

WISER – A Holistic Analysis Of Large Scale Electronics Health Records

(No. T4-2113)

Principal investigator

Amos Tanay

Faculty of Mathematics and Computer Science

Department of Computer Science and Applied Mathematics

Overview

In the current era of big-data medicine, the Electronic Health Records (EHR) analysis is limited due to its disease-oriented approach. We developed WISER (Weizmann Institute of Science Electronic-health-record Research) platform - a suite of algorithms, models, software and workflows that enables the genomics and post-genomics data revolution in healthcare analytics. It brings together an established scientific leadership in genomics and post genomics data analytics technologies and extensive experience with analysis and modelling of traditional real-world EHRs. WISER creates a powerful data environment that facilitate today's medical analytics needs, while building the infrastructure for tomorrow's diagnostic and therapeutic approaches

The Need

Current EHR analysis systems have many drawbacks: The raw data is treated ad-hoc, the infrastructure is "endpoint-driven" although the problem is holistic, integrative and bottom-up and the systems are not "EHR native" – they are based on technologies that were developed in other domains and are slow, indirect and disregard medical, epidemiological and statistics semantics. Furthermore, these solutions are limited with respect to the breadth, depth and multilayer nature of the data needed to support holistic and multi-variate models: the current Machine Learning (ML) point of view is dichotomic and naive – health vs. disease instead of continuous multi-variate risk modelling, the population health models are not compatible with modern quantitative genetics, the poor modelling of pre-disease stages is inhibiting new data-acquisition opportunities and post-genomic opportunities (transcriptomics, metabolomics, epigenomics) are missing.

The Solution

The group of Prof. Tanay developed WISER - Weizmann Institute of Science Electronic-health-record Research platform that optimize modelling, predicting and assessing management of disease and disease outcomes in complex EHRs.

Technology Essence

WISER holistic platform consists of software, workflows and powerful algorithms facilitating accelerated discovery in large, complex, sparse and heterogeneous EHR environment. Prof. Tanay and his team developed a technology that combine state of the art data analysis and machine learning with the ability to continuously formalize and improve system-wide knowledge providing superior performance and scalability. The WISER platform includes highly optimized proprietary database engine for patient trajectory, data-hungry low-level normalization and imputation tools that use semantics as early as possible, automated massive screening for data imperfection, errors, and irregularities, multi-variate quantitative modelling of common lab markers, assisted rapid derivation of

disease risk scores, integrative management of EHR knowledge with global cohort library, adaptive controls, comparative analysis using ML software and EHR algorithms, and interactive UX.

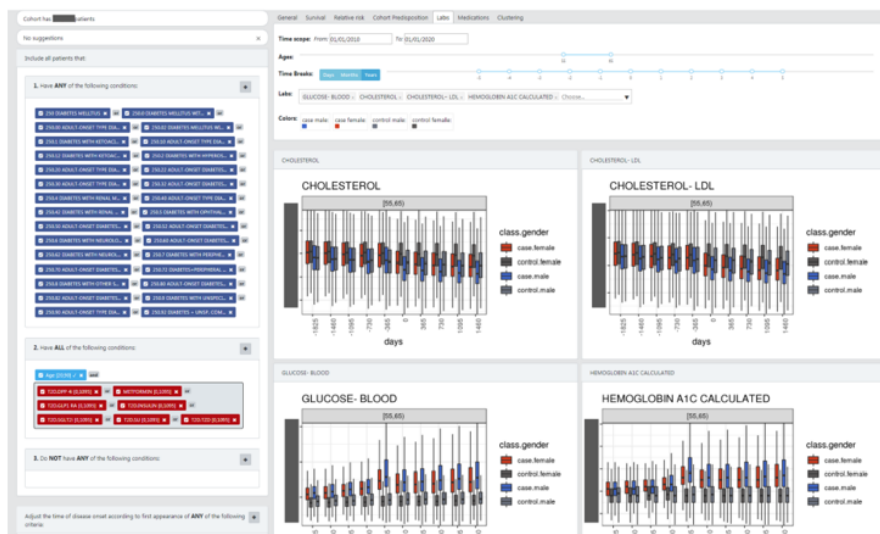


Figure 1: WISER example of lab distributions over time stratified by age / gender and sampled control.

Applications and Advantages

Applications

- Prediction/risk stratification of diseases including follow up analysis and optimization.
- Lab readout standardization, normalization and comparison using historical lab readouts, lab instruments' information and patient distributions (health state, age) from different patient communities.
- Novel personalized metrics for lab readout in healthy individuals enabling to report on patient risks for developing abnormal lab results in the future.
- A foundation for the development and deployment of new precision diagnostics and management tools, from molecular profiling to continuous sensors.
- Smart sampling of patients for novel molecular assays.

Advantages

- Enables feature engineering, normalization, data harmonization, classification and prediction via tailored machine learning algorithms.
- Facilitates the organization of the system know-how such that cohort libraries, normalized data tracks, and computed quantitative measure can be developed to maximize the impact of available information and avoid the currently common cycles of "re-discovery" of biases and confounders in the data.
- Provides advanced workflow and analysis capabilities for non-programmers such as analysts, physicians and decision makers.
- Automated, real-time analysis of medication side effects.
- Supports 10s of thousands of clinical variables over complex cohorts.
- Fully compatible with quantitative genetics and with epigenetic integration into genetic stratification and marker derivation.

Development Status

Prof. Tanay and his group exemplified a multivariate precision modelling of 2.1B billion lab measurements of 92 different lab tests from 2.8 million adults (age range 20 - 90) over a span of 18 years in the Clalit Healthcare services system. Their analysis identified 131 chronic conditions and 5223 drug-test pairs that affected tests distributions, retaining 545 million tests to create a virtual survey of lab tests in healthy individuals. The team also developed personalized models that can predict if currently healthy individuals will present with abnormal test levels within 2-3 years. This can provide advanced risk stratification for overall survival or specific subsequent diseases. It shows that multivariate modelling of lab tests can be used to infer patients' deterioration potential while their lab values are within the currently assumed normal ranges.

Market Opportunity

WISER can greatly enhance EHR analytics. It can be used in:

- Healthcare systems, in particular those caring for large populations going through complex chronic disease trajectories.
- Pharma, providing best-in-class analytics that can integrated multiple EHRs in diverse analysis projects and provide native integration of genetics and post-genomic analysis.
- Medical technology developers, enabling powerful, trustworthy and rapid EHR integration with new data layers created by developers of diagnostic and therapeutic technologies.