SYNTHEMA + ERN-EuroBloodNet

Joint Training Programme on Synthetic Data Generation in SCD and AML

Funded by the European Union

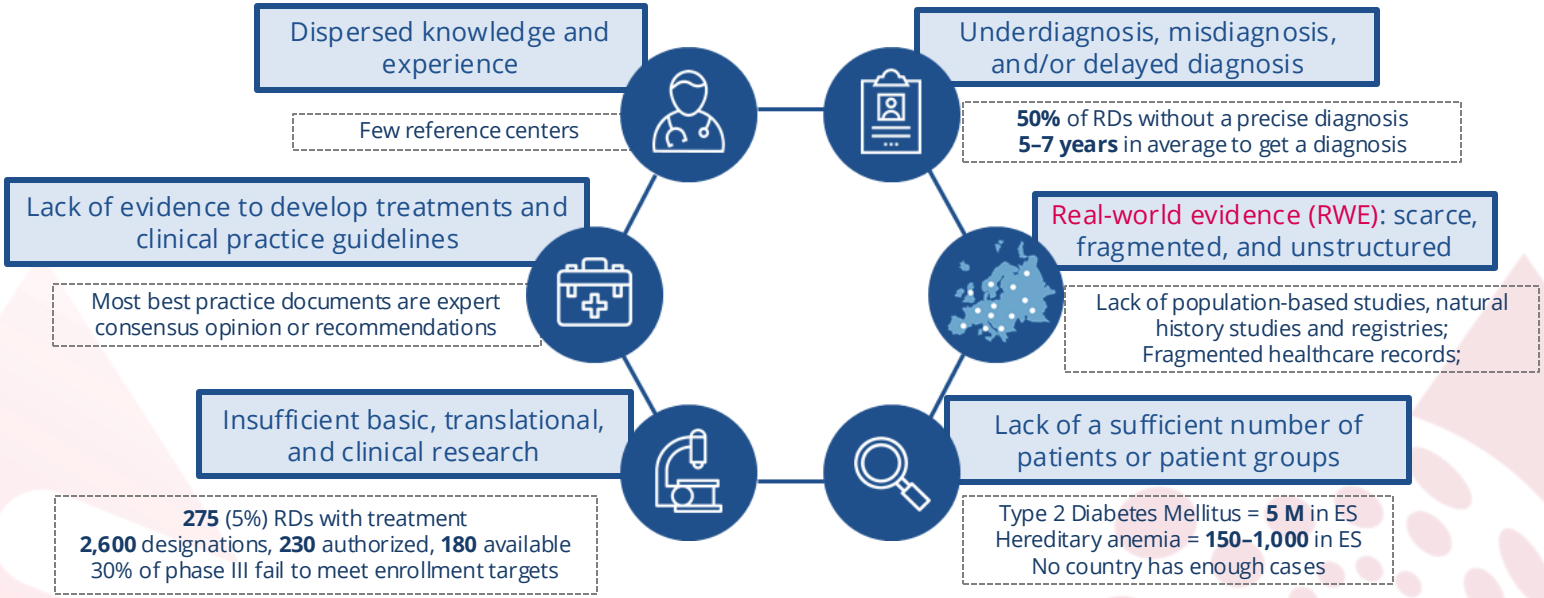
Opportunities of Synthetic Data Generation in Rare Haematological Diseases

María del Mar Mañú Pereira - ERN-EuroBloodNet Scientific Coordinator

#1

29th May 2026

ERN's perspective: the unmet need in RDs: +7000 RDs, 30Mi PLWRD in Europe



<15% of paediatric rare-disease patients can account for up to 50% of hospital care processes and costs

Synthetic data generation in Rare Hematological Diseases

- **Synthetic cohorts for better stratification**

Create synthetic longitudinal datasets combining EHR, genomic, and omics data to improve AML and SCD sub groups classification, identify rare patient subgroups, and enable AI-driven phenotype–genotype matching.

- **Synthetic data for biomarker and endpoint discovery**

Integrate synthetic clinical and omics data to identify prognostic biomarkers, predict disease progression, and develop better clinical endpoints for precision hematology.

- **Synthetic external control arms for faster clinical trials**

Use synthetic patient populations generated based on real-world data from registries as external control arms to optimize recruitment, simulate trials, and accelerate evaluation of new AML and SCD therapies.

- **Synthetic data augmentation to improve quality and reduce bias**

Apply generative AI to address missing or imbalanced data, improve dataset representativeness, and support patient-centered and regulatory-grade real-world evidence analyses.

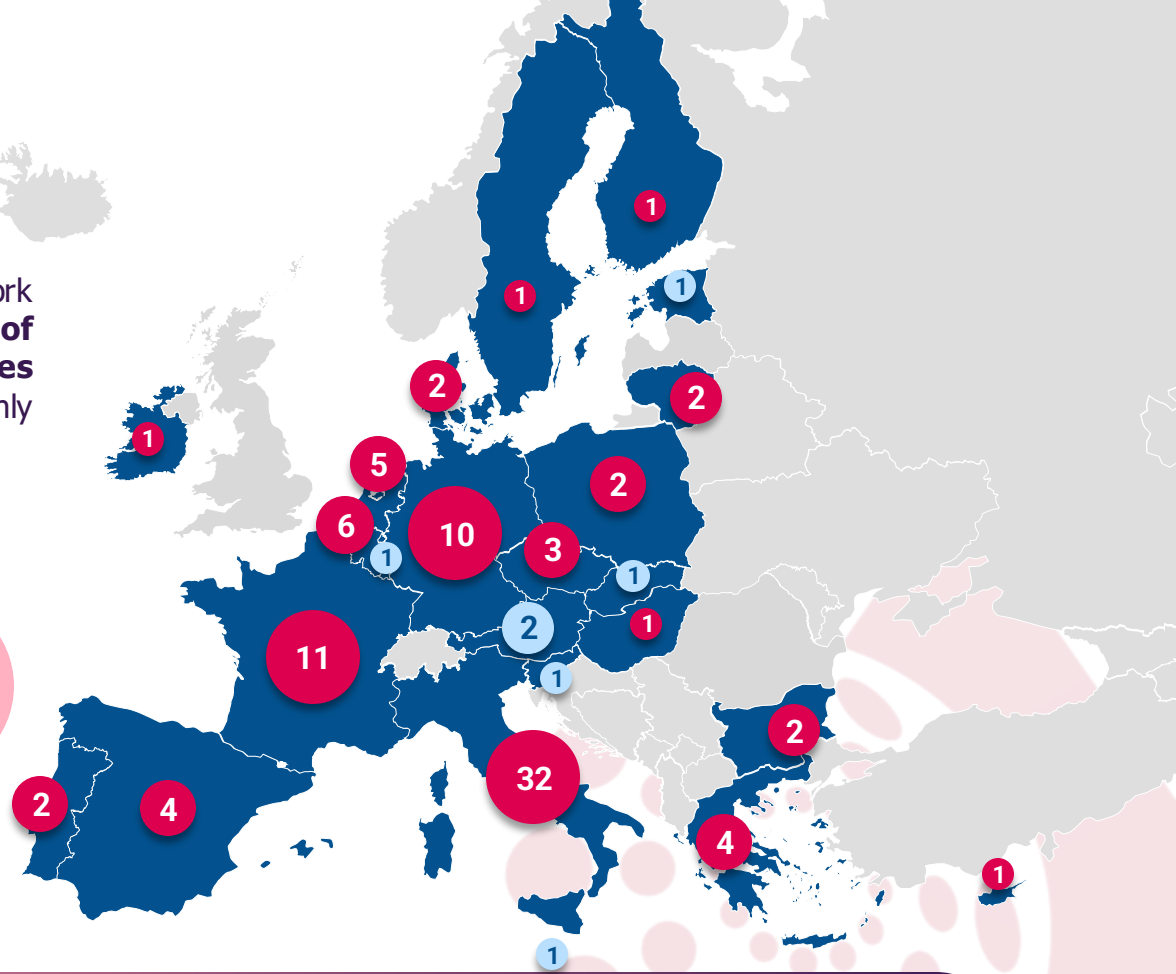
ERN-EuroBloodNet

ERN-EuroBloodNet is a collaborative network aiming to **improve the healthcare services of complex or rare hematological diseases** (RHD) and conditions that require highly specialized treatment in Europe.

