Fraunhofer SCAI

Fraunhofer Institute for Algorithms and Scientific Computing SCAI

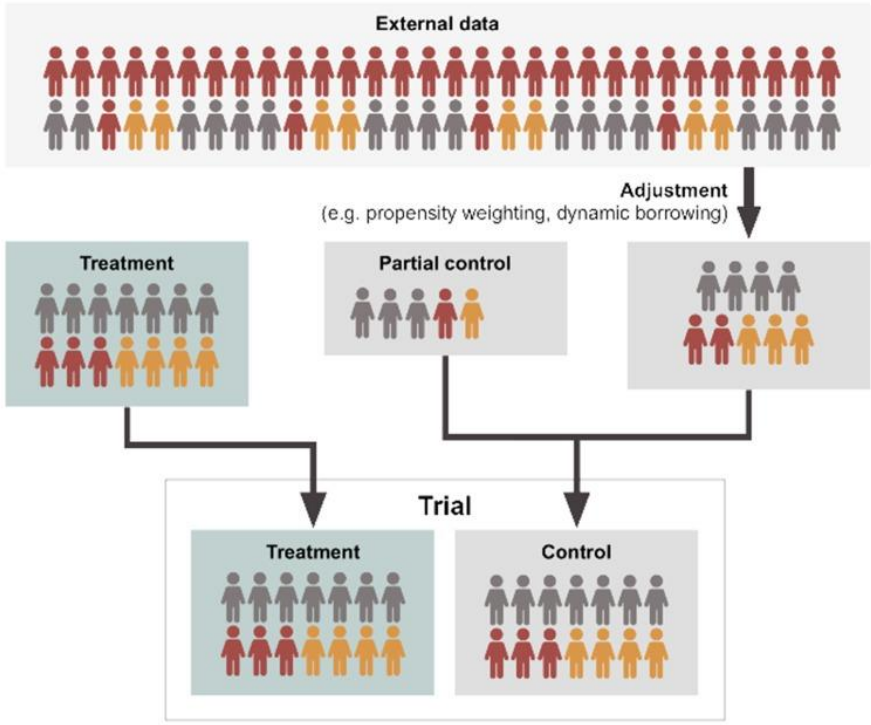
Use of external data to accelerate evidence generation

Digital twins and external controls

Prof. Dr. Holger Fröhlich
Head of AI & Data Science Group, Deputy Head of Department of Bioinformatics
Fraunhofer Institute for Algorithms and Scientific Computing (SCAI)

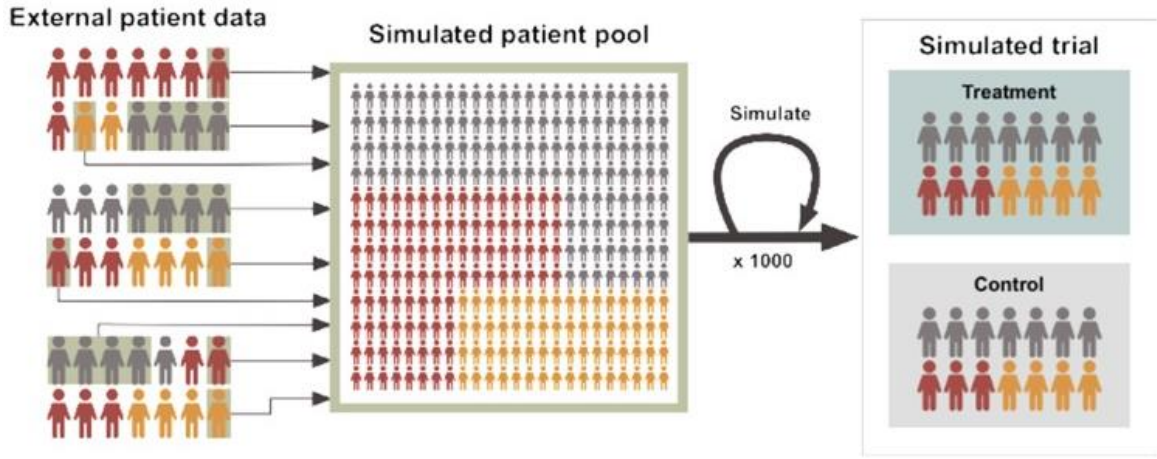
What are external controls and digital twins?

