BRIEF OVERVIEW ON THE COPERIMO*PLUS* PROJECT

COPERIMO stands for COVID-19 Personalized Risk Models. The major aim of the project is the training, the optimization and validation of personalized risk models for SARS-CoV-2 infected subjects.

COPERIMO plus

THE FUNDAMENTAL CONCEPT OF COPERIMO*PLUS*

COPERIMO*plus* was driven by the following fundamental considerations:

- The human response to SARS-CoV-2 infection is very heterologous. In particular, risk for fatal progression of disease in elderly people is high.
- A wide spectrum of co-morbidities and population-level risk factors have been identified (incl. age and obesity)
- However, despite generalizable observations of statistical significance, we observe fatal progression of disease also in younger patients and even in re-infected patients (surviving the first infection and dying of the second one).
- A wide spectrum of treatments is currently being tested in clinical trials; global trial registries list more than 1000 COVID-19 trials (observational and interventional)
- If it was possible to collect a substantial amount of trial and study data and to use them for training of an AI model, we should be able to come up with a system that allows for personalized risk assessment
- …. and that personalized risk assessment is exactly, what COPERIMO*plus* aims at

COPERIMO*PLUS* TASKS

COPERIMO*plus* is organised in different tasks:

- Systematic approaches towards collecting – at global scale - longitudinal, patient-level data.
- Development of "GECCO*plus*" a data model for COVID-19 patient-level data that comprises all variables relevant for machine learning / AI approaches
- Systematic curation and enhancement of interoperability of COVID-19 patient-level data
- Development of adaptive concepts for distributed (federated) learning of COVID-19 risk models: dealing with data protection rules in the most flexible and adaptive way
- Training, optimization and validation of advanced AI models for personalized risk assessment of SARS-CoV-2 infections

DATA WRANGLING

COPERIMO plus

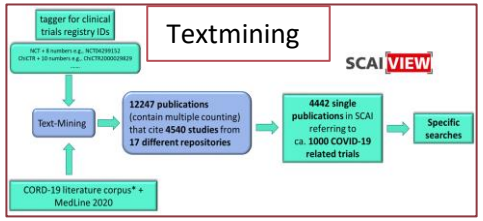
DATA CURATION & DATA WRANGLING

collection of Covid-19 datasets → structured dataset catalogue → data acquisition → detailed data curation → global data model for Covid-19 datasets