89
Members in 18 MS

7
Affiliated Partners in 6 MS

10
ePAGs



Members

Affiliated Partners

ePAG: European Patient Advocacy Group

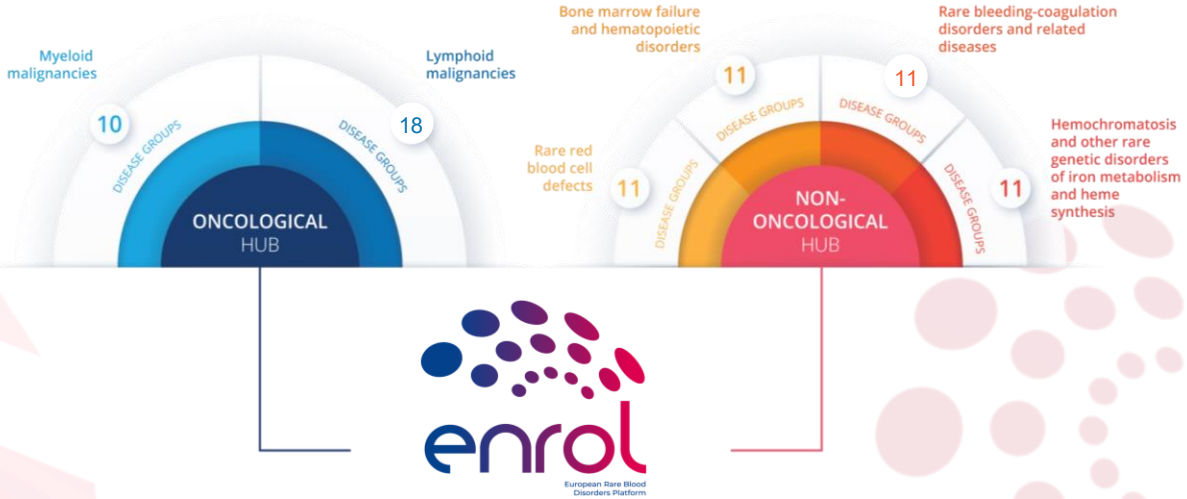
European Rare Blood Disorders Platform - ENROL

Facilitate epidemiological surveillance

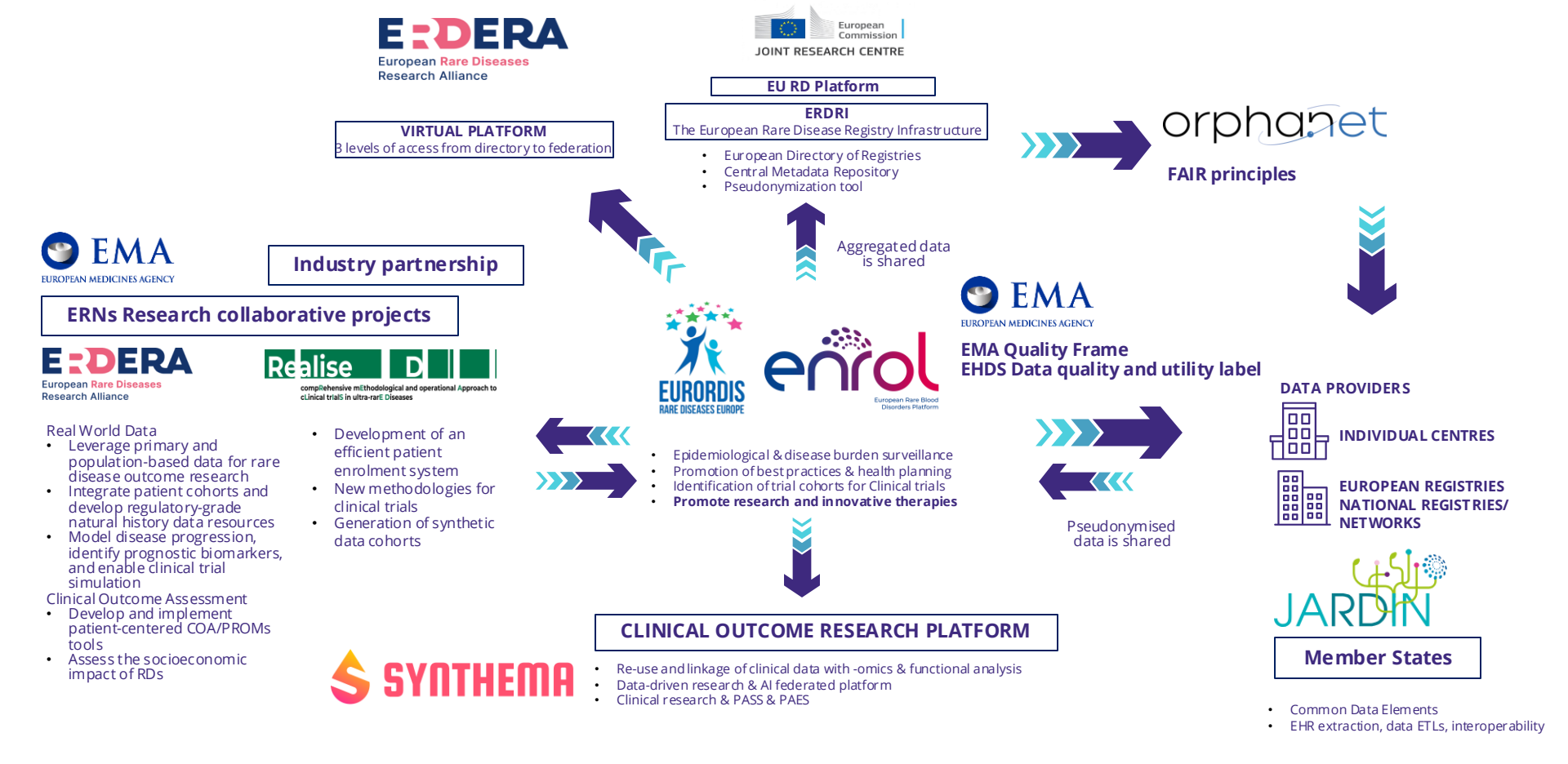
Enhance health planning

Enable the identification of patients' cohorts

Promote research & innovative therapies



ERN's perspective: opportunities to accelarate research through RWE



Synthema Use cases & objectives

Sickle cell disease (SCD)

The disease

SCD is an inherited chronic disorder caused by the presence of abnormal adult haemoglobin.

Red blood cells become rigid and sickle-shaped, breaking down or blocking normal blood circulation, resulting in acute and chronic pain and progressive organ-specific clinical complications.

Kidney disease

Cohort

1,000+ patients with genetic diagnosis for SCD disorder, including paediatric (1+ year-olds) and adult patients.

Acute myeloid leukaemia (AML)

The disease

AML is a type of blood cancer that starts in immature myeloid cells (blasts) in the bone marrow.

These abnormal cells grow and divide too quickly, interfering with normal blood cell production.

Overall survival analysis

Cohort

2,500+ patients with "de novo" AML (2016 WHO classification criteria).

