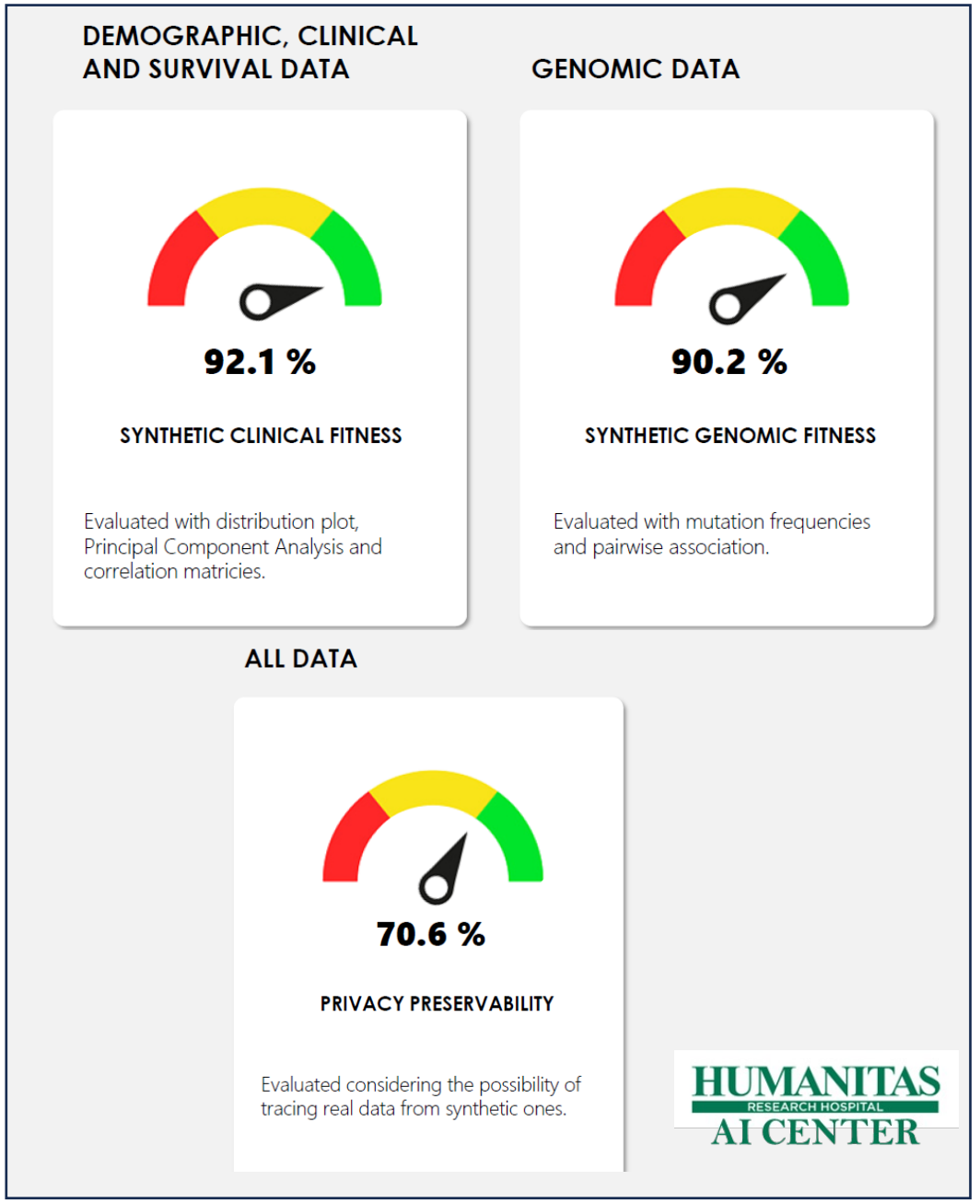
2021 WHO guidance on ethics and governance of AI for health