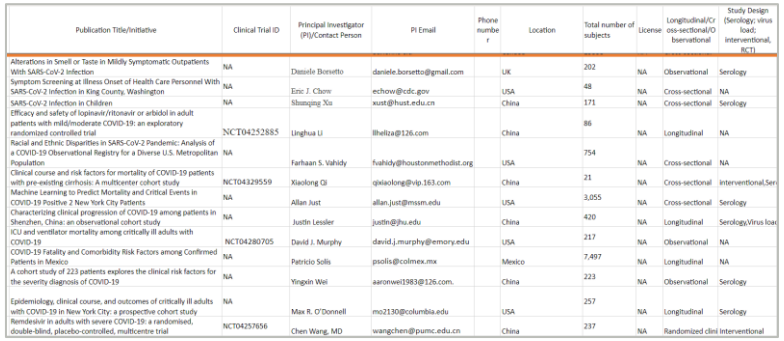
publications

global initiatives e.g.,
www.covidanalytics.io/
https://covid-evidence.org/

completed clinical trails from repositories such as
https://clinicaltrials.gov/

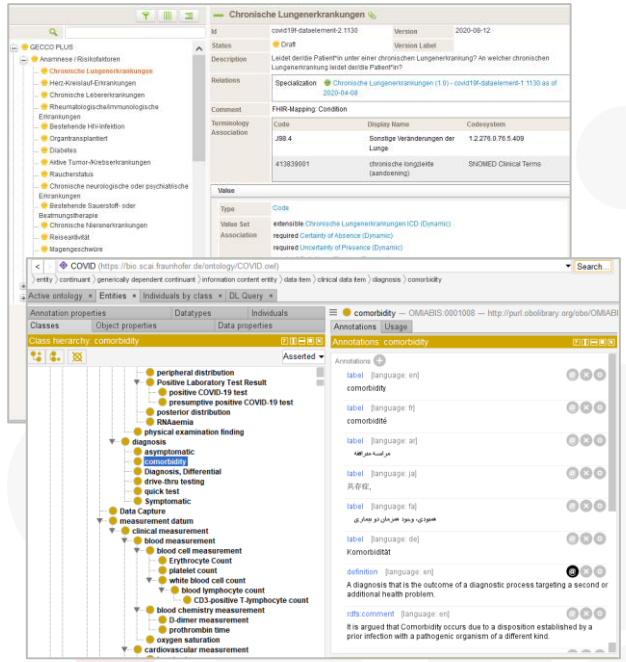


- > 800 entries
- ranked
  - study type
  - availability
  - Patient numbers



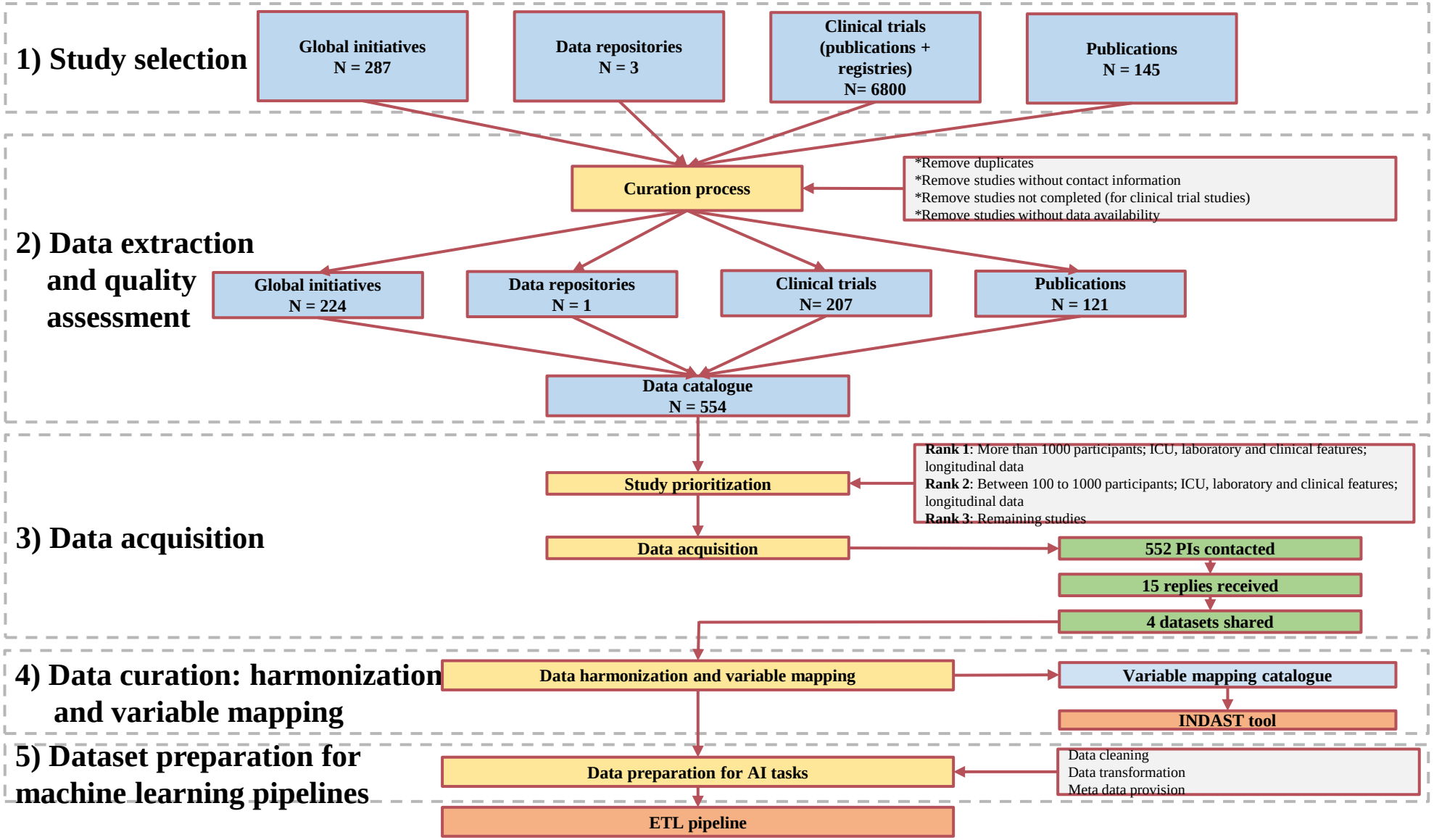
Systematic recording of variables

- endpoints,
- questions and
- study
- protocols
- inclusion / exclusion criteria
- statistical power estimates