Synthema Clinical partners

SCD

Vall d'Hebron
Institut de Recerca
VHIR

UMC Utrecht

Università degli Studi di Padova

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

AML

HUMANITAS
RESEARCH HOSPITAL

CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

HOSPITAL DA LUZ LEARNING HEALTH
TRAINING, RESEARCH & INNOVATION CENTER

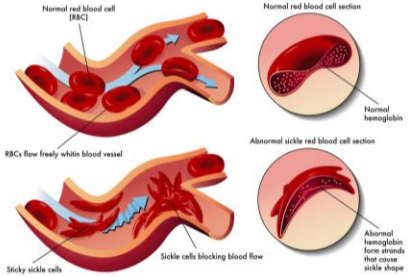
UMC Utrecht

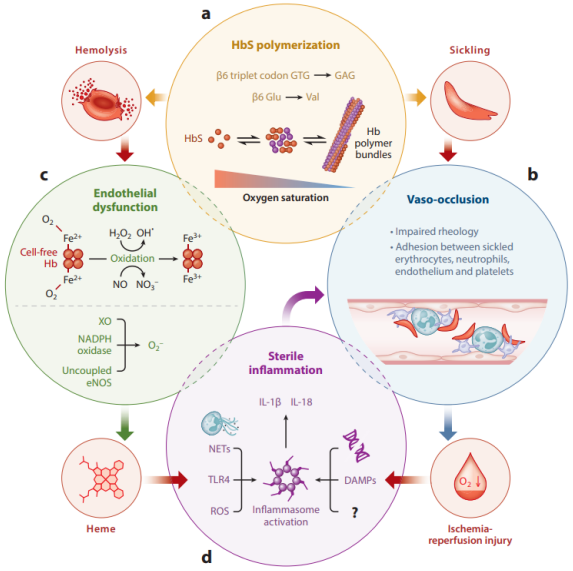
Università degli Studi di Padova

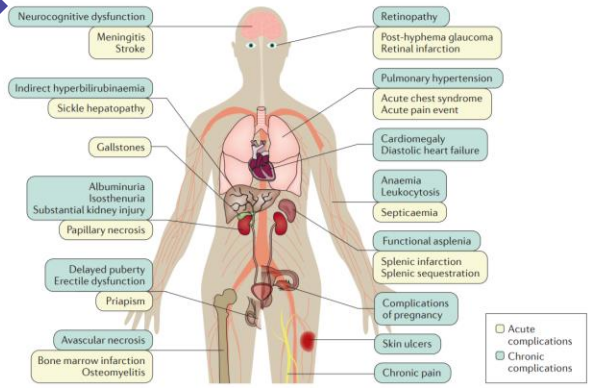
Sickle cell disease Use case

Sickle cell disease Use Case

Disease mechanism & Clinical outcomes







❑ Monogenic disorder autosomal recessive

- Different genotypes
- Genetic modifiers 1000+

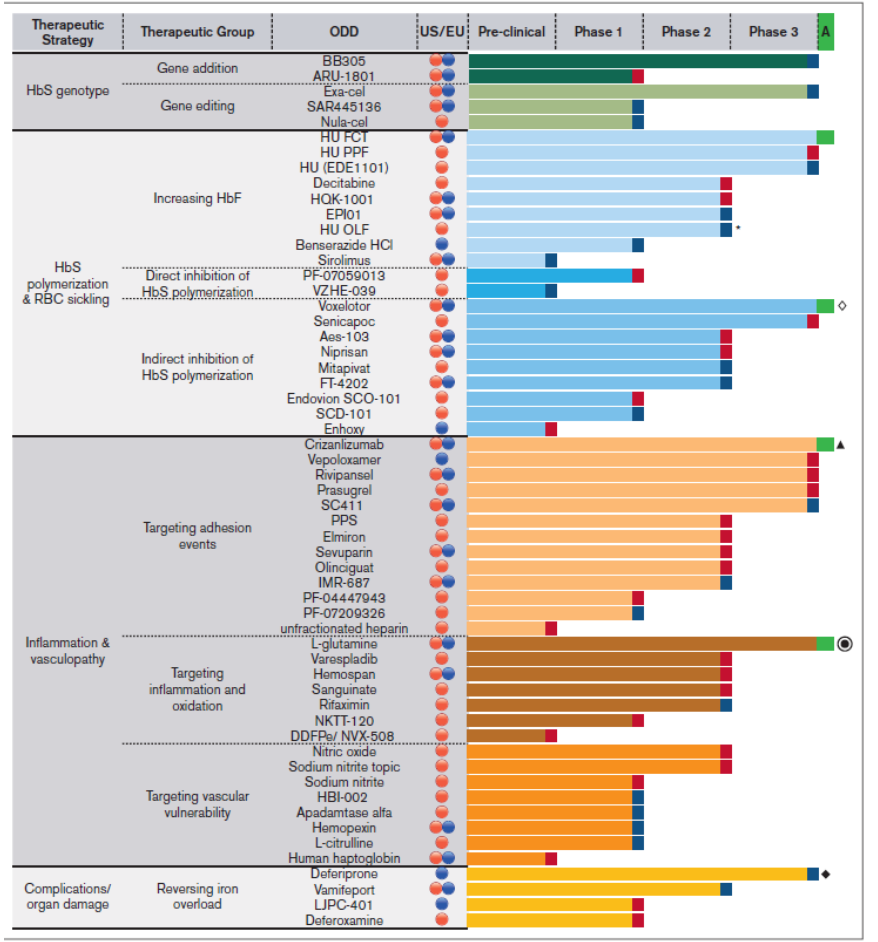
❑ Complex and interconnected physio pathological pathways

❑ High heterogeneity of clinical outcomes

- Chronic disease
- Multi systemic involvement
- Decreased life expectancy
- Low QoL

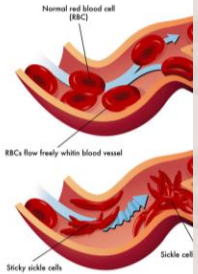
Sickle cell disease Use Case
Treatments

- Curative:
  - Hematopoietic stem cell transplanation
  - Gene Therapy
- Disease modifying:
  - Hydroxyurea: All SCD Patients genotypes SS/Sb0 from 1 yo
- Many new drugs in the pipeline
  - HbS polimerization & sickling
  - Inflamation & vasculopathy
  - Complications / organ damage



Sickle cell disease Use Case

Unmeet medical needs



Disease classification:

- Lack of reliable biomarkers for classification of patients groups according to clinical phenotypes and severity

Prognosis:

- To whom perform HSCT or gene therapy?
- Chronic Organ damage: how to prevent it or detect it very early?

Response to treatment:

- What is the optimal treatment/combination for each patient?

❑ Monoge nic disorder
autosomal re ssive
- Different genotypes
- Genetic modifiers

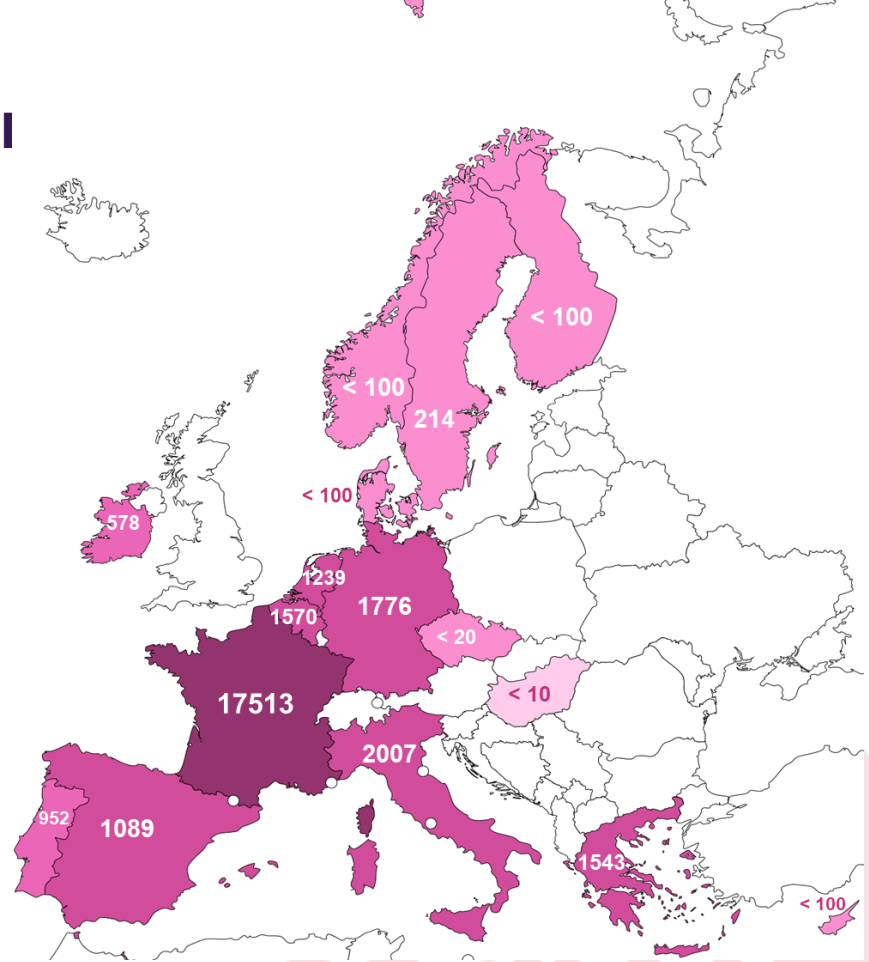
❑ Complex and interconnected physio-pathological pathways

