

# A Data Analytics Approach to Rare Disease Discovery

Kenneth V Lieberman, MD Section of Pediatric Nephrology, Hackensack University Medical Center Hackensack Meridian School of Medicine at Seton Hall University



Hackensack Meridian *Health* Joseph M. Sanzari Children's Hospital

### Background

- The ubiquitous adaptation of EMR's following the mandate in the Affordable Care Act has spawned multiple big data applications in medicine.
- Hypophosphatasia (HPP) is a rare genetic disease of bone formation that causes severe skeletal abnormalities that can result in growth retardation, bone pain, and pathologic fractures as well as dental abnormalities. The disease frequently starts in early childhood. ERT is available.
- Goal: Design and implement a search strategy to find HPP patients in an EMR.

#### Methods

- The overnight backup store for EPIC Chronicles database is searched with the EPIC Clarity tool
- The initial search parameter is the age and sex adjusted cutoff for hypophosphatasia (low AP)
- The initial patient at risk pool is progressively winnowed down by eliminating patients with other entities known to result in low AP values

#### Limitations

 The study technique does not assure that all of the effected patients have been identified

#### Conclusions

 A phenotypic search strategy can be used in the EPIC environment to find previously undiagnosed patients with rare disease

#### Algorithm Development

- 5,986 unique AP determinations on 3,733 patients out of the 1,870,292 patients in HUMC EPIC were enumerated
- Eliminate patients with at least one normal AP value
- 132 candidates <21: 80; >21: 52
- Eliminate patients with Pernicious Anemia
- Eliminate patients with Multiple Myeloma
- Eliminate hypothyroid patients
- Eliminate malnourished patients, mostly cancer and IBD (AP is a nutrition marker
- Go to outside labs
- Examine signs/symptoms/imaging (s/s/i) for confirming features
  Hurdles
  - nur
- Inadequate ICD coding
- Finding hypothyroid patients
  - Lab values
  - Medication record
- Patients with inadequate information
  - Only one AP value
  - No imaging for confirmation

# Challenges

- Physicians are poor utilizers of ICD coding
- Under what auspices to proceed? IRB?
- Is this human subjects research or quality improvement?
- HHS OHRP guidance: https://www.hhs.gov/ohrp/regulationsand-policy/guidance/faq/quality-improvementactivities/index.html
- Publication: SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised guidelines from a detailed consensus process
- Contacting patients (what are their expectations)
- Care team concerns (medical liability)
- Integrate outside clinical data sources





#### Candidate Disease Risk Stratification

- Low risk (few AP values, no suggestive s/s/i): 29
- Medium risk (suggestive s/s/i): 7
- High risk (many AP values, suggestive s/s/i): 5
- Highest risk: 1 young adult admitted with multiple long bone fractures after minor trauma; had many very low AP values; diffuse osteoporosis

## **Next Steps**

- Validate at other EPIC sties
- Develop an algorithm of algorithms
- Expand to other rare diseases
- Pursue state HIE's
- Untrapping free text (unstructured data in clinical documents) through NLP with the application of medical ontologies
- Utilize machine learning strategies to streamline the development process

Acknowledgement: Thanks to Pinkrose Hamilton and the Business Intelligence team