- Real-world data can be used to construct **external controls** for clinical trials



Digital twin simulation

„A **digital twin** is a digital model of an intended or actual real-world physical product, system, or process (a *physical twin*) that serves as the effectively indistinguishable digital counterpart of it for practical purposes, such as simulation, integration, testing, monitoring, and maintenance." *(Wikipedia)*



- Key concept is **Simulation**
- **How to simulate a patient?**

Thorlund et al., Clin Epidemiol, 2020

Fraunhofer SCAI

AI for Simulation of Synthetic Patient Trajectories

Real world cohort

Custom AI algorithms
(VAMBN, MultiNODE)

A mathematical representation of the real world cohort based on an AI model

Synthetic subjects

Quality assurance:
Goodness of fit between synethetic subjects and real world cohort

Data sharing

Facilitate sharing of sensitive data across organizations

Trial design

Simulation

- synthetic control arm
- "what, if" scenarios

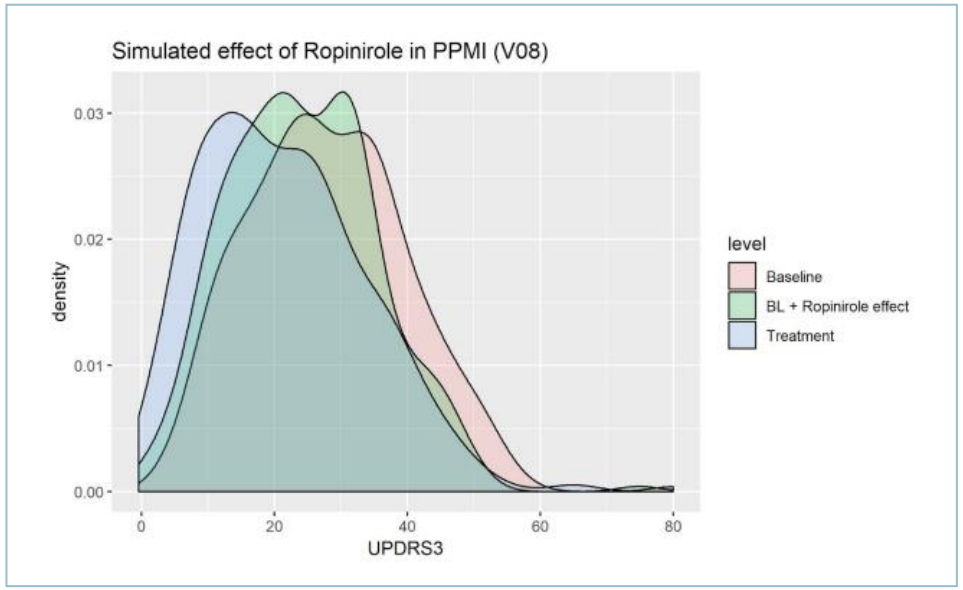
Exploration & Extrapolation

- inclusion / exclusion criteria
- Interpolation / extrapolation
- Statistical power