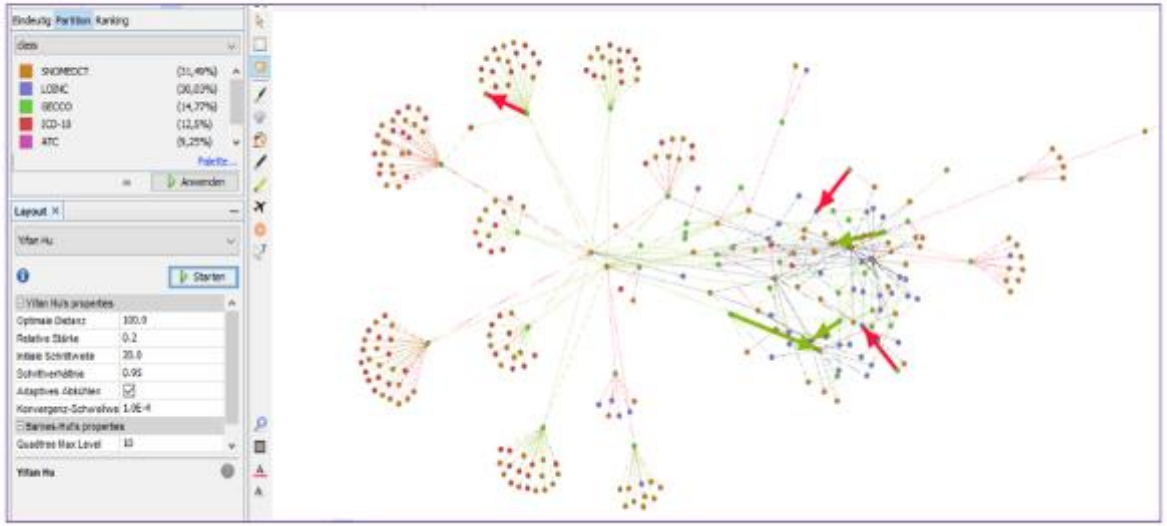
Fraunhofer SCAI · Fraunhofer IME · Fraunhofer IGD · Fraunhofer IKTS · Fraunhofer MEVIS · Fraunhofer IAIS

QUALITY ASSESSMENT AND ”AI-READINESS” OF STUDY DATA



INTEGRATIVE DATA SEMANTICS: COMMON COVID DATA MODEL

Integration of the COVID-19 reference data model GECCO+ with controlled vocabularies ICD-10, SNOMED and LOINC to map&import study data:



Semi-automated Mapping: 5.500 distinct variables of clinical studies are mapped to GECCO++ offering a combination of automated and manual methods

Wegner, P., Jose, G. M., Lage-Rupprecht, V., Golriz Khatami, S., Zhang, B., Springstubbe, S., ... & Hofmann-Apitius, M. (2022). Common data model for COVID-19 datasets. *Bioinformatics*.

