Strategies Toward Identifying Undiagnosed Rare Disease Patients

Douglas Drake, Clinerion & Christopher Rudolf, Volv Global

1. Introduction

Common estimates are that as much as 50% of many rare disease patients are never correctly diagnosed due to heterogenous presentation, lack of immediate medical expertise in recognizing symptoms and access to the specific diagnostic testing necessary. At the same time, rare diseases not only impact the patient, but also their family and social support structure as they often search to understand the cause for the disease and its impact on their life and capabilities.

2. Aims and Strategies

With increasing digitalization, we now have masses of structured, coded information on patients available in electronic health records (EHRs), incl. demographic data, diagnoses (e.g. using ICD-10), medications, procedures and laboratory results.

We explore how EHR-based real-world data enables the ability to find potential patients that have been diagnosed as well as those that have not been correctly diagnosed.

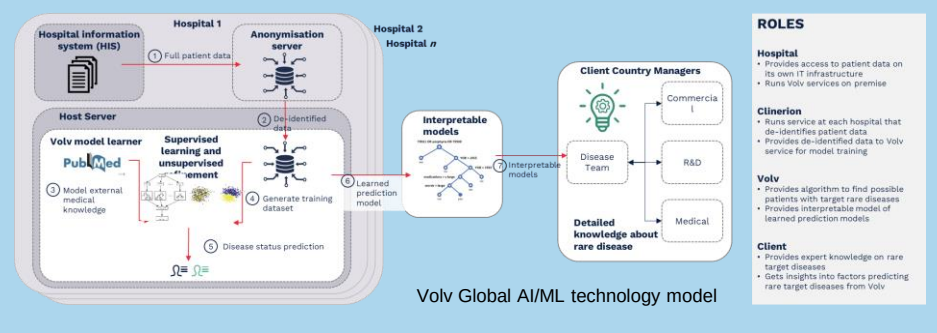
We explore how EHR systems can be used:

1. to identify existing diagnosed patients.
2. to identify undiagnosed patients through related symptoms.
3. to model more heterogenous disease response using an AI/ ML approach to find patients yet undiagnosed.

2.1. Direct Identification via Diagnosis Codes

In the first level, the existing EMR data can be readily used for patient stratification of rare disease patients. Treatment modalities for rare disease patients are expanding through next generation gene therapeutics that directly treat the underlying genetic cause of many of these diseases.

The EHR search example shown below uses Clinerion's Patient Network Explorer, an EHR-based patient network which includes many high-density rare disease regions, and can be used to readily identify and trace rare disease patients' journeys, longitudinally.



Volv Global AI/ML technology model

2.2. Triangulation from Related Symptoms

In the second level, by applying sponsor-based disease symptom models to the EHRs in Clinerion's Patient Network Explorer platform, we can identify undiagnosed rare disease patients by triangulating from data in their anonymized EHR records.

For example, Fabry disease affects several organ functions, including the kidney and cardiovascular systems, and causes pain. By looking at these symptoms in various combinations, and also taking note of (non-)response to treatments for these symptoms individually, we can narrow in on patients for whom the root cause of all their symptoms is the rare disease, Fabry, sidestepping their incorrect diagnosis.

In the table on the right, we show the results in five sites in two different countries.

2.3. Leveraging AI/ML

In the third level, we leverage the power of advanced analytics to learn how to identify Fabry patients without developing a disease symptom model, but by applying artificial intelligence (AI) and machine learning (ML) algorithms.

These AI/ML algorithms develop detailed models based upon all possible heterogenous disease presentations and novel predictive biomarkers.

Clinerion can apply AI/ML technology provided by Volv Global. Their model allows AI/ML across structured and anonymized data, enabling more sophisticated search and pattern-mapping of patient care data and care metrics within the platform. Longitudinal modelling from EHRs permits individual patient journeys to be modelled. The results provide an algorithm which identifies patient phenotypes and predicts disease progression that allows replication in other geographies, on other platforms. For Fabry, the Volv Global model is over 95% accurate; the Volv Global AI methods and results will be published separately.

3. Results

We applied these approaches to a sponsored search for Fabry, Pompe, Gaucher and Mucopolysaccharidosis Type 1 diseases in Turkey and in the United Arab Emirates, using an existing real-world data platform linking an international network of hospitals.

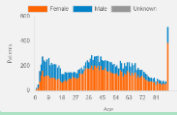
3.1. Direct Identification via Diagnosis Codes

Fabry & Gaucher

Regions	Sites	Patients
EUROPE	22	363
ASIA	24	2.7K
SOUTH AMERICA	9	2.99K

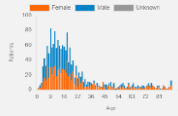
Pompe E74

Regions	Sites	Patients
EUROPE	22	3.11K
ASIA	28	12.5K
SOUTH AMERICA	8	1.08K



MPS1 E76

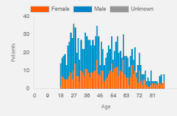
Regions	Sites	Patients
EUROPE	16	148
ASIA	20	1.36K
SOUTH AMERICA	8	121



3.2. Triangulation from Related Symptoms

Pompe E74

Regions	Sites	Patients
ASIA	5	1.34K



3.3. Leveraging AI/ML

Clinerion is developing a machine learning infrastructure that will be aligned with local installations at the partner hospitals, keeping the patient records on-site. Such a federated AI/ML infrastructure model would use local training data to develop and test pattern-learning methodologies and subsequently optimize the global pattern training for the AI/ML model, preserving privacy both at the local and federated levels.

This project in ongoing and has been set up in partnership with key academic AI researchers within Switzerland and Innosuisse, the Swiss Innovation Agency.

4. Conclusions

The project shows the potential for using EHR data for rare disease stratification as well as potential identification of patients incorrectly diagnosed, and outlines a methodology that can be used to identify patients for diagnostic testing using digital technologies.

We have outlined approaches to diagnosing rare disease patients, how EHR data can be used to create a targeted patient outreach and diagnostic effort in concert with the hospital. The use of symptom models to query and retrieve patient cohorts rather than simply querying disease codes can enrich undiagnosed rare disease patients between 15 to 20% from the estimated undiagnosed patient population, resulting in better patient care, stratification and study capability.

We anticipate a future in which anonymized EHRs can be used for better diagnosis of patients, especially rare disease patients, who need accurate diagnoses to access the right treatments.