❑ High heterogeneity of clinical outcomes
- Chronic disease
- Multi systemic involvement
- Decreased life expectancy
- Low QoL

Sickle cell disease Use case

Disease population year 2023 – SCD I

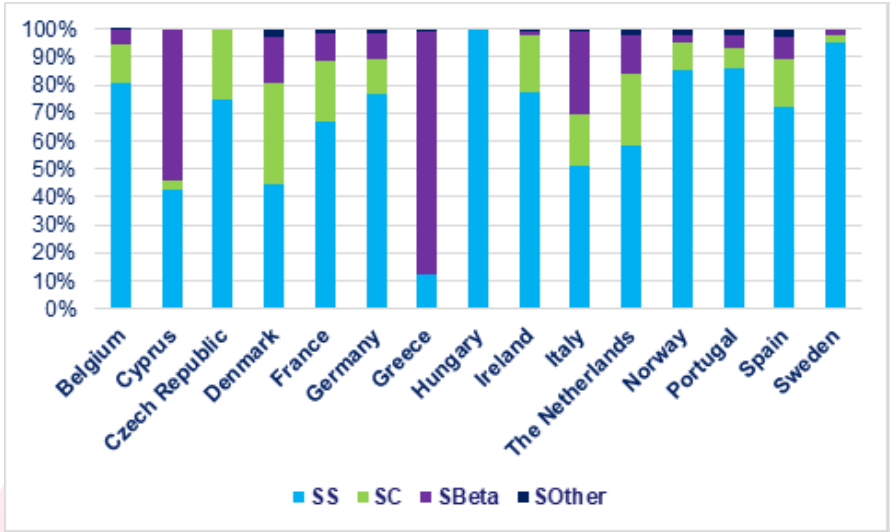
Country	Centres	SCD Total	% Pediatric	%Female
Belgium	14	1.570	53,4	57,0
Cyprus	3	71	23,9	49,2
Czech Republic	4	15	53,3	41,7
Denmark	2	98	36,7	50,0
Finland	1	86	52,3	-
France	30	17.513	40,5	53,1
Germany	35	1.776	82,1	47,9
Greece	23	1.543	9,5	58,4
Hungary	1	6	0,0	50,0
Ireland	1	578	58,5	54,6
Italy	36	2.007	48,2	51,2
Netherlands	6	1.239	43,7	56,7
Norway	5	84	48,8	52,4
Portugal	8	952	76,9	54,3
Spain	53	1.089	75,9	44,3
Sweden	11	214	72,7	48,6
TOTAL	**233**	**28.841**	**45,9**	**53,0**



