FinOMOP Swarm Learning:
Deep learning for patient-specific modelling of Acute Myeloid Leukemia based on longitudinal clinical laboratory data and OMOP

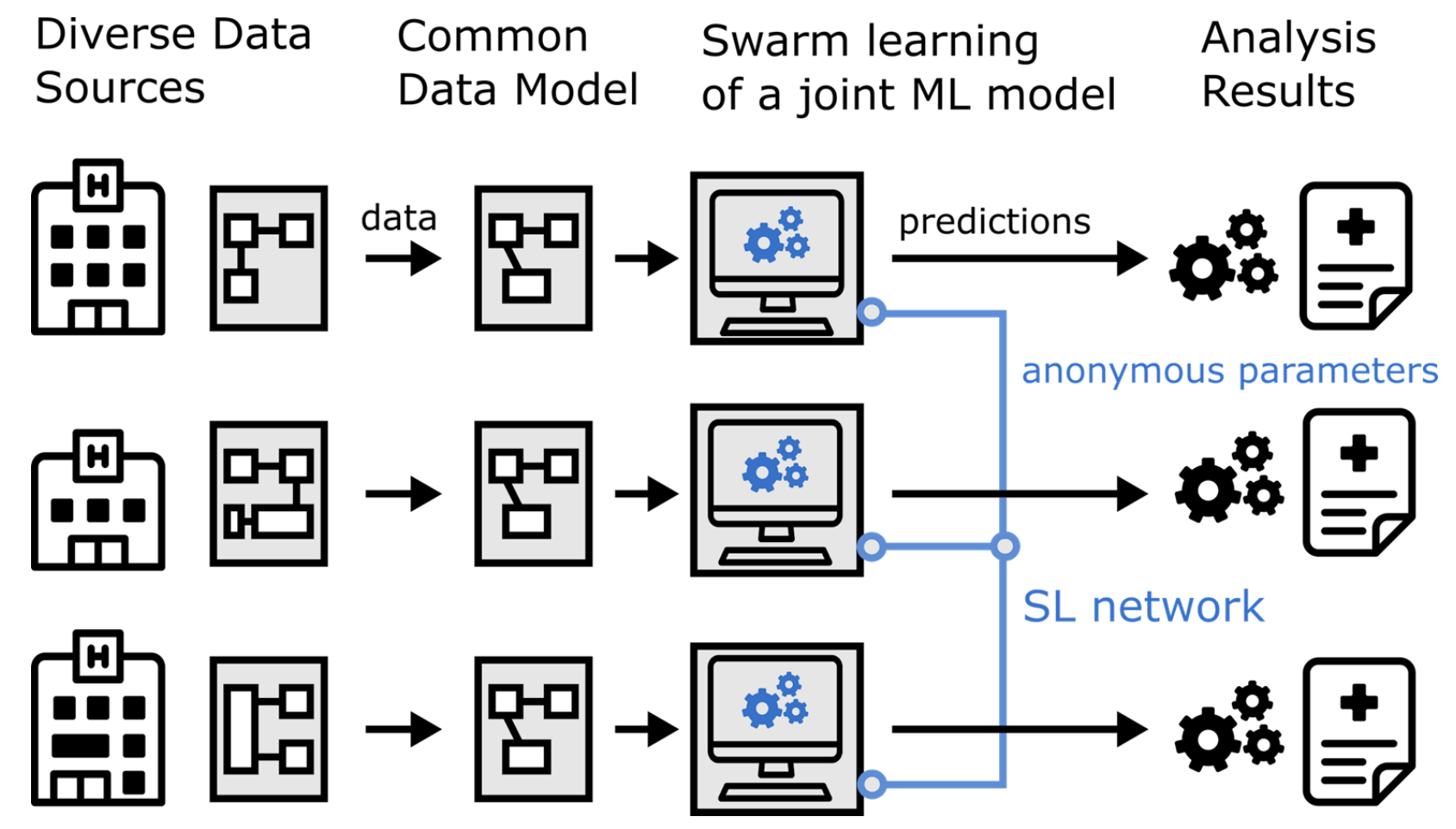
PRESENTER: **Eric Fey**

INTRO:

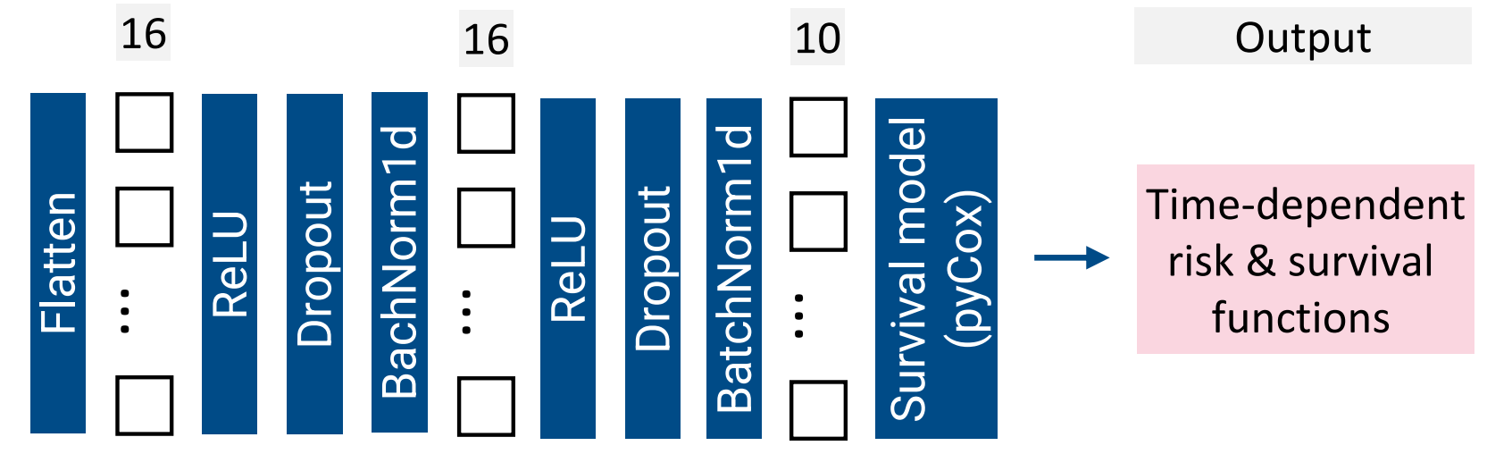
- Deep learning is powerful but data hungry!
- Data sharing hindered by privacy concerns
- Federated learning hindered by heterogeneous data formats
- Solution: Use common data model and learn models *at the edge* without sharing source data using Swarm learning (SL)

METHODS

1. OMOP + Swarm Learning => Joint model



2. Cohort: AML use case
   All patients with AML diagnoses and at least 3 blast measurements within 21 days of diagnosis
3. Data
   Endpoint: Overall patient survival
   Features: Blood count measurements, timepoints up to 21 days after diagnosis
4. Model architecture v1:
   pyTorch, pyCox, DeepHit