https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btac651/6776111

DATA STEWARDSHIP AND DATA CURATION

COPERIMO plus

INTELLIGENT DATA STEWARDSHIP MODULE

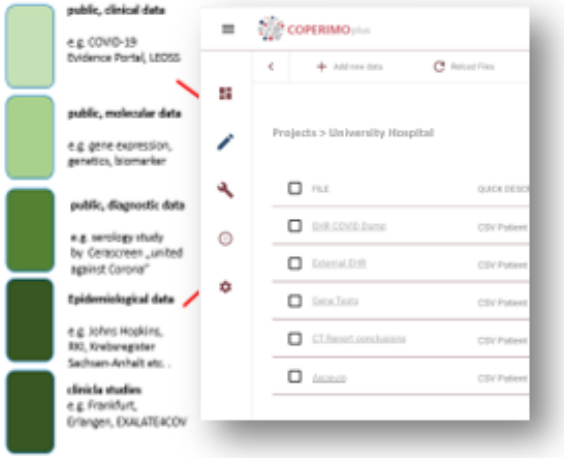
Data Sources

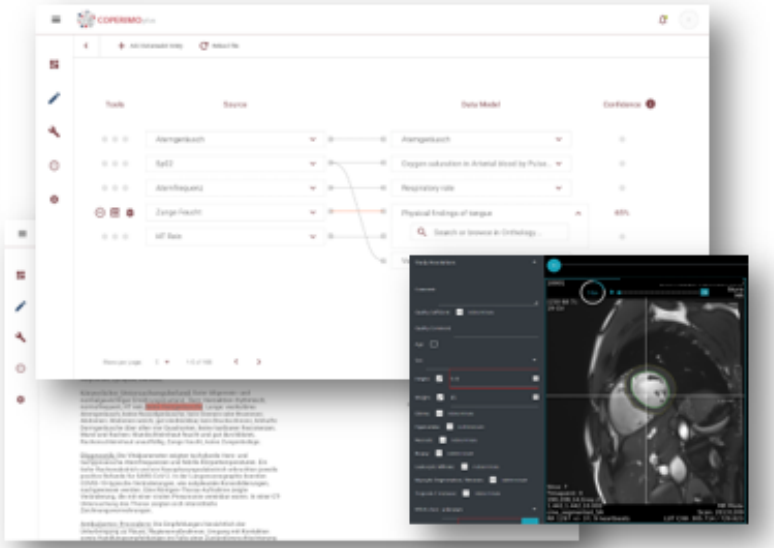
AI Support

Mapping & Data Curation
Annotation & QualityAssessment

Prepare & Foward

Modelling & Risk Analyses





Web Interface

- AI Mapping Tool
- AI Quality Assurance
- Annotation Tasks
- General Data Curation

Wegner, P., Schaaf, S., Uebachs, M., Domingo-Fernández, D., Salimi, Y., Gebel, S., ... & Kodamullil, A. T. (2022). Integrative data semantics through a model-enabled data stewardship. *Bioinformatics*, *38*(15), 3850-3852.

AI FOR PERSONALIZED RISK ASSESSMENT

COPERIMO plus

AI FOR PERSONALIZED RISK MODELS

- Individual patient

SARS-Cov-2 test

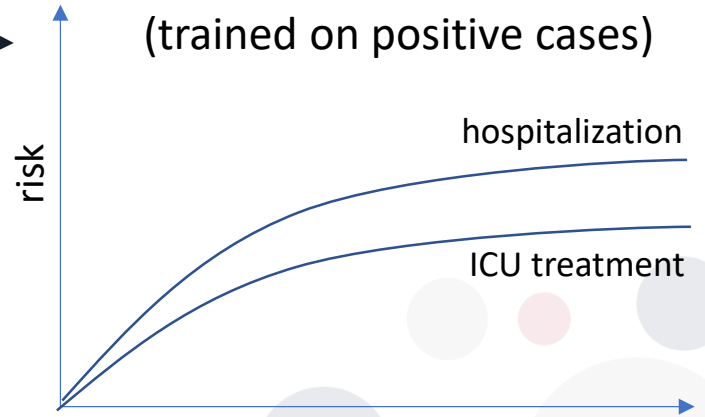
- Comorbidities
- Clinical assessment
- Demographics
- Smoking status
- Biomarkers
- Imaging
- …

Test negative → What, if the patient would have been tested positive?

Test positive → Treat and observe

AI learns from every single patient

AI based risk model
(trained on positive cases)



High risk patient

Low risk patient

Strong protection: (Self-) Quarantine; Repeated testing

Weak protection: back to normal life

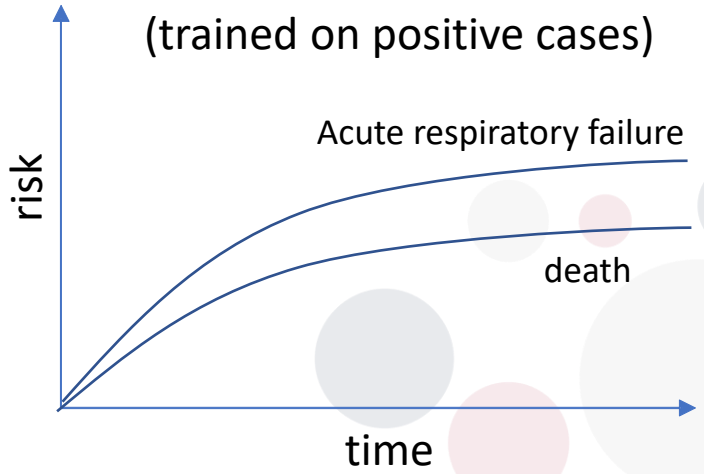
→ Better planning of hospital capacities

AI FOR PERSONALIZED RISK MODELS

- **Better planning of intensive care unit capacities**
  (based on hospital Covid-19 patient data)

  - Risk of artificial ventilation
  - … acute respiratory failure
  - … sepsis
  - … death

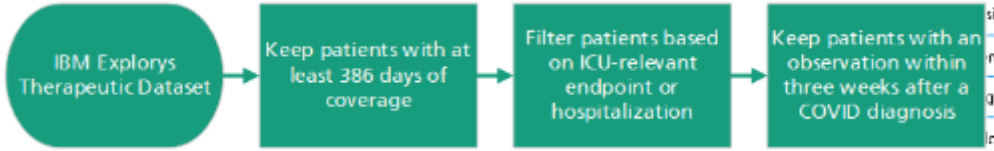
AI based risk model
(trained on positive cases)