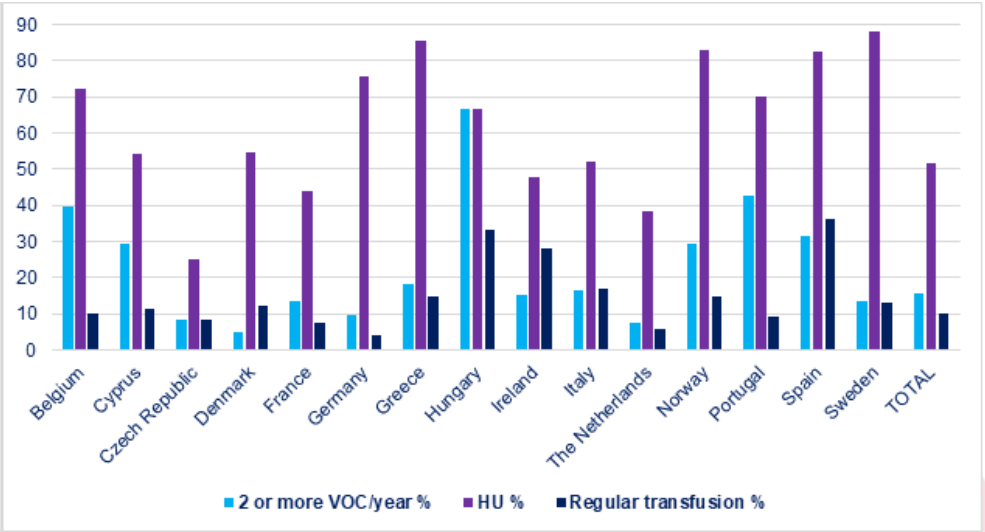
Data source: RADeep

Sickle cell disease Use case

Disease population year 2023 – Clinical outcomes heterogeneity

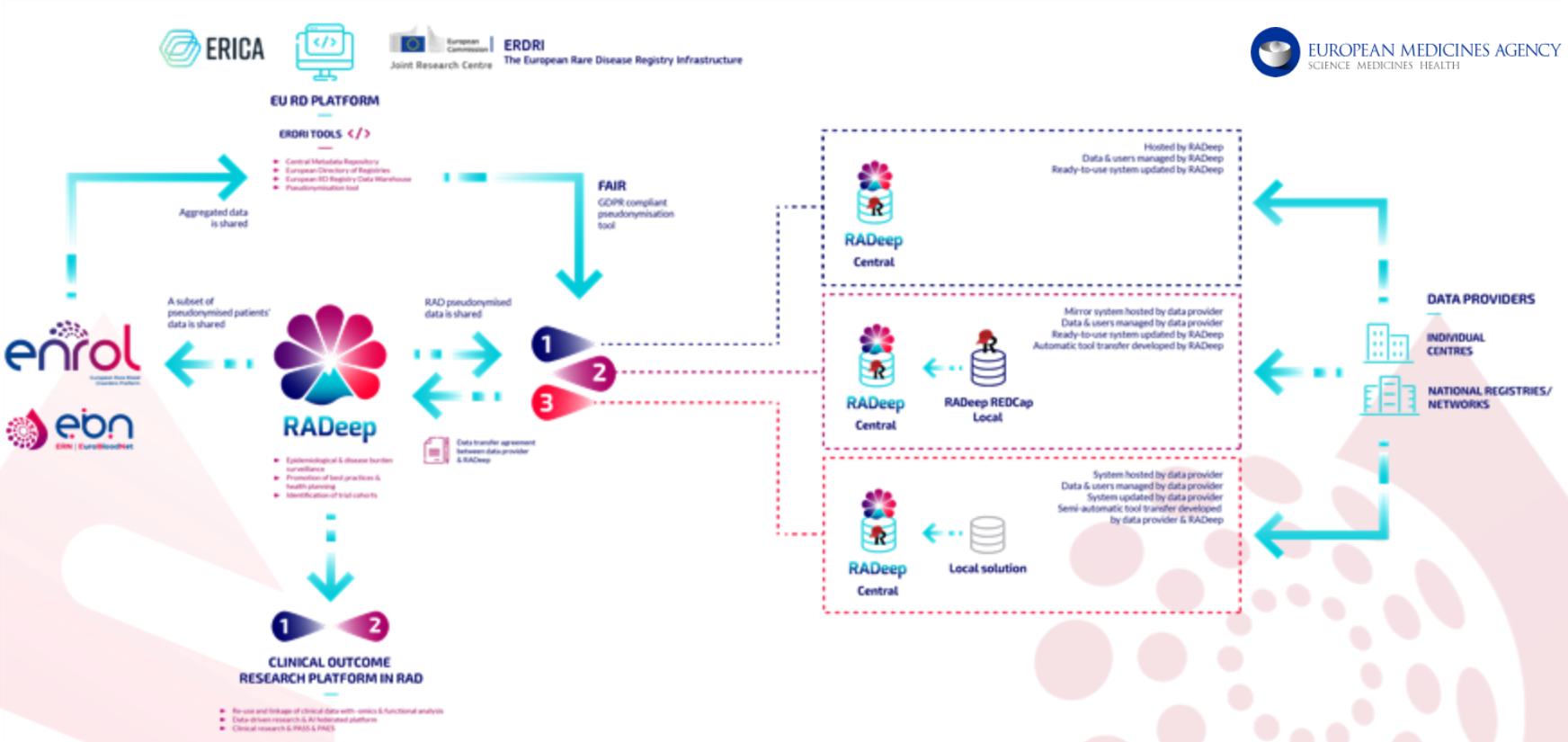


Country distribution of SCD patients by genotype



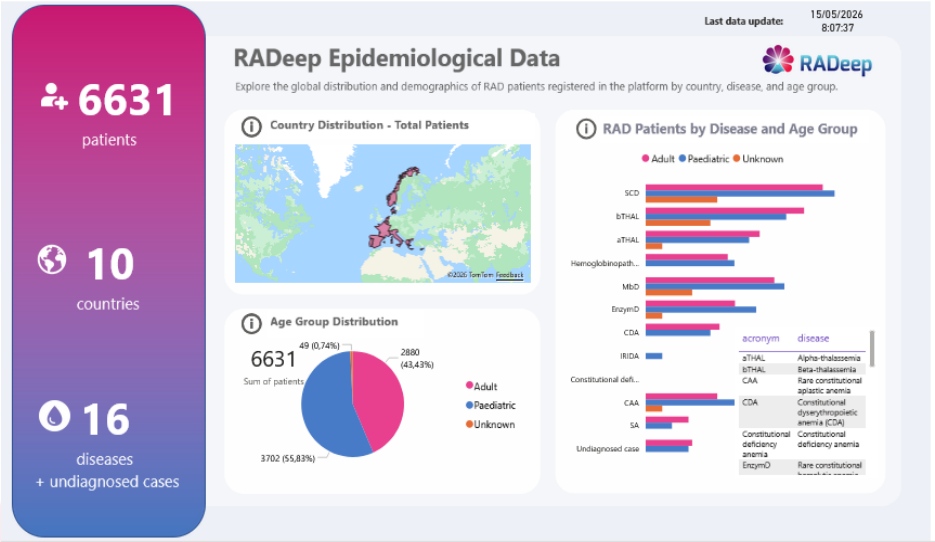
Country distribution of SCD patients by severity: % of patients with 2 or more VOC per year, on Hydroxyurea (HU) treatment, on regular transfusion treatment

SCD - Strategy for data collection: RADeep

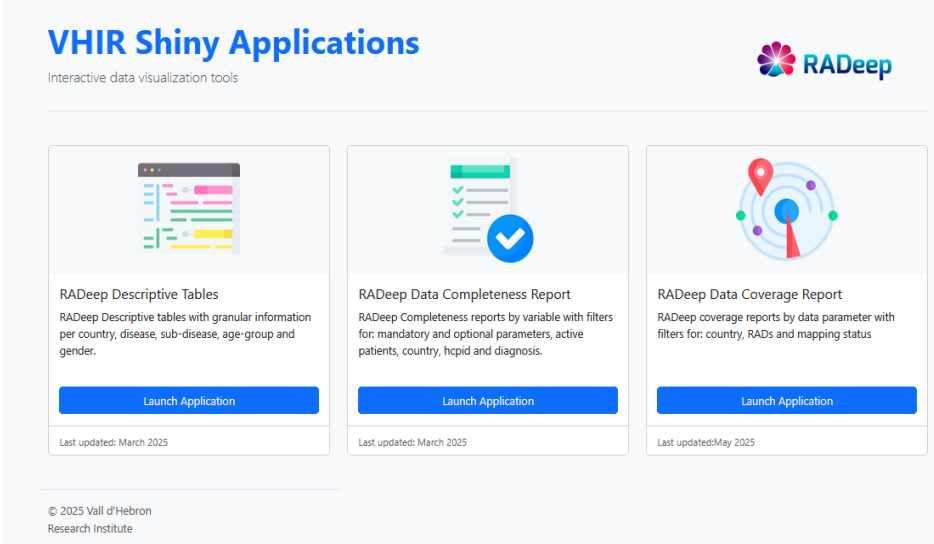


RADeep: EMA Data Quality Framework

Public dashboard



Private dashboards



RADeep - RADiANT

ERN- EuroBloodNet
Rare Anemia Disorders registry

GDPR / EHDS

Federated platform registry

GDPR / EHDS / Drug and MD Regulations / AI Act

Federated platform: 4 nodes in 4 countries

11 Ongoing collaboration agreements involving
180 (117 active) HCPs in 10 EU countries:
12 Member States

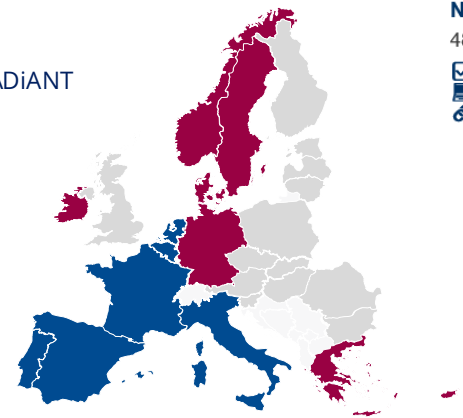
- Belgium
- Cyprus
- Denmark
- France (2)
- Greece
- Italy
- Portugal
- Spain
- The Netherlands
- + Norway

4 in negotiation
- *France*
- *Ireland*
- *Sweden*
- *Switzerland*

Countries onboard in RADiANT
4 EU countries

- France
- Italy
- Spain
- The Netherlands

2 in negotiation
- *Belgium*
- *Portugal*



1,289 enrolled Sickle Cell Disease patients across Europe

Netherlands (SCORE, 4 sites)
487 patients (45% PED)
Deployment
Virtual Machine
Connected to Central Node

Spain (INTEGRA, 10 sites)
334 patients (76% PED)
Deployment
Virtual Machine
Connected to Central Node

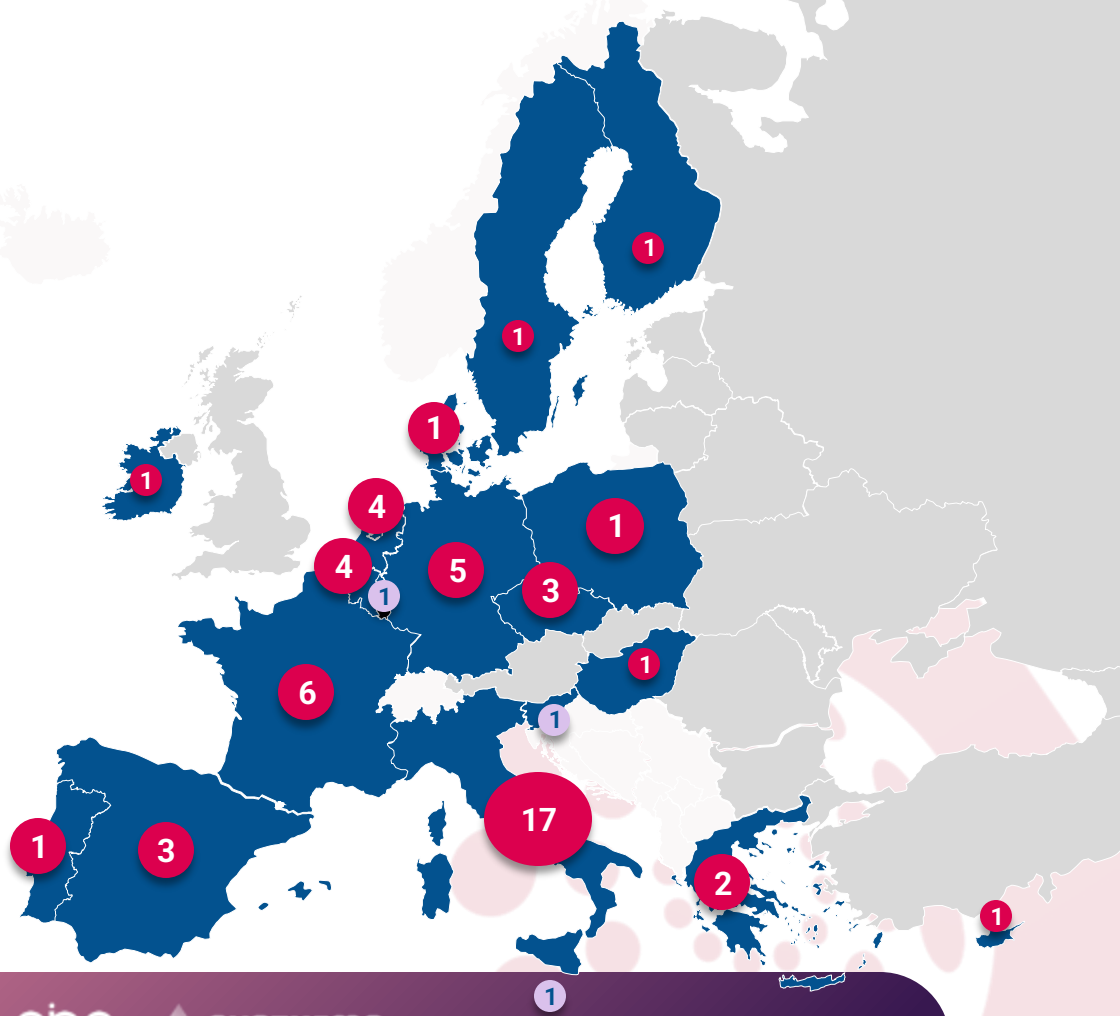
RADiANT
Central Node
Federated AI
Data Harmonisation
Secure Computation