https://www.who.int/publications/i/item/9789240029200

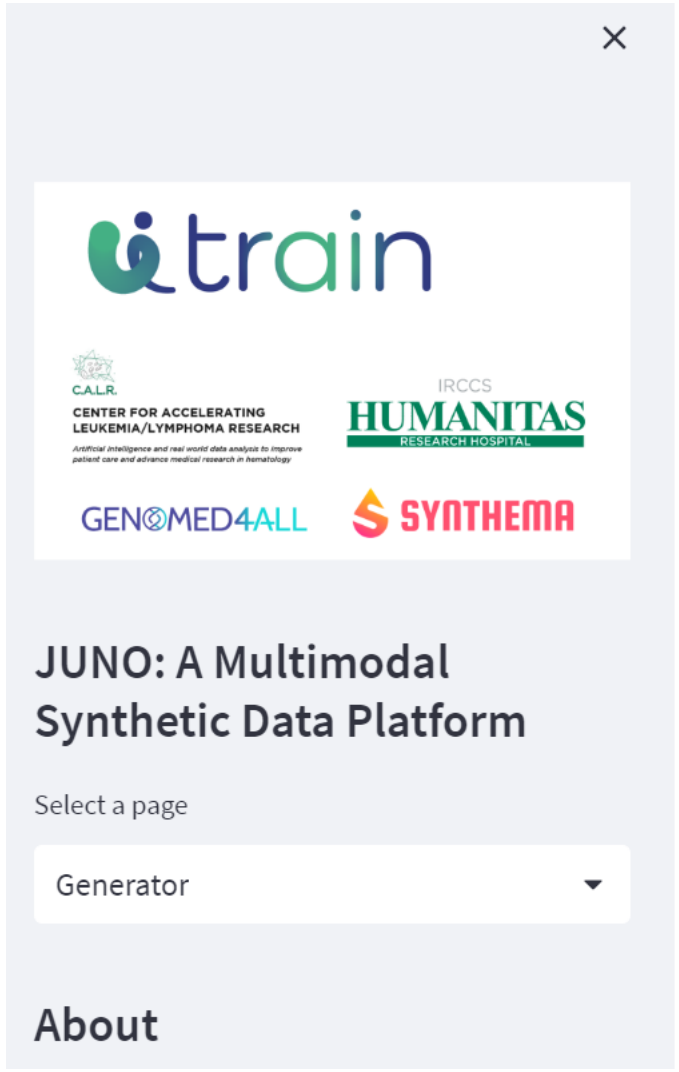
We have to address three important topics for a right deployment of AI in hematology:

- Transparency of models: interpretability and explainability
- Reliability of models: independent validation of generated AI-models
- Protection of data and data sharing: compliance with GDPR (EU)

D'Amico S et al. Journal of Clin Oncol CCI 2023 Jun;7:e2300021



Generation of multimodal longitudinal synthetic data



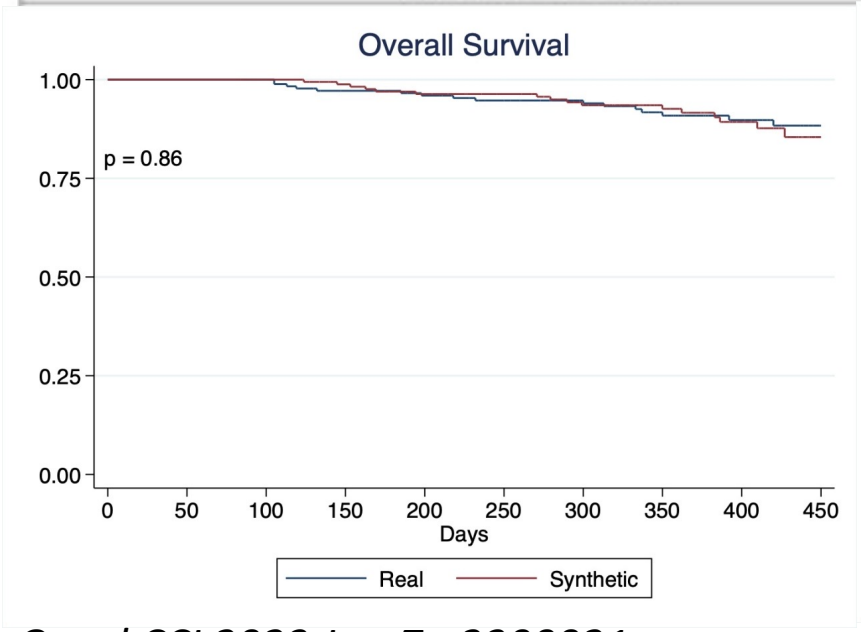
Synthetic Data to accelerate clinical research in Hematology