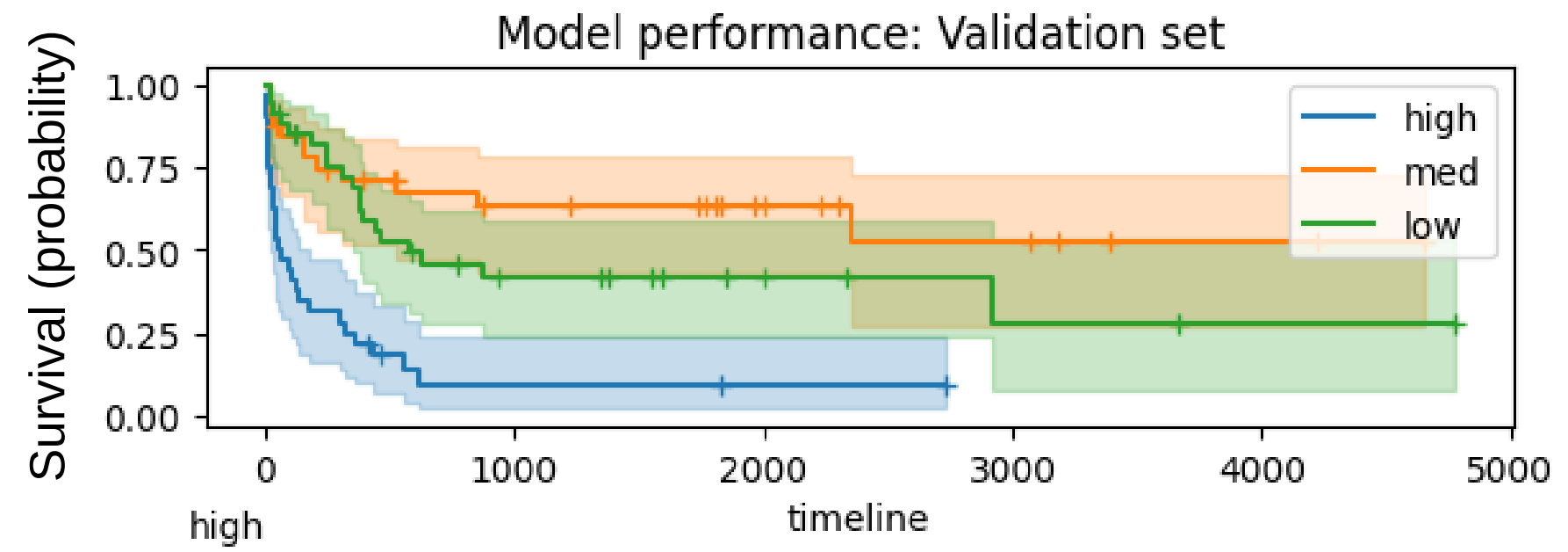
RESULTS

Developed FinOMOP-SL framework:

- Findata compliant implementation at three sites: **Helsinki**, **Turku**, **Tampere**
- Integrating OMOP & swarm learning