AI learns from every single patient

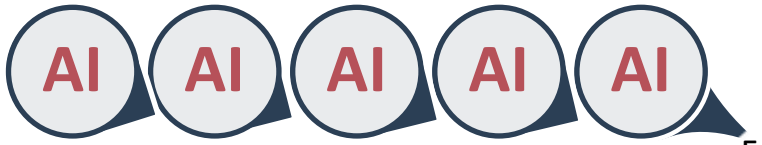
Results (IBM Data): Clinical Routine Data Allow for Predicting Acute Respiratory Manifestation and Hospitalization

Datasets cleaned and pre-processed



Subsets	No. of Patients	No. of COVID Patients
Raw	4563769	207876
Sufficient Coverage	3632185	101046
Acute Respiratory Manifestation	651356	13625
Hospitalization	1263146	12748

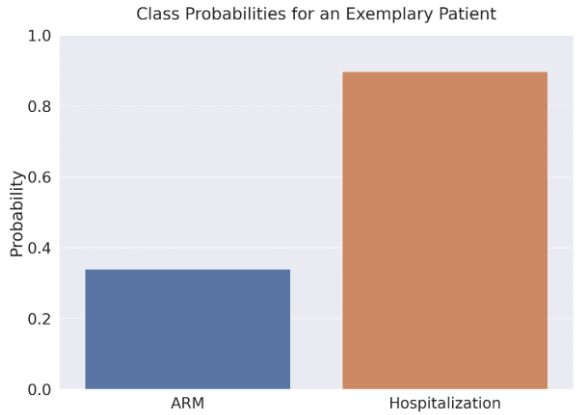
Wide range of AI algorithms implemented and evaluated

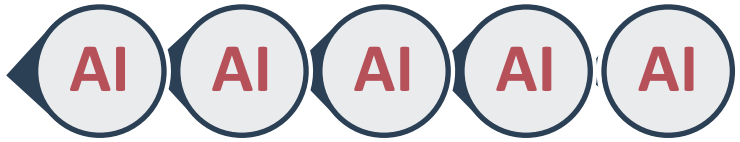


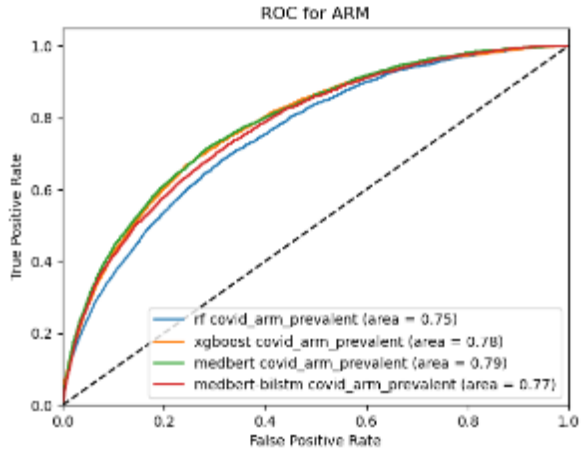
Feature extraction via AI (MedBERT)

Fine tuning / comparison of AI models

Personalized risk predictions







Results (LEOSS):
AI Based Mortality Prediction for German Hospital Patients

COPERIMO plus

Data: 5677 patients (80 days follow-up)

Age	
18 - 25 years	181
26 - 35 years	472
36 - 45 years	540
46 - 55 years	907
56 - 65 years	1,125
66 - 75 years	980
76 - 85 years	1,231
Gender	
Male	3,228
Female	2,218
Ethnicity	
Caucasian	4,224
Unknown	941
Asian & Pacific Islander	155
African & African American	98
Hispanic or Latino	6

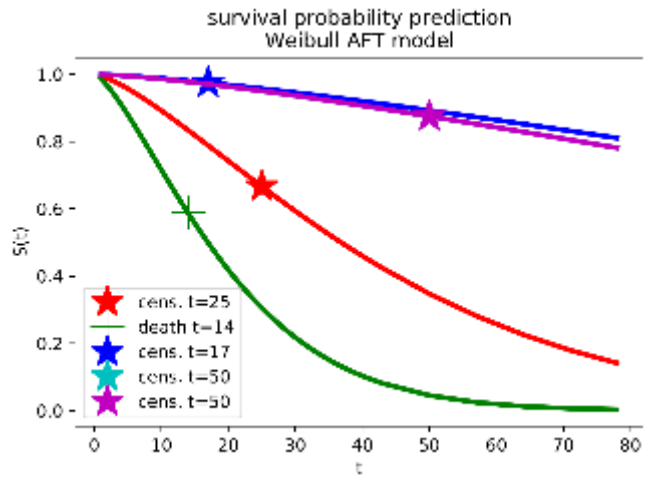
AI

Implementation & comparison of different AI models

	Harrell's C	IBS
Cox	0.77	0.14
DeepSurv	0.73	0.13
RSF	0.74	0.13
Weibull AFT	**0.79**	**0.12**
XGB SE	0.76	0.21