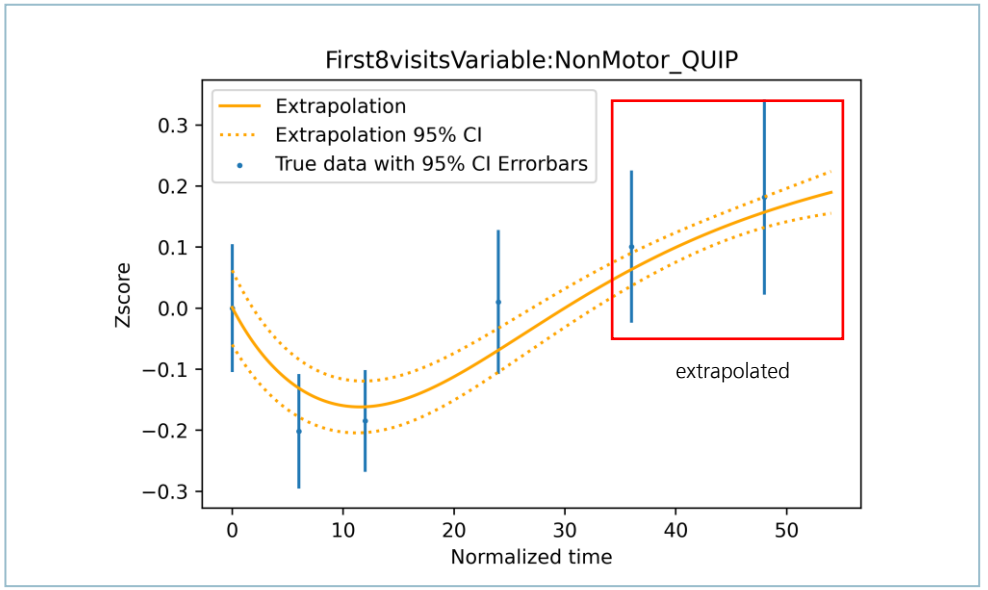
Gootjes-Dreesbach,… & Fröhlich, Frontiers in Big Data: Medicine and Public Health, 2020
Sood,… & Fröhlich, Scientific Reports, 2020
Wendland,… & Fröhlich, npj digital medicine, 2022

Utility of AI Generated Synthetic Data

- Trained AI models allow
  - Simulating counterfactual scenarios, e.g. treatment with a given drug
  - Statistical power calculations
  - Interpolation / extrapolation between visits, also on continuous time scale



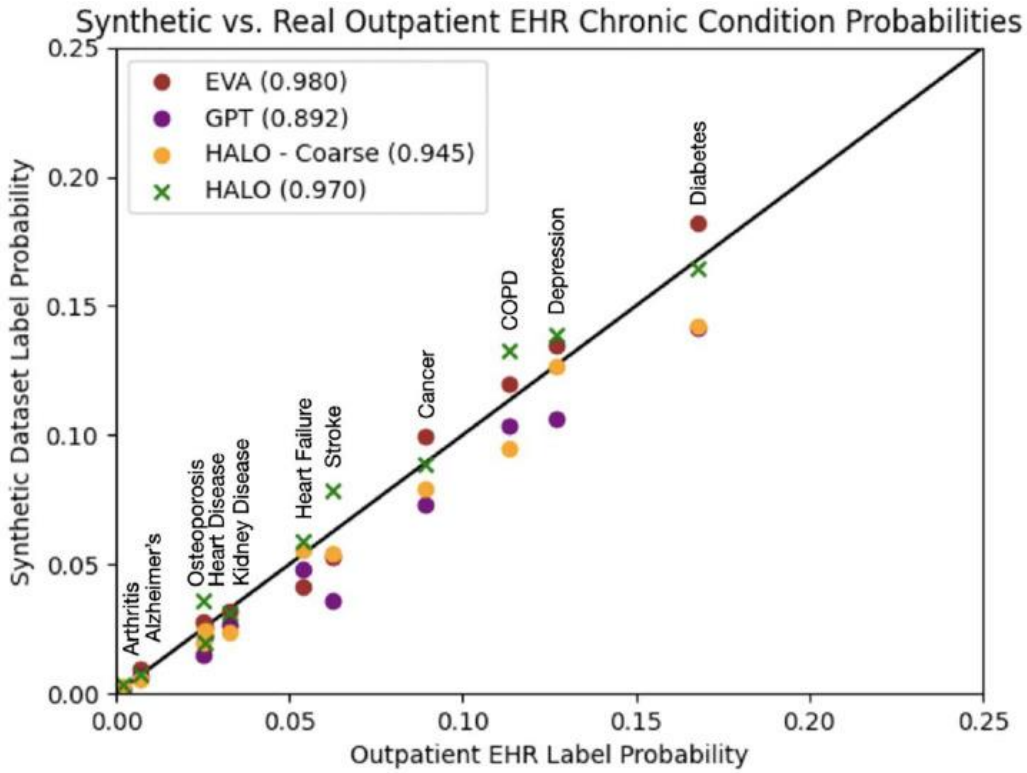
Gootjes-Dreesbach, … & Fröhlich, Frontiers in Big Data: Medicine and Public Health, 2020