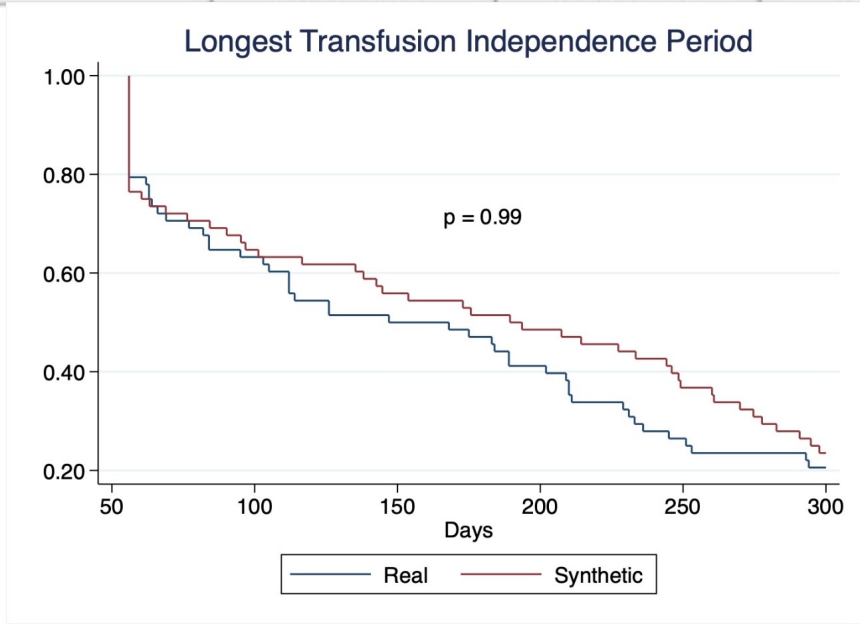
Comparing endpoints of clinical trials using **real** and **synthetic** control arms.

Real-world efficacy and safety of luspatercept in patients with transfusion-dependent anemia due LR-MDS-RS, who had an unsatisfactory response to or are ineligible for erythropoietin-based therapy: a multicenter study by 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





D'Amico S et al. Journal of Clin Oncol CCI 2023 Jun;7:e2300021

Comparing endpoints of clinical trials using real and synthetic control arms

Real case study

Longitudinal Synthetic Data Generation by Artificial Intelligence to Accelerate Clinical and Translational Research in Breast Cancer

Real-world data (RWD) are critical for breast cancer research but are limited by privacy concerns, missing information, and data fragmentation.

A dataset of 1052 HER2-positive and triple-negative breast cancer (TNBC) patients from the i2b2 platform was used. Advanced generative models.

The synthetic datasets exhibited high fidelity (score 0.94) and ensured privacy, with temporal patterns validated through time-series analyses and Uniform Manifold Approximation and Projection embeddings. Synthetic data replicated the endpoints of the APT trial, demonstrating its feasibility for generating synthetic control arms with preserved statistical properties of the real dataset.

AI-generated longitudinal SD effectively address key challenges in RWD use in breast cancer. This approach can improve translational research and clinical trial design while ensuring robust privacy protection. Integration with platforms like i2b2 highlight their scalability and potential for broader applications in oncology.

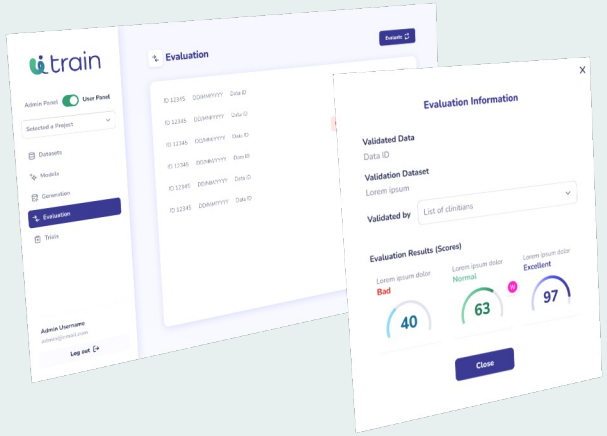
Gantt chart with events concerning procedures, treatments, clinical characteristics and last follow-up of a real patient (top) and a synthetic one (bottom). The timeline is based on visits. It is also indicated in parentheses the time (in days) elapsed since the diagnosis for the nth visit. The two patients, with the same clinical features identified at the first visits, follow the same course of treatments and procedures, at different times, but respecting the time deltas between the administration of ATC treatments