AI

Personalized risk predictions



AI

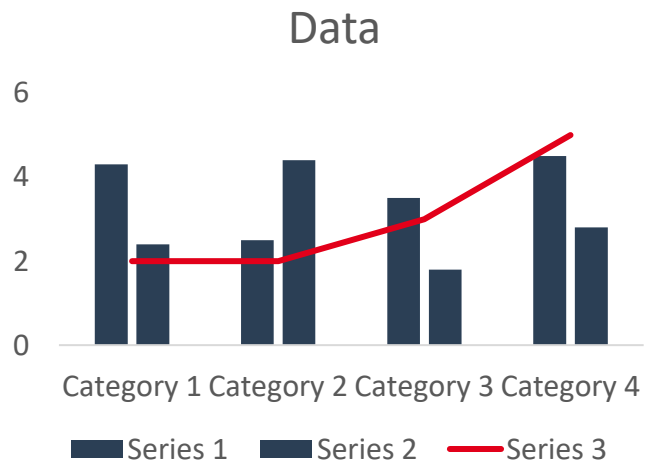
Model explanation & interpretation

rank	domain	feature	importance
1	Demographic	Age	6.5%
2	Symptoms	Asymptomatic	2.9%
3	Vital	SO2 Oxygen Saturation	2.9%
4	Hemato. Lab	Hemoglobin	2.6%
5	Lab	Troponine T	2.6%
6	Symptoms	Muscle Aches	2.4%
7	Lab	Ferrit	2.4%
8	Lab	CRP	2.3%
9	Hemato. Lab	Platelets	2.2%
10	Demographic	Gender	1.9%
...	...	...	...
46	Comorbidities	Dementia	0.8%
		Total	29.4%

Modality	Cumulative Importance	Number of Features
Lab	23.5%	25
Symptoms	20.5%	41
Comorbidities	13.2%	43
Vital	11.2%	11
Hemato. Lab	11.1%	8
Demographic	8.7%	3
Treatments	4.2%	5
CT_Xray	3.1%	12
Urine	2.0%	8
Other	1.6%	2
Smoking	1.0%	2
Total	100%	160

Clinical Studies, Data Categories and Contract Management

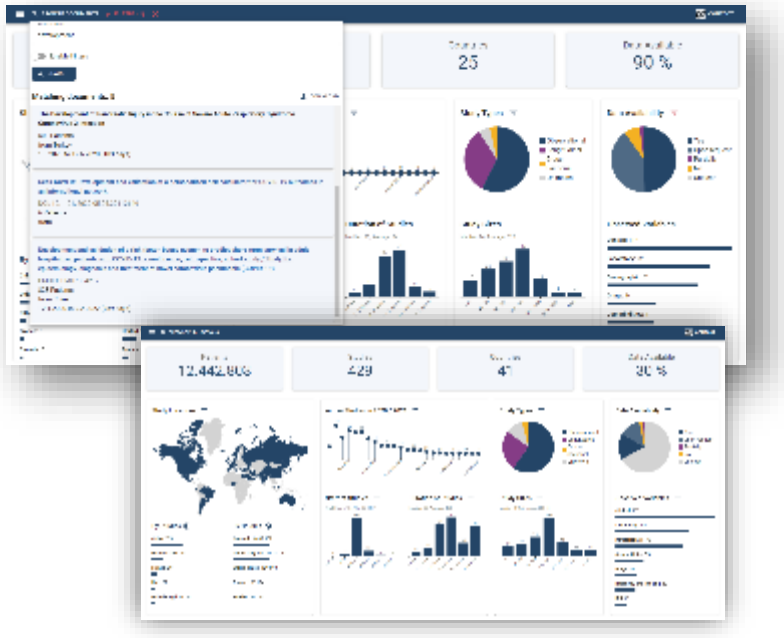
- Data acquisition
  - Providing due diligence catalogue according to defined criteria
  - Evaluation of access, cost, clinical relevance, quality and quantity of the data
  - Support regarding contracts and/or freedom to operate approvals in cooperation with central administration of FhG