Wendland, … & Fröhlich, npj digital medicine, 2022

Fraunhofer SCAI

AI Generated Synthetic EHRs



Synthesize *Extremely High-dimensional* Longitudinal Electronic Health Records via Hierarchical Autoregressive Language Model

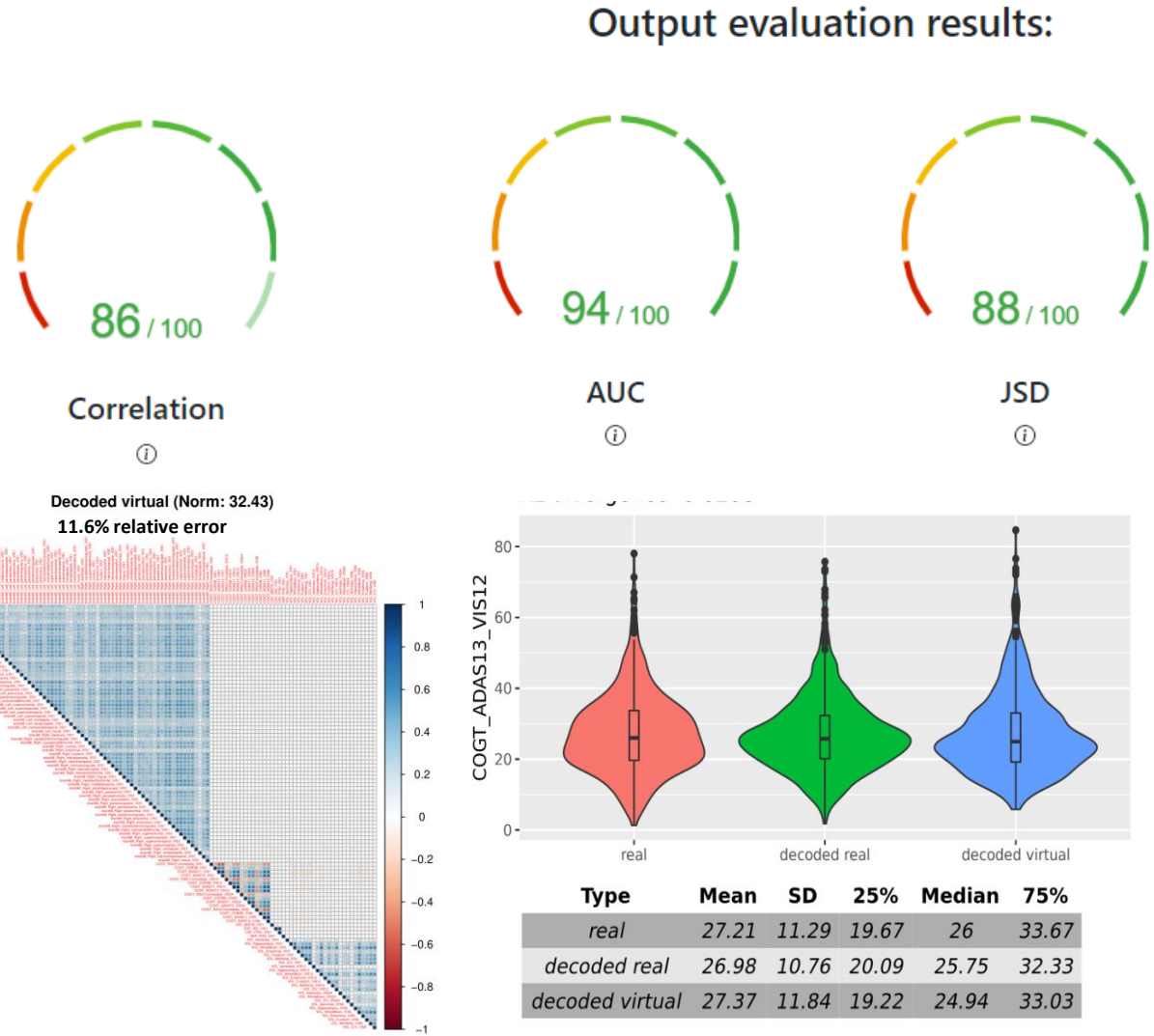
Brandon Theodorou[1], Cao Xiao[2], Jimeng Sun[1*]
University of Illinois Urbana-Champaign.[1]
Relativity Inc.[2]
* To whom correspondence should be addressed: jimeng@illinois.edu

