Meet the panel

- **Elisabetta Vaudano**
  Principal Scientific Manager, Innovative Medicines Initiative (IMI)
- **Graciela Muniz Terrera**
  Senior Lecturer in Biostatistics and Epidemiology at the University of Edinburgh, participant in IMI's EPAD project
- **Martin Hofmann-Apitius**
  Head of Business Area Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing SCAI, coordinator of IMI's AETIONOMY project
- **Holger Fröhlich**
  Head of AI & Data Science Group; Deputy Head of Department of Bioinformatics, Fraunhofer; Professor at University of Bonn; previously, Head of Data Science Enablement, UCB

Cohorts – a beginner's guide

What is a cohort?

➢ A group of people selected on the basis of certain characteristics, e.g. age / gender / occupation / exposure to a specific risk factor like cigarette smoke

Why do researchers need cohorts?

➢ To study how common diseases are, their causes, what happens to patients over time…valuable also to identify individuals to enroll in clinical trials of novel treatment.

How do cohorts studies work?

➢ By comparing two or more groups over time and studying them for signs of disease. E.g. if smoking is suspected to cause a disease, you could compare a cohort of smokers to a cohort of non-smokers.

A bit more jargon-busting…

Cohort studies can be….

➤ Prospective

A group of similar individuals who differ with respect to certain factors under study is followed over time to determine how these factors affect rates of a certain outcome. For example, one might follow a cohort of middle-aged truck drivers who vary in terms of smoking habits, to test the hypothesis that the 20-year incidence rate of lung cancer will be highest among heavy smokers, followed by moderate smokers, and then non-smokers.

➤ Retrospective

Groups of individuals who are alike in many ways but differ by a certain characteristic (for example, female nurses who smoke and ones who do not smoke) is followed in terms of a particular outcome (such as lung cancer). Data on the relevant events for each individual are collected from existing records and can immediately be analyzed to determine the relative risk of the cohort compared to the control group. This is fundamentally the same methodology as for a prospective cohort study, except that the retrospective study is performed post-hoc, looking back.

Questions we'll ask (and hopefully answer) in this session...

➢ What's the problem with standard cohorts with real people?

➢ What is a virtual cohort?

➢ What are the advantages of a virtual vs a standard cohort

➢ How do you create a virtual cohort?

➢ How are scientists using virtual cohorts today?

➢ What does the future hold for virtual cohorts?