Italy (1 site)
130 patients (75% PED)
Deployment
Virtual Machine
Connected to Central Node

France (2 sites)
494 patients (81.6% PED)
Deployment
Virtual Machine
Connected to Central Node

ERN | EuroBloodNet

RBC subnetwork



Total Members & APs: 55

- Members (52)
- Affiliated Partners (3)

Kidney Injury- Real World Evidence

RADeep Statistical Analysis

SCD Cross-sectional descriptive analysis - Secondary outcomes

▼ **Table of Contents**

▼ **List of Tables**

Missing data

- Statistical analysis includes only participants without missing data on the variables of interest for each analysis. For the avoidance of doubt, no imputation was performed in cases of missing data for this statistical analysis.
- In all cases "unknown" category is treated as missing value.
- In the case of data providers giving information through the Local solution (not RADeep- REDCap), for the porpoise of this analysis, in the cases only the occurrence of cases ('Yes' answer) could had been mapped, the missing value was consider as 'No'.
- To describe categorical variables the results in each stratum are described using number of observations or subjects with a certain stratum category (n) out of the total number of observations or patients in each stratum (N). In the case of continuous variables, the median with IQR and mean with SD will be shown.

Period of data collection

- Time frame of the analytic population: Index date or last follow-up of the patient between 2021-12-01 and 2025-11-30

Secondary outcomes

Portugal was excluded from this section as it has started entering information on November 2025 due to delays with the Ethics Committee. Currently three hospitals have received the Ethics Committee approval but data completeness is not enough to include this country in the analysis of secondary outcomes.

For chronic organ damage, consider that in the case of data providers giving information through the Local solution (not RADeep- REDCap), for certain parameters, a value of 'No' has been assumed when the opposite option 'Yes' was not selected. For further information, review the coverage report.

▼ Clinical outcomes

Clinical outcomes results are presented in total by country and sex at birth.

To describe categorical variables the results in each stratum are described using number of observations or subjects with a certain stratum category (n) out of the total number of observations or patients in each stratum (N). In the case of continuous variables, the median with IQR and mean with SD will be shown.

Results are shown by total SCD patients and stratified by age pediatric/adult and the categories 0-11, 12-17, 18-44, 45 years or more, sub diagnosis (Sickle cell anaemia and Sickle cell-haemoglobin C disease syndrome) and genotype (SS, SB0, SC, SB+).

Kidney Injury- Real World Evidence

- Kidney Injury in SCA patients

Table 57. SCA patients (N= 2826) - Chronic organ damage – by country and sex at birth

Group	Total									Male								Female							
	Overall N = 2,826[1]	Belgium N = 832[1]	Cyprus N = 2[1]	Denmark N = 23[1]	France N = 743[1]	Greece N = 5[1]	Italy N = 426[1]	Netherlands The N = 277[1]	Spain N = 518[1]	Overall N = 1,374[1]	Belgium N = 397[1]	Cyprus N = 2[1]	Denmark N = 8[1]	France N = 339[1]	Greece N = 3[1]	Italy N = 220[1]	Netherlands The N = 144[1]	Spain N = 261[1]	Overall N = 1,452[1]	Belgium N = 435[1]	Cyprus N = 0[1]	Denmark N = 15[1]	France N = 404[1]	Greece N = 2[1]	Italy N = 206[1]	Netherlands The N = 133[1]	Spain N = 257[1]
Microalbuminuria	188/311 (60.5%)	44/44 (100.0%)	0/2 (0.0%)	0/1 (0.0%)	79/173 (45.7%)	0/1 (0.0%)	18/27 (66.7%)	47/63 (74.6%)	0/0 (-)	70/125 (56.0%)	17/17 (100.0%)	0/2 (0.0%)	0/0 (-)	26/61 (42.6%)	0/1 (0.0%)	6/12 (50.0%)	21/32 (65.6%)	0/0 (-)	118/186 (63.4%)	27/27 (100.0%)	0/0 (-)	0/1 (0.0%)	53/112 (47.3%)	0/0 (-)	12/15 (80.0%)	26/31 (83.9%)	0/0 (-)
Renal insufficiency	77/275 (28.0%)	1/2 (50.0%)	0/2 (0.0%)	1/1 (100.0%)	72/172 (41.9%)	0/1 (0.0%)	1/29 (3.4%)	2/63 (3.2%)	0/5 (0.0%)	25/115 (21.7%)	1/1 (100.0%)	0/2 (0.0%)	0/0 (-)	23/61 (37.7%)	0/1 (0.0%)	1/14 (7.1%)	0/32 (0.0%)	0/4 (0.0%)	52/160 (32.5%)	0/1 (0.0%)	0/0 (-)	1/1 (100.0%)	49/111 (44.1%)	0/0 (-)	0/15 (0.0%)	2/31 (6.5%)	0/1 (0.0%)

Table 58. SCA adult patients (N= 1087) - Chronic organ damage – by country and sex at birth

Group	Total									Male								Female							
	Overall N = 1,087[1]	Belgium N = 365[1]	Cyprus N = 2[1]	Denmark N = 22[1]	France N = 317[1]	Greece N = 4[1]	Italy N = 83[1]	Netherlands The N = 163[1]	Spain N = 131[1]	Overall N = 444[1]	Belgium N = 145[1]	Cyprus N = 2[1]	Denmark N = 8[1]	France N = 115[1]	Greece N = 3[1]	Italy N = 35[1]	Netherlands The N = 79[1]	Spain N = 57[1]	Overall N = 643[1]	Belgium N = 220[1]	Cyprus N = 0[1]	Denmark N = 14[1]	France N = 202[1]	Greece N = 1[1]	Italy N = 48[1]	Netherlands The N = 84[1]	Spain N = 74[1]
Microalbuminuria	173/268 (64.6%)	40/40 (100.0%)	0/2 (0.0%)	0/1 (0.0%)	76/156 (48.7%)	0/1 (0.0%)	11/12 (91.7%)	46/56 (82.1%)	0/0 (-)	64/99 (64.6%)	15/15 (100.0%)	0/2 (0.0%)	0/0 (-)	25/51 (49.0%)	0/1 (0.0%)	3/3 (100.0%)	21/27 (77.8%)	0/0 (-)	109/169 (64.5%)	25/25 (100.0%)	0/0 (-)	0/1 (0.0%)	51/105 (48.6%)	0/0 (-)	8/9 (88.9%)	25/29 (86.2%)	0/0 (-)
Renal insufficiency	77/233 (33.0%)	1/2 (50.0%)	0/2 (0.0%)	1/1 (100.0%)	72/156 (46.2%)	0/1 (0.0%)	1/14 (7.1%)	2/56 (3.6%)	0/1 (0.0%)	25/88 (28.4%)	1/1 (100.0%)	0/2 (0.0%)	0/0 (-)	23/51 (45.1%)	0/1 (0.0%)	1/5 (20.0%)	0/27 (0.0%)	0/1 (0.0%)	52/145 (35.9%)	0/1 (0.0%)	0/0 (-)	1/1 (100.0%)	49/105 (46.7%)	0/0 (-)	0/9 (0.0%)	2/29 (6.9%)	0/0 (-)