Fraunhofer SCAI

How realistic is synthetic data?

- Syndat: A web-based tool for quality assessment
  - Statistical distributions of individual variables should be similar to real data
  - ML classifier (Random Forest) should misclassify majority of synthetic as real patients
  - Correlation structure of synthetic data should be similar to real data

- Users can upload their own synthetic data

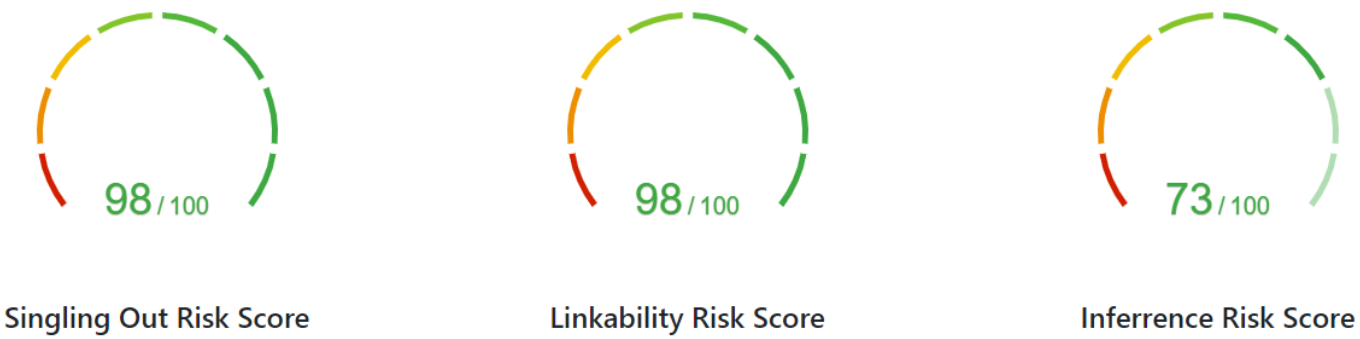


https://syndat.scai.fraunhofer.de/

How privacy preserving is synthetic data?