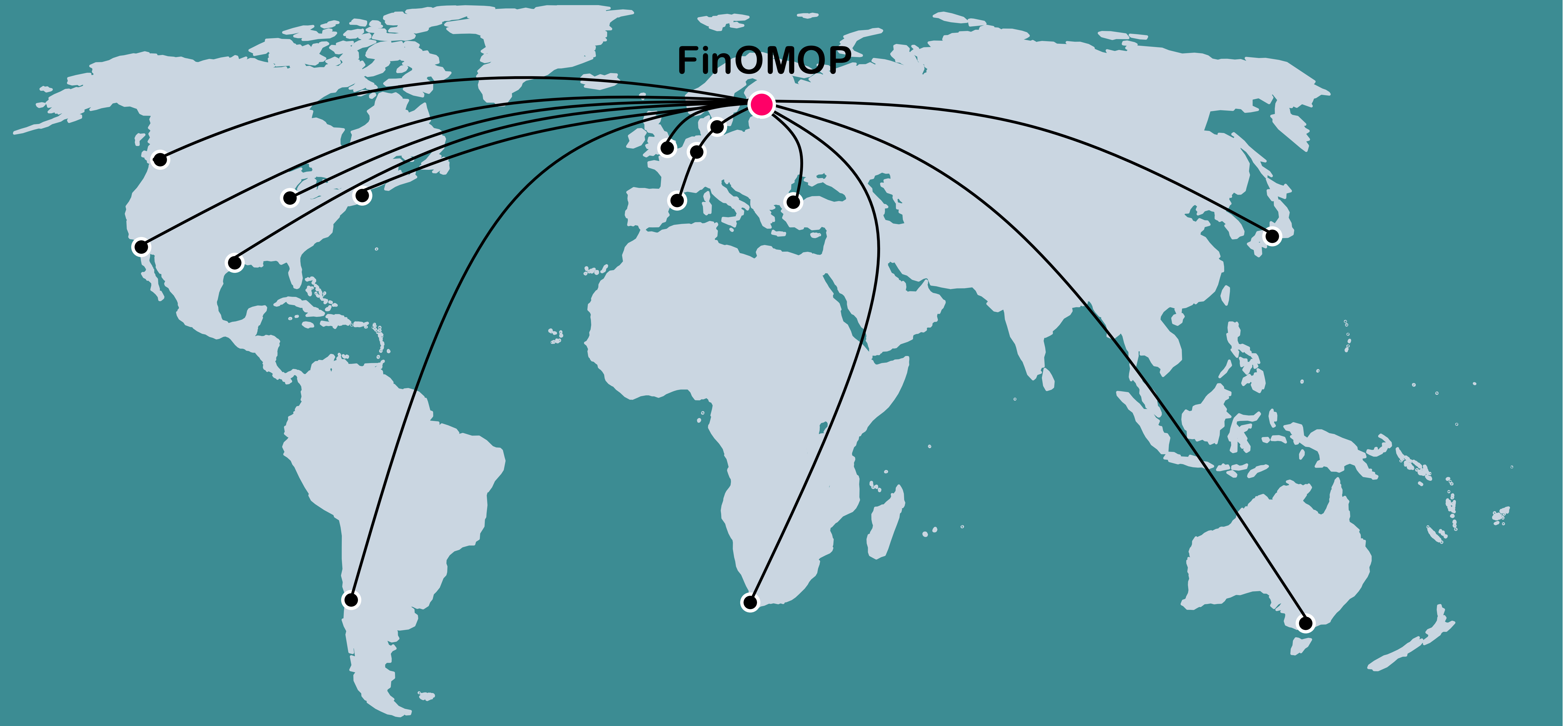
Proof of concept: use case AML

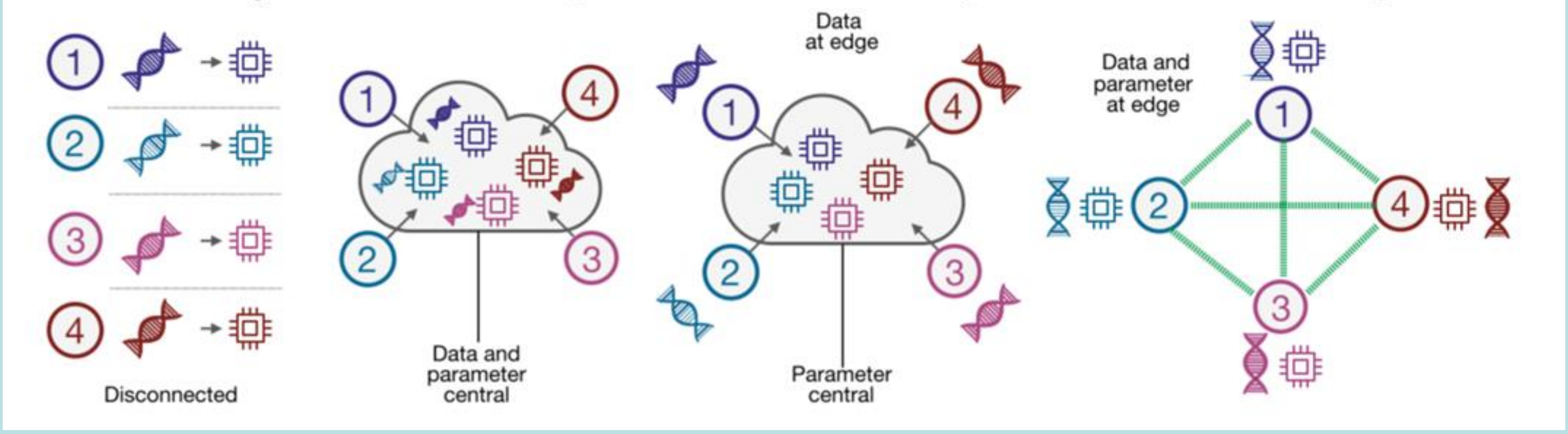
- Predictive power model v1, so far only trained on HUS data::



high						
At risk	31	2	1	0	0	0
Censored	0	2	3	4	4	4
Events	1	28	28	28	28	28
med						
At risk	32	15	9	5	2	0
Censored	0	6	12	15	18	20
Events	0	11	11	12	12	12
low						
At risk	34	10	5	2	1	0
Censored	0	6	11	13	14	15
Events	0	18	18	19	19	19

Build deep, predictive models for precision medicine together in global networks.





- Limited dataset
- Bias
- Low accuracy
- Disconnected

+ Enlarged dataset
+ Better accuracy
- Aggregation
- Monopolisation

+ Enlarged dataset
+ Better accuracy
+ No data movement
+ No central custodian
+ Shared insights