Table 59. SCA pediatric patients (N= 1739) - Chronic organ damage – by country and sex at birth

Group	Total									Male								Female							
	Overall N = 1,739[1]	Belgium N = 467[1]	Cyprus N = 0[1]	Denmark N = 1[1]	France N = 426[1]	Greece N = 1[1]	Italy N = 343[1]	Netherlands The N = 114[1]	Spain N = 387[1]	Overall N = 930[1]	Belgium N = 252[1]	Cyprus N = 0[1]	Denmark N = 0[1]	France N = 224[1]	Greece N = 0[1]	Italy N = 185[1]	Netherlands The N = 65[1]	Spain N = 204[1]	Overall N = 809[1]	Belgium N = 215[1]	Cyprus N = 0[1]	Denmark N = 1[1]	France N = 202[1]	Greece N = 1[1]	Italy N = 158[1]	Netherlands The N = 49[1]	Spain N = 183[1]
Microalbuminuria	15/43 (34.9%)	4/4 (100.0%)	0/0 (-)	0/0 (-)	3/17 (17.6%)	0/0 (-)	7/15 (46.7%)	1/7 (14.3%)	0/0 (-)	6/26 (23.1%)	2/2 (100.0%)	0/0 (-)	0/0 (-)	1/10 (10.0%)	0/0 (-)	3/9 (33.3%)	0/5 (0.0%)	0/0 (-)	9/17 (52.9%)	2/2 (100.0%)	0/0 (-)	0/0 (-)	2/7 (28.6%)	0/0 (-)	4/6 (66.7%)	1/2 (50.0%)	0/0 (-)
Renal insufficiency	0/42 (0.0%)	0/0 (-)	0/0 (-)	0/0 (-)	0/16 (0.0%)	0/0 (-)	0/15 (0.0%)	0/7 (0.0%)	0/4 (0.0%)	0/27 (0.0%)	0/0 (-)	0/0 (-)	0/0 (-)	0/10 (0.0%)	0/0 (-)	0/9 (0.0%)	0/5 (0.0%)	0/3 (0.0%)	0/15 (0.0%)	0/0 (-)	0/0 (-)	0/0 (-)	0/6 (0.0%)	0/0 (-)	0/6 (0.0%)	0/2 (0.0%)	0/1 (0.0%)

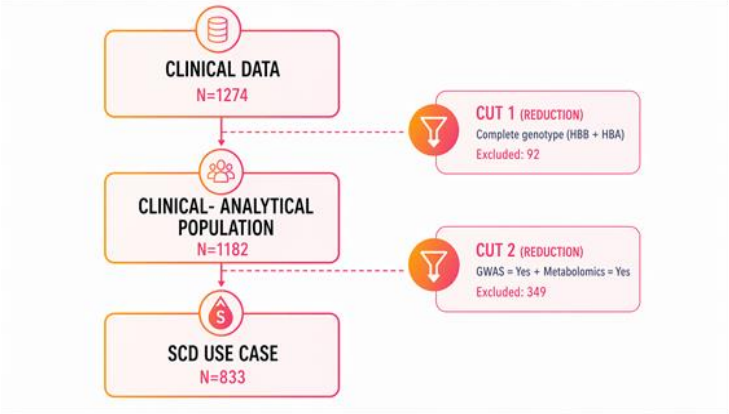
Kidney Injury

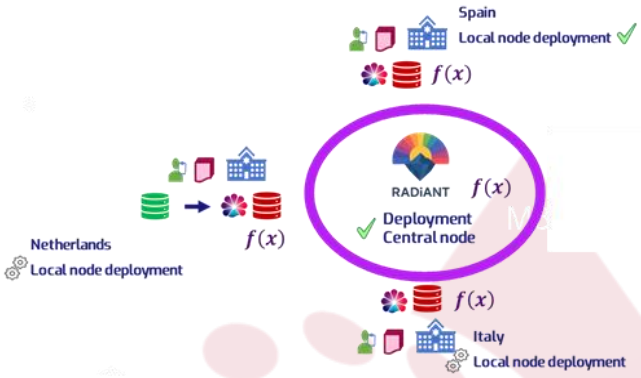
Use case: Augmentation of SCD patients with Kidney injury (KI) based on real world data

1. Patient collection → 2. Population Selection Flow → 3. Federated Architecture Deployment

1,445 SCD Patients enrolled

- France (2 site) = 494 (37.6% PED)
- Netherlands (SCORE 4 sites) = 487 (45% PED)
- Spain (INTEGRA 10 sites) = 334 (76% PED)
- Italy (1 site) = 130 (75% PED)

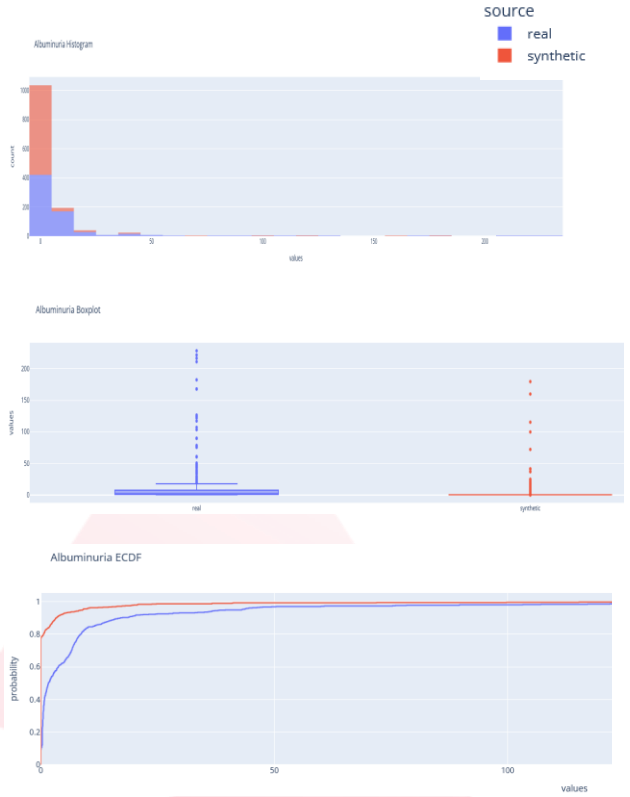




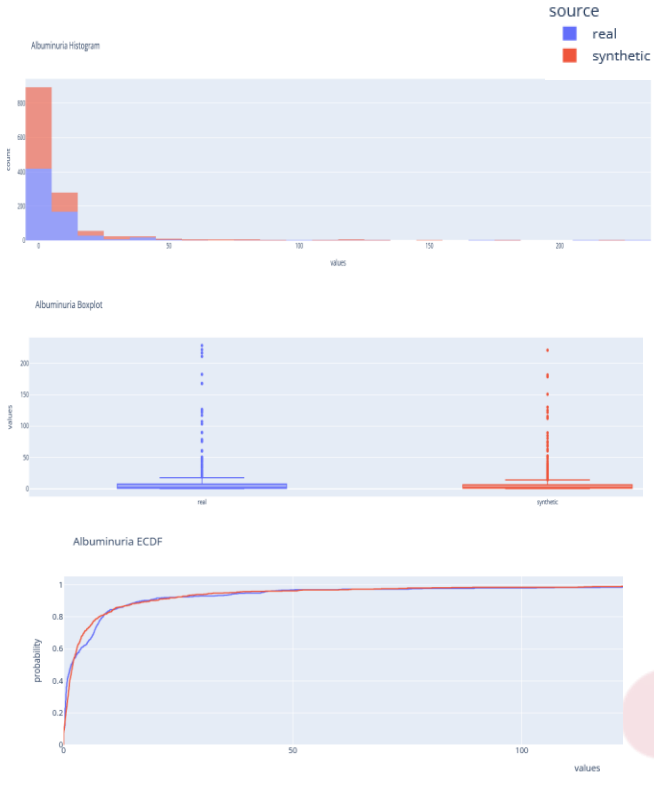
- 64 clinical variables related to KI
- 10 SNPs related to HbF
- 13 metabolites related to glycolysis, energy metabolism, TCA cycle, purine turnover, oxidative stress, nitric oxide, vascular function, pentose phosphate pathway, HU treatment marker