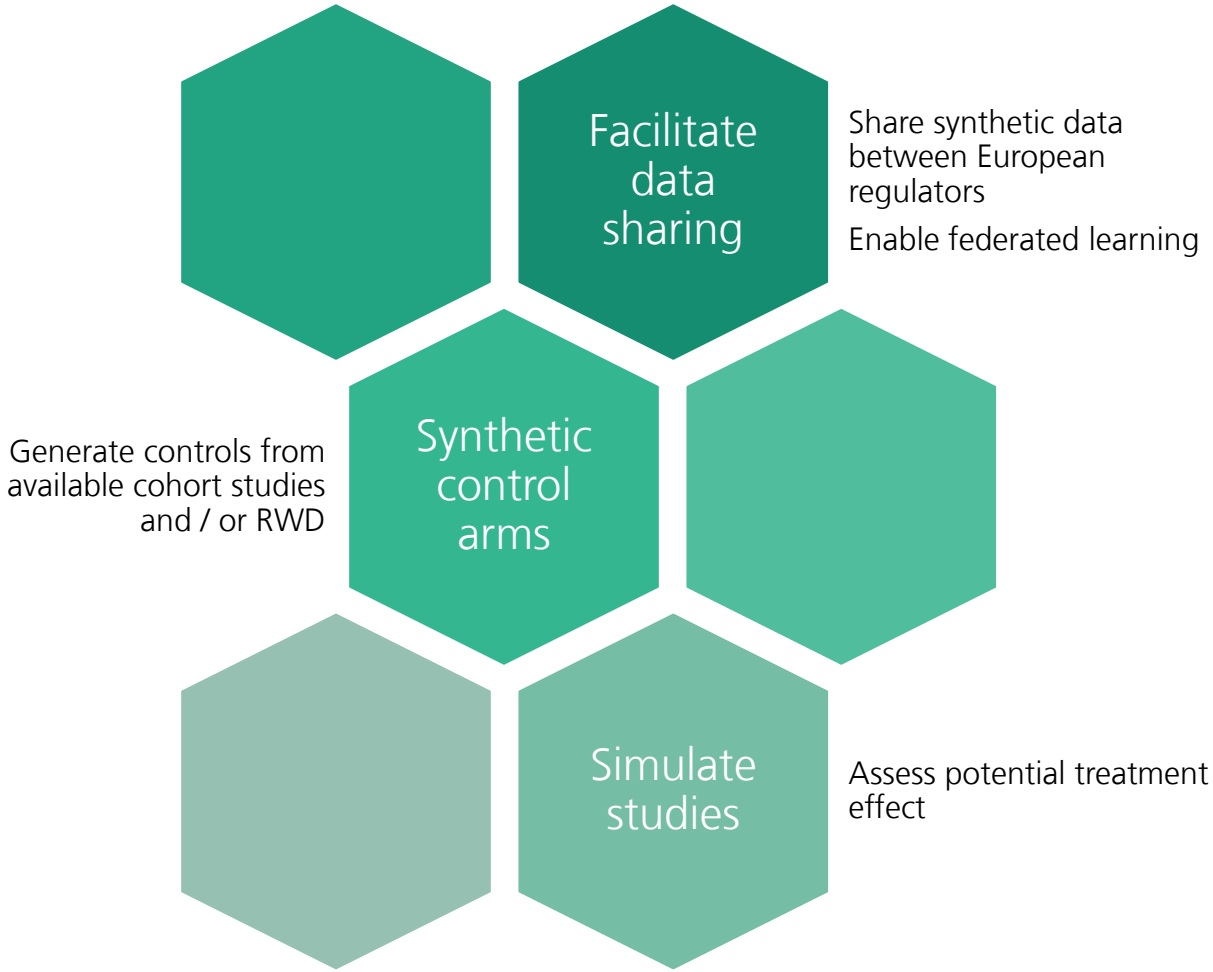
- **Singling out risk**: Can we single out a real individual based on a rare combination of attributes in the synthetic data?
- **Linkability risk**: Can we link together two or more records (either in the same dataset or in different ones) belonging to the same individual or group of individuals?
- **Inference risk**: Can an attacker confidently guess (infer) the value of an unknown attribute of a real data record?

Giomi et al., A Unified Framework for Quantifying Privacy Risk in Synthetic Data, arXiv, 2022



https://syndat.scai.fraunhofer.de/

Summary: Where is the potential benefit for regulators?



Fraunhofer SCAI

Fraunhofer Institute for Algorithms and Scientific Computing SCAI

Contact

Prof. Dr. Holger Fröhlich
Head of AI & Data Science Group
Deputy Head of Department of Bioinformatics
Fraunhofer Institute for Algorithms and Scientific Computing (SCAI)

Tel. +49 2241 14-4206
holger.froehlich@scai.fraunhofer.de