Zazzetti E et al. Journal of Clinical Oncol CCI 2025; in press

Train Synthetic Data Expertise

SynthEthics

Multimodal Clinical Synthetic Data Generation Platform



NTT DATA

Technological partnership for software infrastructure and deployment

Already **installed** in

Humanitas Hospitals

HUMANITAS

European Horizion Projects

synthia · SYNTHEMA · GEN@MED4ALL

Diseases registries

RISM · Azienda Ospedaliera Policlinico Bari

Universities and research center

Yale University · King's College London

Medical foundation and clinical trial institution

Fondazione CARIPLO · AIRC · fondazione GIMEMA · FIL Fondazione Italiana Linfomi

Industrial companies

Already **validated** in

- **Hematology** Rare blood cancers, Lymphomas
- **Oncology** Breast Cancer, Lung Cancer, Colorectal cancer
- **Gynecology** Ovarian cancer
- **Fertility**
- **Neurology** Multiple Sclerosis, Alzheimer Diseases
- **Gastroenterology** Ulcerative colitis, Crohn's disease
- **Hepatology** Hepatitis C
- **Rheumatology** Rheumatoid arthritis

Synthetic data and next-generation clinical trials in hematology – a European perspective

GEN MED 4ALL

SYNTHEMA

Realise RD

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

synthia

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Draft Concept Paper on the Development of a Reflection Paper on the Use of External Controls for Evidence Generation in Regulatory Decision-Making

- **Definition** of an external control arm
- The appropriate **clinical and regulatory setting and minimal requirements** for external controls
- Operational and feasibility aspects
- Planning, design, conduct, analysis and reporting of studies for which external controls are used and related **methodological aspects** such as **considerations on minimization of bias and confounding**, the definition of estimands, target trial emulation, type 1 error control, sample size
- **Prospectively planned external control comparisons vs comparisons conducted when results are already available (either trial data, external control or both)**
- **Data quality**: relevance, reliability, extensiveness, timeliness
- **Source(s)** of the external data

Synthetic data and next-generation clinical trials in hematology

Phase II study of anti-GD2 Chimeric Antigen Receptor-Expressing T cells (GD2-CART01) in pediatric patients affected by relapsed/progressing Neuroblastoma

Prof F Locatelli – Dr.ssa F Del Bufalo

SynthEthics

Next-Gen Clinical Trials in Hematology. A perspective

1. DATA SOURCES

- Registries and repositories on hematological diseases already certified or to be certified by “EMA qualification registry initiative” for regulatory purposes (EuroBloodNET, and others)
- Previously conducted clinical trials by academic networks or pharma

2. INNOVATIVE TECHNOLOGIES FOR DATA ACCESS, HARMONIZATION AND PRIVACY

- EU-funded consortia developing innovative methodologies for data access, data harmonization, common data models and privacy (Synthema, GenoMed4all, Synthia)

3. INNOVATIVE DESIGN FOR CLINICAL TRIAL

- EU-funded consortia developing innovative methodologies for the conduction and the evaluation of the quality of next-gen clinical trials including external control arms (Realise-D consortium and others)

4.1. BIULDING THE PLATFORM

and

4.2 SUBMISSION TO EMA CHMP

- Procedure on “Qualification of novel methodologies for medicine development”

Scientific coordination by hematological community

Early discussion with EMA **Monitoring/advisory by EMA** **Interaction with CHMP**

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, **CIO Humanitas**
- Giulia Maggioni, Luca Lanino, Alessia Campagna, Gabriele Todisco, **Clinicians**