ESOF2020
EUROSCIENCE OPEN FORUM
TRIESTE

Why we need Virtual Cohorts

Sometimes, data are like Ghost Ships ….

Euro Science

Fondazione Internazionale Trieste

Patient-Level Data: Essential for Translational Biomedical Research

Patient-level data are an essential cornerstone of any translational biomedical research

- Studies with patients are needed to describe diseases and disease progression
- Studies with patients are formally required to test the safety and efficacy of new drugs and other interventions
- Studies with patients give us insights on what “biomarkers” we can use to describe the disease state a patient is currently in
- Studies with patients are the cornerstones of any translational biomedical research

Sometimes, Study Data are like Islands

Common features of these Islands?

Some Clinical Data Sets are like Ghost Ships ….

They appear in grant applications again and again, but you will never be able to work with them …..

… other Data Sets are perfectly siloed ….

- Variable catalogue ?
- Summary statistics ?
- Interoperability ?
- Shared metadata ?
- MERGE dataset ?
- Pre-Processed ?
- Curated ?
- Imputed ?
- ….

Getting Access to a Study means …. NOTHING

Even when we get access to studies, this does not mean that we can work with the data …

- Getting access to study data does not mean, that you can use them
- Cannot share data within the consortium (everybody pre-processes again and again)
- Sometimes, important variables (visits in time series) are not in the package
- We have to chase special data owners (e.g. missing parts of the data set shared)
- Data may be incomplete, of limited quality, or simply not comparable to other studies ….

… and when you solved all the technical problems ….

- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- Data protection
- .....

GDPR, Declaration of Human Rights (1948)!, et al. ……

GDPR

The General Data Protection Regulation

….
ROADBLOCKS for translational biomedical research

No way out? ....

Our way to respect data privacy and to enable translational research at the same time:

Synthetic Data

When we were children ….

.. We were told: “**you must not play with money, not with food and not with patient data** …”

- DOCH !

The German word “Doch!” has a meaning close to “sure!” or “Yes, of course!” …. but it is much stronger and it disproves a previous statement.

We love to use that term in disputes when somebody says “you can´t do that” and we simply say: “doch”.

A bit like Obamas “yes, we can …”

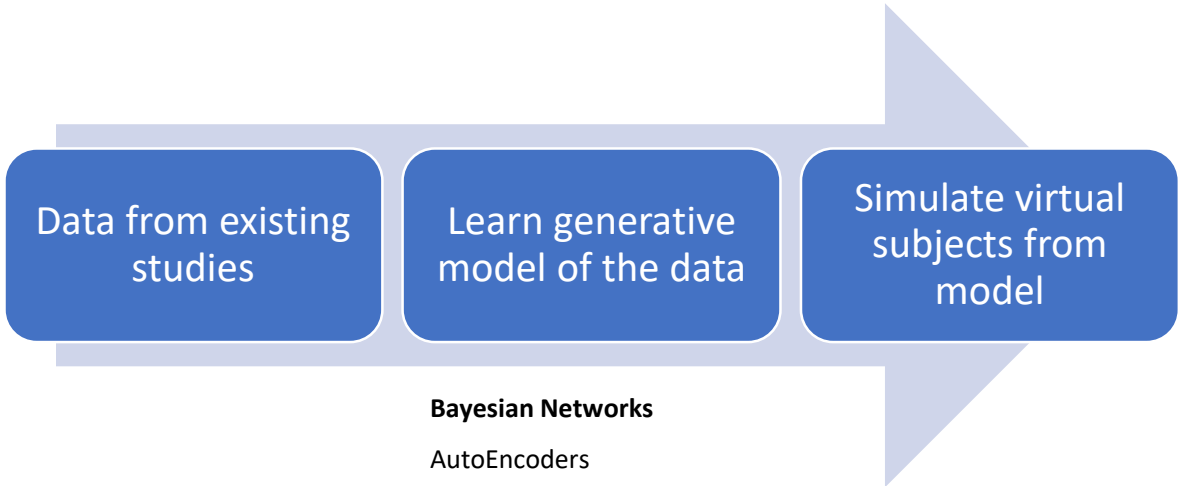
Actually, we need to play with data ….. all the time ….

- **Data scientists** need to play with data for **methods development**
- Scientists doing **clinical trial simulation** needs data to play with
- **Mechanism-based stratification** needs data to play with
- **Deep learning (AI)** needs lots of data to play with
- **Students** need data that can be freely shared to improve their data science skills

Generating Toy Data to play with: Virtual Cohorts

- **Synthetic** data sets
- **Instructed** by reality
- **No patient data privacy rights**
- Very **close to reality**
- **Allow to "publish"** clinical data
- **Allows to share** clinical data
- Allow for **global meta-cohorts**
- Can integrate **a priori knowledge**
- Can be used to **ask "unethical" questions**
- Can be used to **mix and merge**

Learning Synthetic Data from Real World Studies



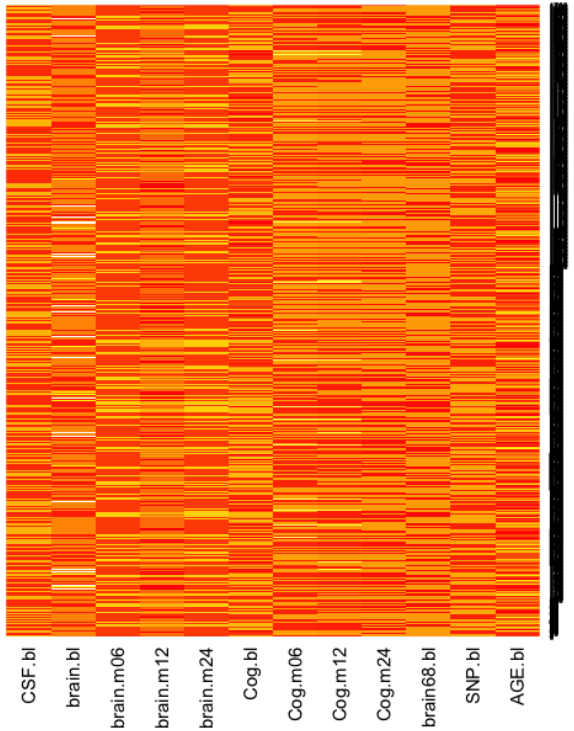
Khanna, Shashank, et al. "Using Multi-Scale Genetic, Neuroimaging and Clinical Data for Predicting Alzheimer's Disease and Reconstruction of Relevant Biological Mechanisms." *Scientific reports* 8.1 (2018): 11173.

Thousands of Virtual Alzheimer Patients …

Virtual Patients generated in the disease area of Alzheimer´s Disease are shown.

In the heat map shown, the features of 689 real-world Alzheimer patients and the features of 1000 virtual patients have been clustered.

Try to tell them from each other ….



Thank you for your attention …

Acknowledgement:

Prof. Holger Fröhlich

Colin Birkenbihl

Meemansa Sood

Shashank Khanna

Prof. Nikolaus Forgó

Dr. Marc Stauch

ESOF2020
EUROSCIENCE OPEN FORUM
TRIESTE

EuroScience

Fondazione Internazionale Trieste