Reducing diagnostic delays in Acute Hepatic Porphyria (AHP) using a machine learning approach

Kat Schmolly, MS2, Simon Beaven, MD/PhD, Olive View-UCLA Medical Center and Pfleger Liver Institute, UCLA

CTSI Summer Research Fellowship Program 2021, David Geffen School of Medicine at UCLA. This is part of an ongoing joint venture study with UCSF.

Study Question

Can routinely collected clinical data available in electronic health records systems be used to identify undiagnosed patients who may be suffering from acute hepatic porphyria?

Background

- Acute intermittent porphyria (AIP) is a rare genetic disease (~ 1:20,000) that manifests primarily as periodic attacks of crippling abdominal pain, often presenting in young women in a cycle-dependent fashion
- Dx: abnormal accumulations of the heme precursor porphobilinogen (PBG) in urinalysis during acute attack
- Suspected neurotoxic effects of PBG over time causing irreversible chronic comorbidities: liver and kidney disease, chronic neuropathy
- AIP can be prevented with RNA silencing drug
- Acute phases treated with IV heme

- Access to healthcare and specialist referrals needed to diagnose
- Typical diagnostic delay of 10-15 years

- Joint venture project between UCLA and UCSF to analyze tens of thousands of EMRs to identify clinical patterns consistent with porphyria in machine-driven contextual review to help identify potential misdiagnosed, latent and undiagnosed AIP patients – reducing diagnostic delays, unnecessary testing, associated costs and morbidity.

Methods

- Step 1: Existing EMR data from AIP patients within the UCLA and UCSF Health Network who have been clinically evaluated for AIP (incl diagnosed and excluded) to be used as positive and negative control for generation of MLA.
  - Labels used for pattern detections are identified manually as well as via auto-extraction
  - Examples of labels used for pattern detection: ALA/PBG, clinical notes, genetic information, abdominal CT scans, ED/UC visits for GI/Abdominal pain, GI/OBGYN referrals
  - Labels will be added throughout study as they appear relevant to patterns
- Step 2: We will test the algorithm by identifying and evaluating potentially undiagnosed AIP patients with confirmatory urine ALA/PBG testing and genetic testing.

Results

- This study is still in progress and is expected to conclude by Summer 2022
- At this time, we are creating the algorithm using preliminary patient data available from the UCSF and UCLA patient populations.
- 144 confirmed AIP patients were found to form the positive control baseline along with 57 evaluated and excluded patients forming the negative control baseline.
- 905 possibly undiagnosed AIP patients were preliminarily identified and will be further evaluated to confirm or exclude a dx of AIP.

- The study population will continue to grow as our CTSI is extracting data from the EMR and the Atlas Genome database.

Discussion

This study is in progress until Summer 2022

- **Further investigations:** For those patients with whole genome sequencing available, an internal case-control study will be performed to try to identify modifier genes that affect the phenotypic presentation of AIP.

- **Limitations:** Patient pools for studies of rare diseases are typically smaller in size, potentially affecting algorithm detection rate. Further, lack of access to outside records of existing UCLA/UCSF patients can affect pattern detection. The creation of Machine Learning Approaches for detection of other rare diseases requires specialist manpower and existing detailed knowledge of the diseases and their clinical presentation. Transferability and application of this algorithm to other EMRs and health systems outside of UCLA/UCSF (Epic) is unknown at this time

Implications

If this machine learning approach proves successful for increasing detection of AHP/AHIP, similar rapid, effective pattern recognition algorithms from the EMR could then be applied to many other rare diseases, including other types of porphyrias and rare abdominal pain syndromes (i.e.. familial Mediterranean fever (FMF), IgG4-related disease). Algorithm widgets could then be deployed as part of EMR systems in primary care clinics and areas without access to specialists to reduce diagnostic delays and unnecessary testing, medical expenses, comorbidity development and patient suffering.