Risk assessment of cardiovascular diseases for all citizens

Cardiovascular diseases are the most common cause of death in the world. More than a third of deaths in Finland are caused by cardiovascular diseases. The current objective is to create an assessment, based on health data, of each person's risk of illness before they consult a doctor.

Andrea Ganna, EMBL Group Leader from Finnish Institute of Molecular Medicine and instructor from Harvard Medical School, wants to establish a nationwide, personalised risk assessment as foundation for planning public health interventions. The assessment is based on the health, demographic and genetic information of the citizens. The assessment, which uses artificial intelligence, improves the allocation of preventive treatments with a lower cost than today.

"Nordic countries and specifically Finland have a unique opportunity and setting, since they have been collecting health and demographic data for years. But the way they have used this data in the past is somewhat outdated. Only very specific correlations and associations in the data have been looked at. However, new methods, such as AI, are emerging, which now allow us to push for a much bigger and ambitious vision."

Andrea Ganna and his group is developing artificial intelligence (AI) approaches to model health trajectories.

"You have a certain health trajectory and have taken certain medication. We ask if there are other people who have followed a similar path. There may be thousands out there. We leverage those people and ask- what happened to those? Let's take that experience and bring it back to you to help you to reduce your disease risk. We can use all this data in a more comprehensive way to help public health and give more information to patients and doctors for decision-making."

Risk assessment before visiting a doctor

Andrea Ganna is interested in epidemiology, genetics and statistics. He has been focusing on leveraging large-scale epidemiological data sets to identify socio-demographic, metabolic and genetic markers of common, complex diseases. In Boston he worked with large-scale exome and genome sequencing data.

According to Ganna , cardiovascular diseases are ideally suited for analysis by artificial intelligence.

"Accurate identification of individuals at high risk is one of the cornerstones of primary prevention of cardiometabolic diseases," he says. "However, at the moment, risk factor assessment for cardiometabolic diseases requires patients to go to the doctor for lipid measurement."

Lipid is the umbrella term used for all fatty acids or their derivatives that circulate in the blood. The body stores fats from food for later use. A diet rich in fats will cause them to attach to the walls of the arteries, leading to cardiovascular and arterial diseases. Lipid measurement is effective, but some members of the population are unaware that they belong to a risk group.

Ganna wants to revolutionize primary prevention by providing risk assessment before an individual even steps into the doctor's office.

"Some simply don't go to the doctor's and so lots of people are missed. However, since all the data on medication and diagnoses has already collected, we can identify high-

Finnish school children. Probably the most important population is younger individuals who do not see a doctor very often. Genetics is particularly valuable because it can capture disease risk at an earlier age than other risk factors. The FinnGen study will utilise samples collected by a nationwide network of Finnish biobanks. The study is based on combining genome information with digital health care data from national health registries.

risk patients without them going to doctor. We can make a risk map of cardiovascular diseases of the whole country by including every individual."

Calculating the risk is done by modelling longitudinal histories of diseases and medications with the gene data, family and demographic data.

"We are trying to understand how genetics interact with data regarding medications, diagnoses, demographics, and familial risk. This can provide an unprecedented holistic view of an individual's health status."

Ganna gives an example.

"When you break your leg, you go to the doctor. However, today the doctor is just looking at the leg, although during the same visit other information could also be obtained. We can inform the doctor about other risks the patient has based on the collected data. We can precompute the other risks of the patient, for example if this patient has also high risk for cardiovascular diseases. Thus, during the visit, the doctor can also give advice or refer the patient to a specialist."

Genetics is useful

Ganna chose to come to Finland because of the large genetic project, FinnGen. The FinnGen project will record the genomes of half a million Finns. The project, launched in August 2017, utilises samples collected by all Finnish biobanks. The data from genomes is combined with the information in national health care registers. FinnGen is one of the very first personalised medicine projects of this scale and the public-private collaborative nature of the project is exceptional.

"Finland also has a favourable legislation, giving access to nationwide population data. For me, this is a unique setting", says Ganna.

Ganna and his research group integrate registry-based information with genetic information from large biobank-based studies (e.g. FinnGen) to help identify groups of individuals that can most benefit from existing pharmacological interventions.

"Probably the most important population is younger individuals who do not see a doctor very often. Current risk factors do not work well in this group. Genetics is particularly valuable because it can capture disease risk at an earlier age than other risk factors," says Ganna.