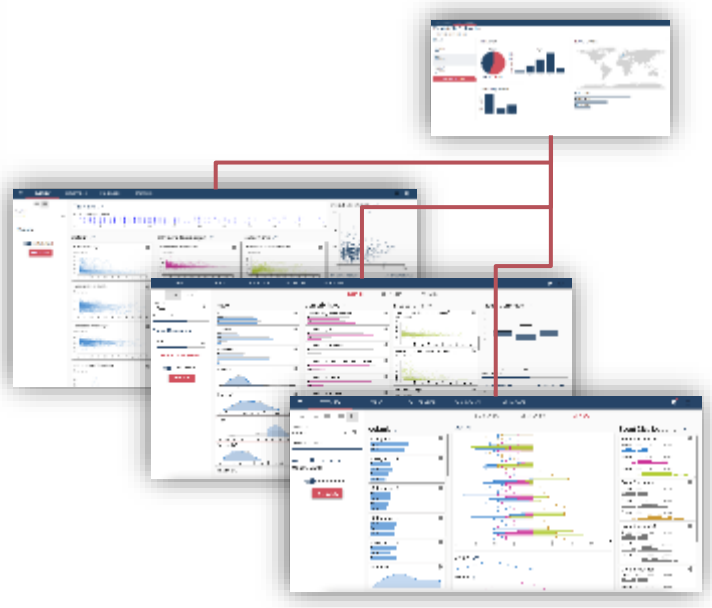
COPERIMO plus

VISUALIZATION AND COMPARATIVE ANALYSIS

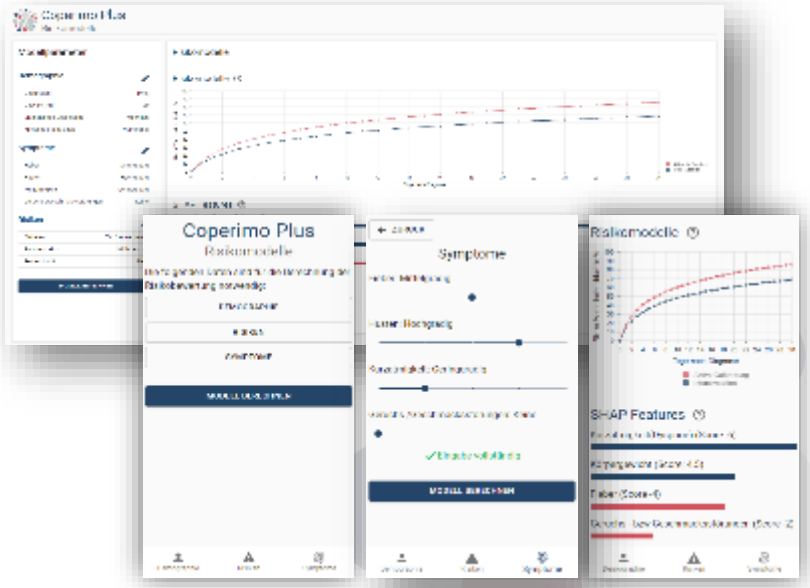
UNSER PLAN
VISUALIZATION & ANALYTICS



Data Acquisition Dashboard



Dataset Exploration

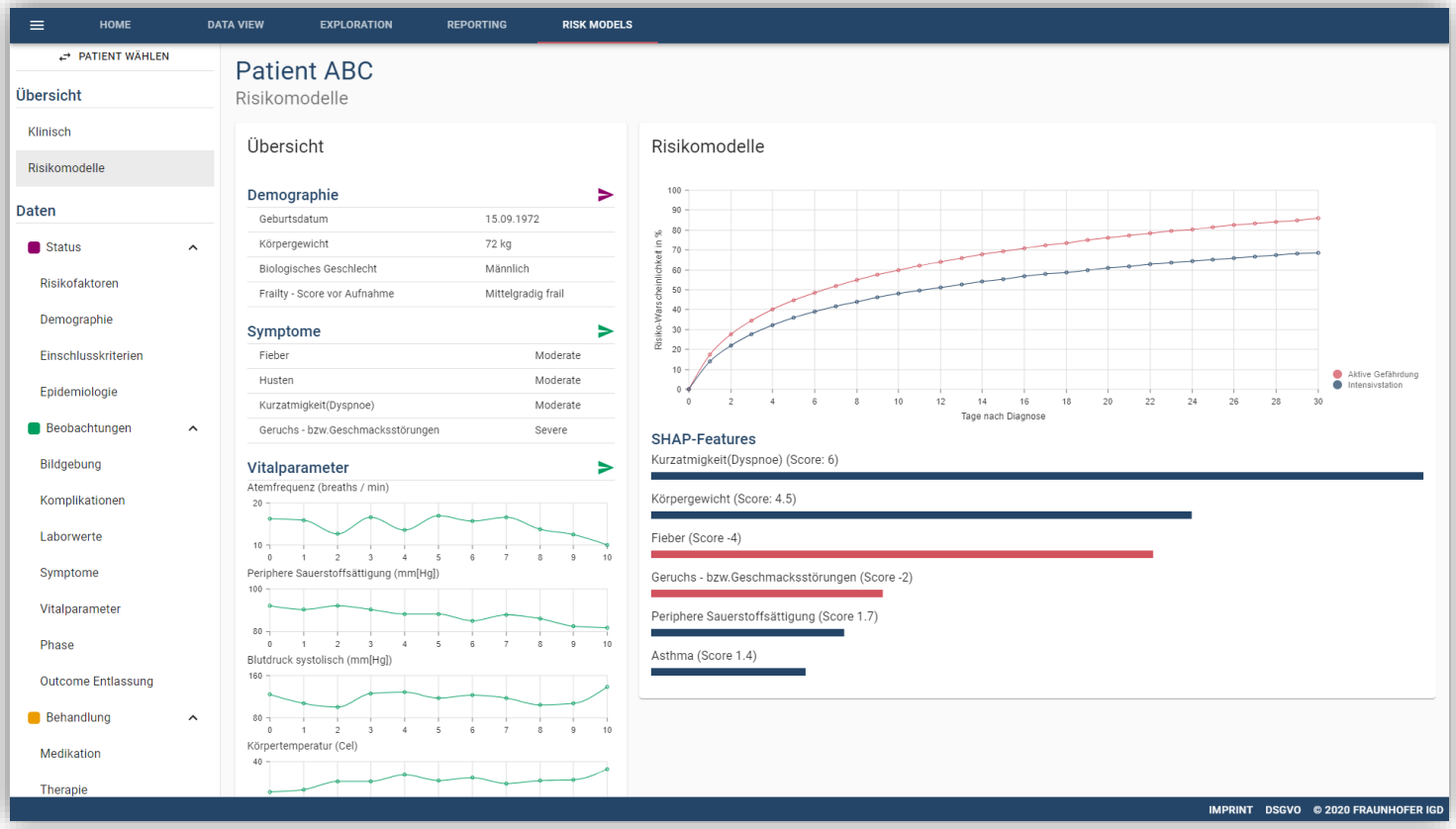


Risk Models UI

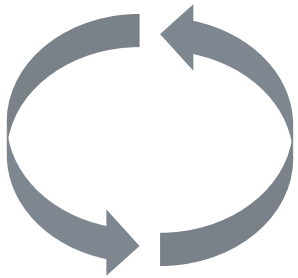
Visual Analytics

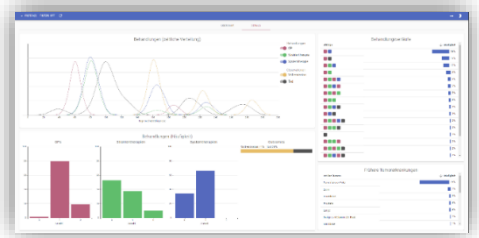
COPERIMO plus

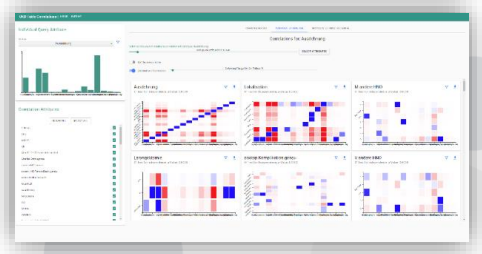
Accessible Interface for personalised Risk Models



Cohort Data Analytics







Fraunhofer SCAI
Fraunhofer IME
Fraunhofer IGD
Fraunhofer IKTS
Fraunhofer MEVIS
Fraunhofer IAIS

SUMMARY

- COPERIMO*plus* will use advanced AI to estimate risk at individual patient level
- COPERIMO*plus* will enable personalized risk management
- The project has established a coherent workflow for data acquisition, data curation and AI modeling
- The project partners with clinical researchers across Germany, Europe and at global scale
- COPERIMO*plus* will do all this “in the open“, there will be no IP claims coming out of this project.
- COPERIMO*plus* will align with related initiatives at global scale; we invite others to collaborate with us
- We aim at covering a wide and diverse spectrum of subjects, including minorities
- Whilst we guarantee the privacy of patient-level data, we will share the resulting models without restrictions
- Our expectation is that COPERIMO*plus* helps to pave the way to future AI strategies aimed at adaptive responses to SARS-CoV-2 or any other, future pandemic