

Exploring electronic phenotyping for clinical practice

Authors: Makhlysheva A., Nordsletta A.T., Tayefi M., Chomutare, T.

This fact sheet provides an overview of what electronic phenotyping is, how electronic phenotyping algorithms are developed, and what they can be used for.



An individual's observable characteristics such as height, eye color, blood type and diseases, in biology are called phenotype.

The person's phenotype results from the interaction of one's genotype, environment and social influence.

Photo: Colourbox.

What is electronic phenotyping?

Electronic phenotyping is the characterization of an individual's condition based on data.

In the electronic health record (EHR) system, clinicians register unstructured and structured data, and the two are complimentary. Unstructured data are represented by patient's signs and symptoms, radiology and pathology reports, discharge summaries and family history. International Classification of Diseases (ICD) codes, laboratory results and medications are examples of structured data.

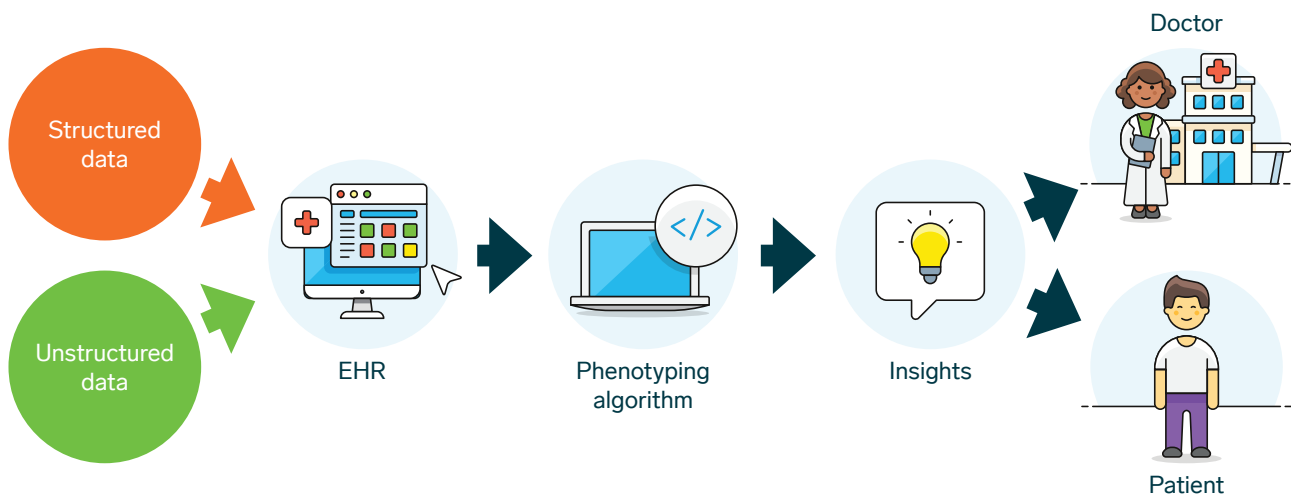
Electronic phenotyping uses EHR and any other machine-readable data to characterize a patient's condition. These data can include genomic data, diagnostic images, structured and unstructured

clinical data, patient-generated data, and environmental data among others. The more data sources we have, the more complete electronic phenotyping can be.

Electronic phenotyping can be used to

- identify people with specific conditions
- public health and safety surveillance
- administrative purposes
- clinical research studies
- precision medicine (PatientsLikeMe)

EHR-driven phenotyping algorithms transform raw EHR data into meaningful features to classify or predict individuals' phenotypes. This information gives insights on whether the individual has a specific medical condition or is at risk for developing one. Combining phenotype and genotype data can characterize patients more precisely.



Methods for electronic phenotyping

Machine learning (ML) and natural language processing (NLP) are used for electronic phenotyping.

ML commonly refers to a collection of techniques for extracting knowledge from large data sets, and use of these techniques for solving classification, prediction and estimation problems. In phenotyping, ML is used to enable correct predictions for target diagnoses, based on observed features from corresponding samples, therefore, reducing efforts needed from humans.

NLP is a combination of ML and linguistics and is used to extract features from clinical notes. This may increase the ability of the phenotyping algorithm to correctly recognize patients having the diagnosis.

References

1. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *BMJ*. 2015 Apr 24;350:h1885. Available from: <http://www.bmj.com/content/350/bmj.h1885.abstract>
2. Pendergrass, S. A., & Crawford, D. C. (2019). Using electronic health records to generate phenotypes for research. *Current Protocols in Human Genetics*, 100, e80. doi:10.1002/cphg.80
3. Alzoubi, H., Alzubi, R., Ramzan, N., West, D., Al-Hadhrami, T., & Alazab, M. (2019). A Review of Automatic Phenotyping Approaches using Electronic Health Records. *Electronics*, 8(11), 1235. MDPI AG. Retrieved from <http://dx.doi.org/10.3390/electronics8111235>

For more information, contact:

Alexandra Makhlysheva, Senior Adviser
alexandra.makhlysheva@ehealthresearch.no