AMMO BAR

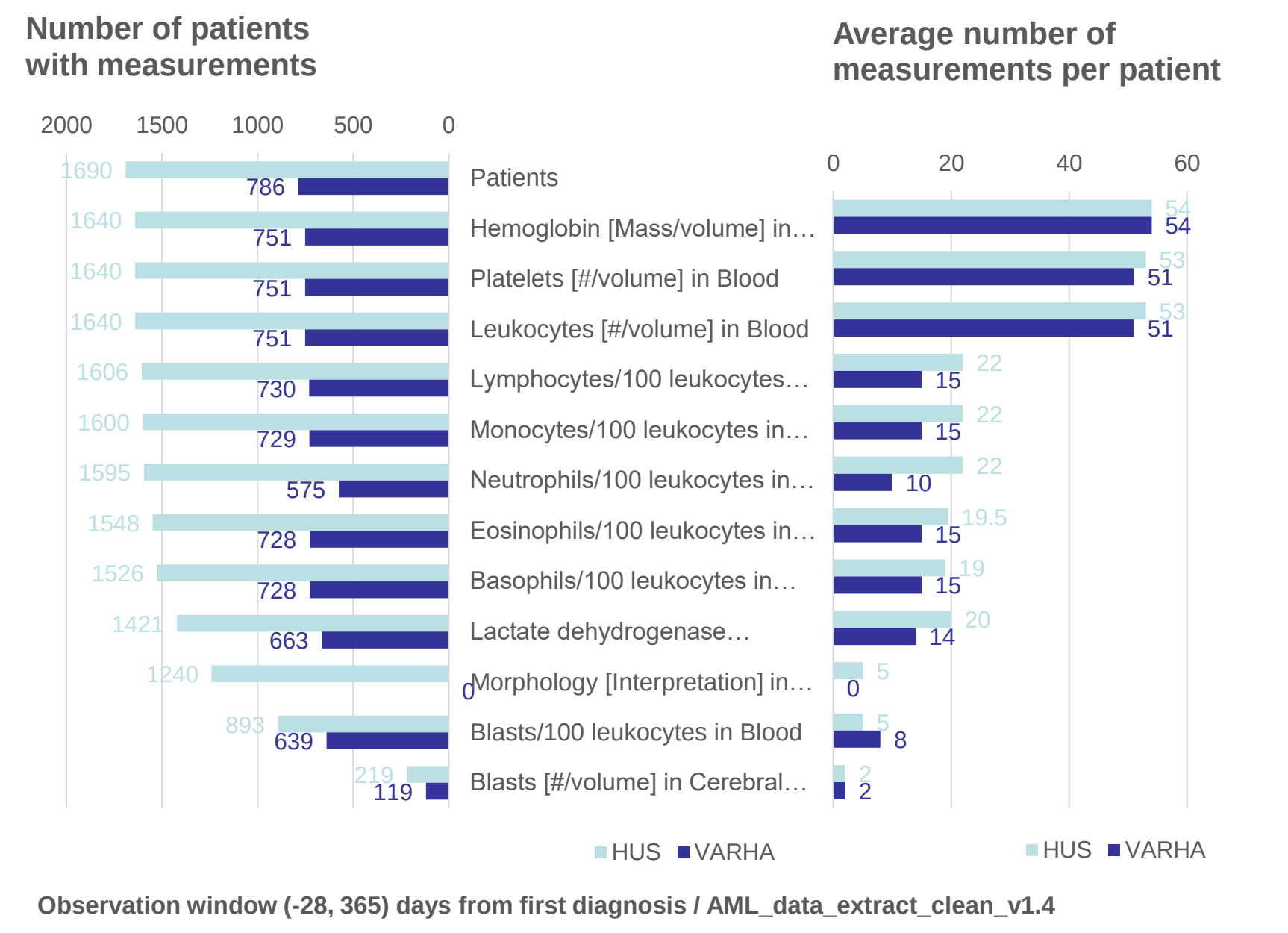
CONCLUSIONS

- Routine short-term follow up data – blood count measurements (21 days) – can predict long-term prognosis in AML
- Federated/Swarm learning of a joint model feasible
- Routine data available at all three sites

Next:

