P1924

Population-wide disease trajectories in ischaemic heart disease

A.D.H. Haue[1], I.F.R. Jorgensen[2], A.H.C. Christensen[3], C.S. Simon[2], M.L.H. Haakonsen[2], F.R. Russo[2], C.L.R. Rodriguez[2], A.A.O. Orozco[2], T.S. Siggard[2], K.B. Banasik[2], P.E.W. Weeke[1], D.W. Westergaard[2], P.L.M. Moseley[2], H.B. Bundgaard[1], S.B. Brunak[2]

[1]*Rigshospitalet - Copenhagen University Hospital, Copenhagen, Denmark;* [2]*University of Copenhagen, Novo Nordisk Foundation Center for Protein Research, Copenhagen, Denmark;* [3]*Herlev Hospital - Copenhagen University Hospital, Hjertemedicinsk Afdeling, Copenhagen, Denmark*

Background: Defining cardiovascular disease (CVD) phenotypes from large, longitudinal electronic health records (EHRs) through analysis of patient similarities and dissimilarities is a strategy with new possibilities to practice precision medicine. We carried out a pilot data screen to quantify and characterise selected CVD phenotypes from EHRs comprising medical histories of 6,986,632 individuals spanning 21 years.

Purpose: The overall aim is to define temporal CVD phenotypes by data-driven characterisation employing bioinformatics approaches. We have defined temporal CVD phenotypes, by data-driven characterisation identifying statistically significant temporal disease trajectories that allow for future integration of lab test results, drug prescriptions and genomic data.

Methods: Data was assessed by computing temporal disease trajectories made from selecting certain indicator diagnoses. Inclusion criterion was admittance to a Danish hospital during 1995–2016. All data points were indexed by a unique key for each individual in a registry based national infrastructure that is stable over a life time. Encryption was performed pre-analyses to acquire research prone patient IDs (PIDs). Diagnostic codes were annotated from EHRs according to the WHO ICD-10, tests according to the Nomenclature for Properties and Units (NPU), procedures according to the Danish Health Care Classification (SKS) and drug prescriptions according to the Anatomical Therapeutic Chemical Classification System (ATC).

Results: The largest subsets in the case population were cardiac arrhythmias (I44–49) and chronic ischaemic heart disease (I20-I25) counting 582,180 and 579,619 patients. respectively. Mapping of temporal disease trajectories leading to cardiac arrest (I46) one of four major CVD complications, demonstrated that the majority of cases matching chronic ischaemic heart disease (I25) who present with cardiac arrest (I46) do not have any intermediate diagnosis. This kind of trajectory illustrates the deep phenotypic spectrum of the most common type of I25 patients. Conversely, no direct disease trajectories were observed between patients diagnosed with cardiac arrest (I46) following myocardial infarction (I21) or heart failure (I50) (see figure). Overall, the population-based reference phenomes of the selected CVD diagnoses from the dataset used was verified using detailed EHR from a subset amounting to approximately 2.6 million patients.

Conclusion: Mining of data from patients with chronic ischaemic heart disease by computing distinct disease trajectories leading to cardiac arrest provide a promising framework for establishing computational phenotypes. The multimorbidity trajectory approach allows us to define the longitudinal phenotype in the big data set. We argue that inclusion of additional data types including large-scale genomic analyses for sub-group stratification will elucidate disease mechanisms facilitating implementation of precision medicine.

Ischaemic heart disease trajectories