Kidney Injury- Synthetic data

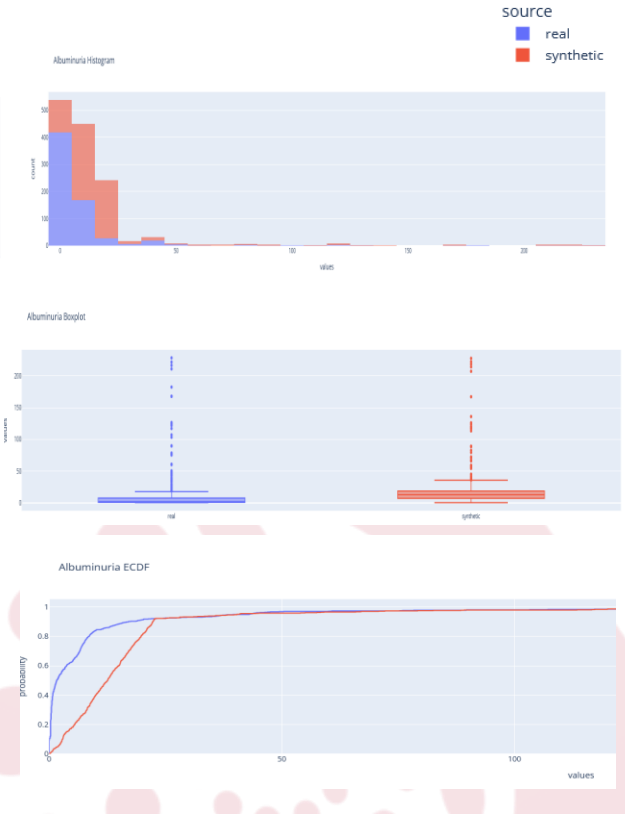
- CTGAN



- VAE



- Bayesian networks



Validation repositories

SCD – Statistical Fidelity and Clinical Utility

Repository link: https://github.com/synthema-project/scd-fidelity-utility-metrics

Fidelity Metrics

- Hellinger distance
- Mann-Whitney & Chi-Squared tests
- Correlation preservation
- ML Distinctiveness (AUC-ROC)

Utility Metrics

- Feature importance rank correlation
- Odds ratios & SMD
- Range validation compliance
- Branching logic compliance

Clinical Validation

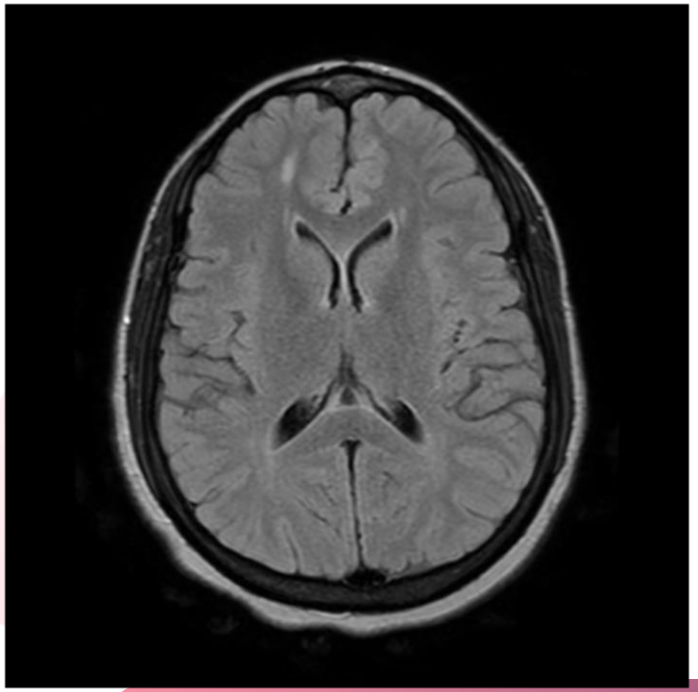
- Albuminuria distribution tests
- Kidney injury × genotype
- Kidney injury × country
- Domain-specific SCD rules

Multi-Country Evaluation metrics

- UBL – aggregated evaluation
- LBL – per-country evaluation
- Config-driven (JSON)
- Batch processing support

Validation of Synthetic SCD MRI

Evaluation Pipeline Overview

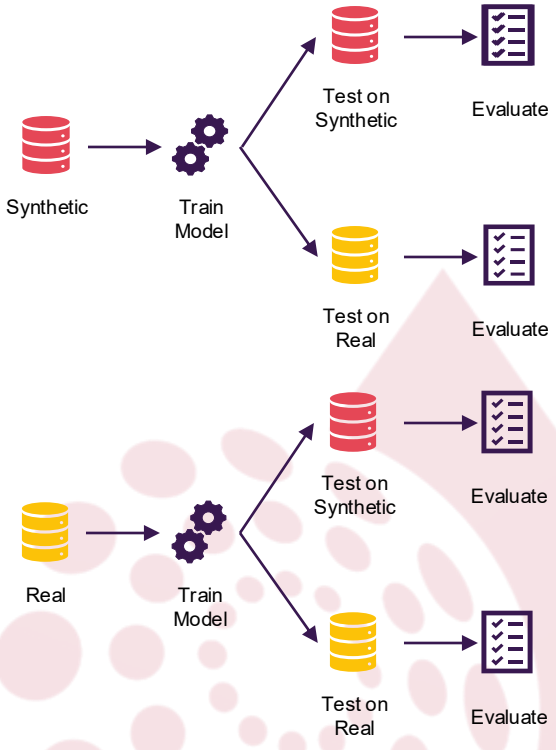


- **Dataset** FLAIR + T1 MRIs with possible presence of White Matter Hiperintensities (WMH)
- **Segmentation Tools** WMH segmentator model developed in GenoMed4All
- **Generator Tools** 3D-Diffusion Model, Movie-based generator
- **Evaluation** Comparison of segmentation tools results onto real and synthetic MRIs

Validation of Synthetic SCD MRI

Evaluation Schema

- Train segmentation model onto synthetic/real data and test it onto synthetic/real data comparing performances
- Blinded calssification of synthetic and real imagesto identify any perceptual discrepancies.
- Compute standard metrcis (i.e. Frechet Inception Disntace) to measure similarity between real and synthetic data



Thanks!

Any questions?

Keep in touch!

eurobloodnet.eu /ERNEuroBloodNet @ERNEuroBloodNet @erneurobloodnet.bsky.social

synthema.eu /synthema @SYNTHEMA_EU @synthemaeu.bsky.social

Funded by
the European Union

Acknowledgements

European Reference Network
for rare or low prevalence complex diseases
Network Hematological Diseases (ERN EuroBloodNet)

Funded by the European Union

This project is supported by the European Reference Network on Rare Haematological Diseases (ERN-EuroBloodNet)-Project ID No 101085717. ERN-EuroBloodNet is partly co-funded by the European Union within the framework of the Fourth EU Health Programme.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

SYNTHEMA

Funded by the European Union

SYNTHEMA is an initiative funded by the European Union's Horizon Europe Research and Innovation programme under grant agreement No. 101095530.