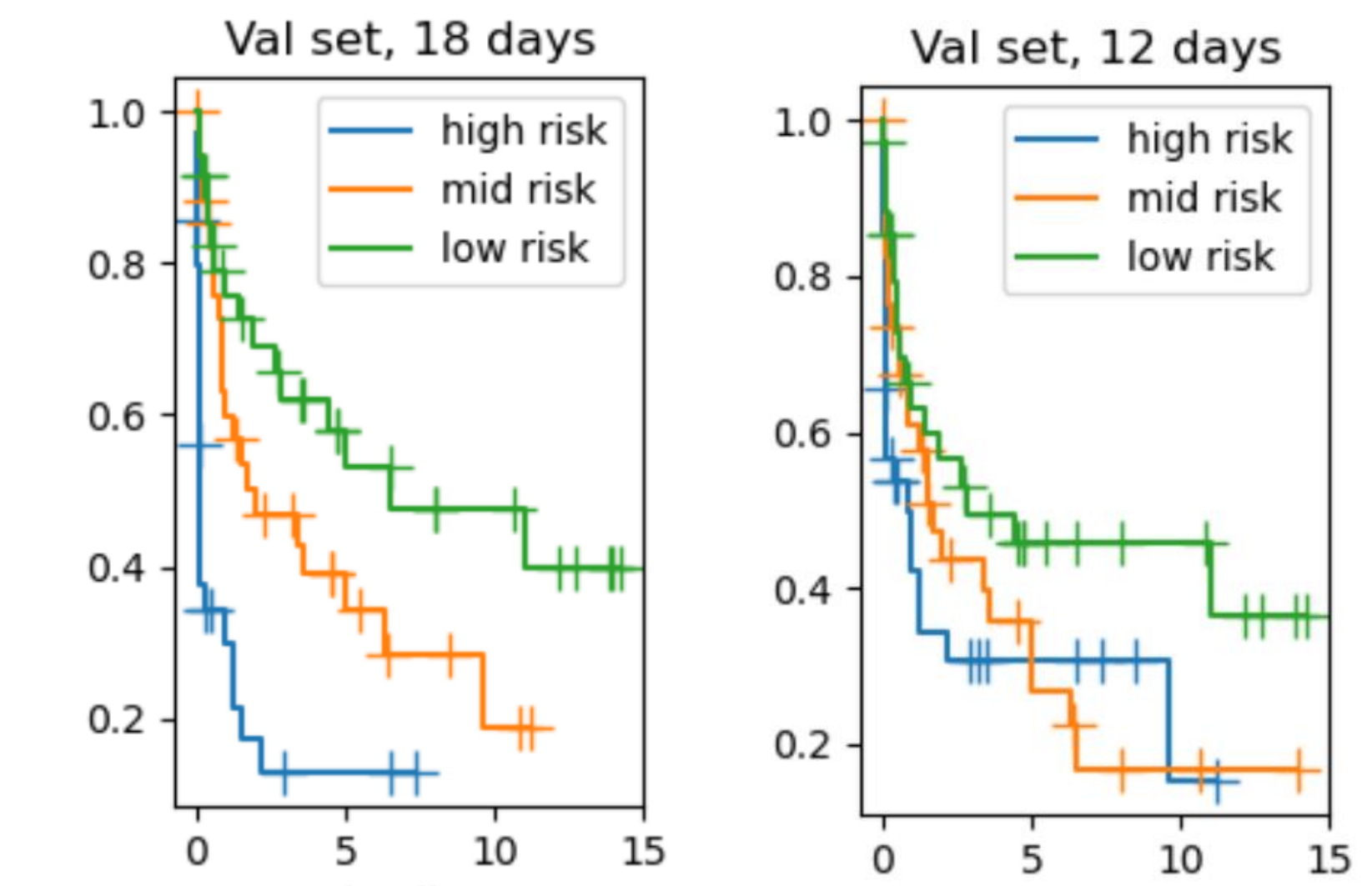
- Estimate and include pCR & RFS
- Integrate genomic data
- Advanced model architectures, e.g. autoencoders + survival models

Data availability HUS & VARHA (Turku)



Observation window (-28, 365) days from first diagnosis / AML_data_extract_clean_v1.4

Performance based availability of timepoints



Val set, 18 days

high risk				
At risk	34	2	0	0
Censored	0	6	8	8
Events	1	27	27	27
mid risk				
At risk	35	7	2	0
Censored	0	8	11	13
Events	0	20	22	22
low risk				
At risk	35	11	7	0
Censored	0	10	13	19
Events	0	14	15	16

Val set, 12 days

high risk			
34	5	1	0
0	8	11	12
1	22	23	23
mid risk			
35	6	2	0
0	7	9	11
0	22	24	24
low risk			
35	9	6	0
0	9	12	17
0	17	17	18

Eric Fey[1,2] (eric.fey@hus.fi), Salma Rachidi[1], Alexey Ryzhenkov[1], Valtteri Nieminen[2], Tomi Mäkelä[1], Oscar Brück[1,2], Kimmo Porkka[1,2]
[1]HUS University Hospital, Finland
[2]iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki, Finland