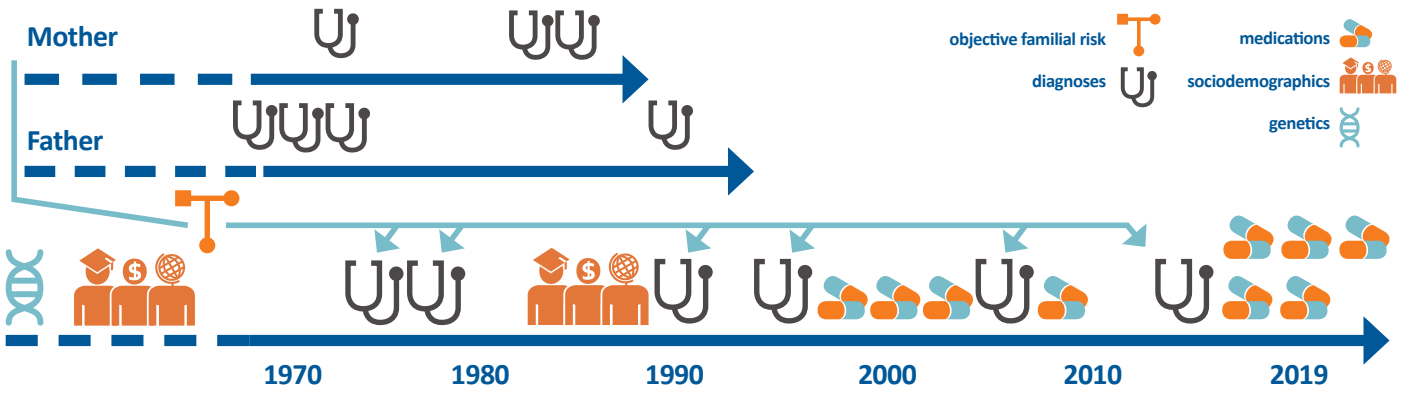
"The first step is to understand how people perceive this information. We have to ensure that doctors use the data in a right way and what can be done with it. Cardiovascular diseases are good example, since their treatment is preventive."

Deep and machine learning

Ganna aims to integrate national and regional registries with deep and machine learning.

"Traditional methods have an advantage since they are relative simple and easy to interpret, but they simple do not scale. In the past 20 years, more than 500 million medical diagnoses have been made of Finns. We are talking about huge data sets. Every year there are millions and millions of new medication purchases and diagnoses. To scale and to leverage this massive data, deep learning methods are needed."

Artificial neural networks are efficient machine learning algorithms, which can be



Health trajectories might allow to identify individuals at high risk for cardiometabolic diseases when using information available nation-wide.

used in pattern recognition. Recurrent neural networks can use their internal memory to process sequences of inputs. This makes them applicable to tasks such as unsegmented recognition. Ganna wants to expand long short-term memory recurrent neural networks to data.

"You can imagine the sequence of health events that we are trying to model as "text" in which each word is a different disease, medication, sociodemographic event etc. across the lifetime. These are naturally suited to model the sequential happening of events, for example they are used to predict the next most likely word in a text message."

Deep learning methods need large supercomputing infrastructure.

"CSC has created a secure environment for this computation. Without a secure supercomputing environment, we could not carry out this project. To be successful, we need, on one the hand, research and development, and, on the other hand, a powerful computing environment."

Personal data is protected

Patient data is important for research, but personal data is also protected. For example, VEIL.AI application created by FIMM anonymises patient data better and faster than traditional methods, and retains information more effectively. If necessary, the application can also produce synthetic, fully anonymous statistical data, which cannot be traced back to any individual.

"We need to guarantee individuals' privacy but, at the same time, we need to integrate a lot of personal data to really leverage the power of artificial intelligence/deep learning approaches to better target public health interventions. Generating synthetic health trajectories will help to respect privacy and, at the same time, to combine a lot of personal information within Finland, but also across Nordic countries."

"My hope is that personal data that is routinely collected in healthcare can help and benefit everyone. My hope is that that this information can help doctors to make better decisions, but also help patients in motivating life style changes. Thus everyone is helping everyone."

Ari Turunen

MORE INFORMATION:

Institute for Molecular Medicine Finland (FIMM)
The mission of the Institute is to advance new fundamental understanding of the molecular, cellular and etiological basis of human diseases. This understanding will lead to improved means of diagnostics and the treatment and prevention of common health problems. Finnish clinical and epidemiological study materials will be used in the research.
www.fimm.fi

CSC – IT Center for Science
is a non-profit, state-owned company administered by the Ministry of Education and Culture. CSC maintains and develops the state-owned, centralised IT infrastructure.
http://www.csc.fi
https://research.csc.fi/cloud-computing

ELIXIR
builds infrastructure in support of the biological sector. It brings together the leading organisations of 21 European countries and the EMBL European Molecular Biology Laboratory to form a common infrastructure for biological information. CSC – IT Center for Science is the Finnish centre within this infrastructure.
http://www.elixir-finland.org
http://www.elixir-europe.org

ELIXIR FINLAND
Tel. +358 9 457 2821s ● e-mail: servicedesk@csc.fi
www.elixir-europe.org/about-us/who-we-are/nodes/finland

www.elixir-finland.org

ELIXIR HEAD OFFICE
EMBL-European Bioinformatics Institute
www.elixir-europe.org