De-identified Aggregate Electronic Health Record Data as a Resource for Understanding Patient Trajectories for Pediatric Leukemia

Wood, N[1]; Glynn, E[1]; Martin, S[1]; Lewing, K[1]; Caragea, D[2]; Noel-MacDonnell[1], J; Hoffman, M[1]

[1]Children's Mercy Kansas City, Kansas City, Mo.; [2]Kansas State University, Manhattan, KS.

Background

- Cancer patients have complex medical encounters with lab work and diagnostic tests followed by multi-modal treatment regimens. They can experience unique complications, co-morbidities, recurrence and even death. They may also have heath care experiences before their oncologic diagnosis that provide early warning signs to their risk.
- Electronic health record (EHR) data provides a detailed collection of information describing these patient interactions with the health care system. De-identified EHR data provides a useful alternative to registries, billing and claims data (Figure 1).
- Assembling the digital timelines for a large cohort of patients can enable researchers to compare trajectories of patients with poor outcomes (relapse, morbidity and mortality) to those with favorable outcomes (remission). This resource enables the analysis of patient data preceding their cancer diagnosis as well as long term, non-cancer, outcomes after treatment.

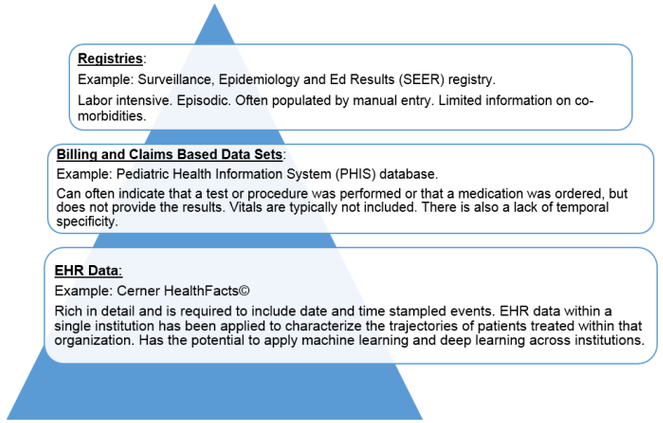

Figure 1. Comparison of secondary databases

Methods

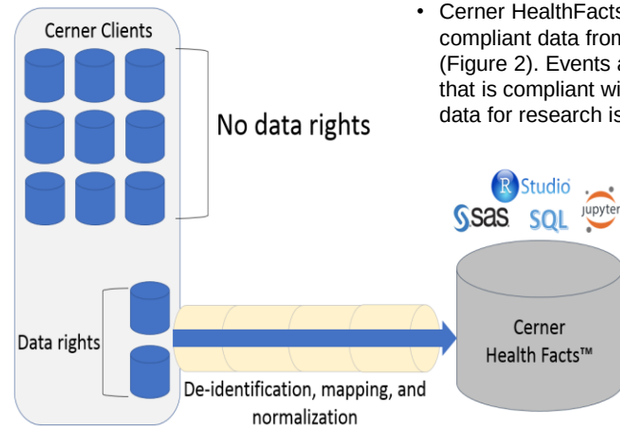

Figure 2. EHR Data Flow: Cerner Health Facts™

- Cerner HealthFacts© (HF) contains de-identified, HIPAA compliant data from Cerner clients that agree to participate (Figure 2). Events are date and time stamped in a manner that is compliant with HIPAA (PHI). The amount of available data for research is vast (Table 1).

Data Type	Current Release
Unique patients	65 million
Total laboratory results	4.3 billion
Total facilities	853
Total medication orders	734 million
Total diagnoses	489 million

Table 1. Data included in the HF database.

- HF includes laboratory data, inpatient medications, demographics, surgical data, and a wide variety of "clinical events," including vitals and billing data.
- We identified ICD9 and ICD10 codes associated with ALL and AML and queried the data to determine how many children are represented with these conditions from 0 to 18 years old.
- A representative patient with ALL was identified. Date and time for the occurrence of tests, procedures, vitals and other were plotted. Select results were plotted over time.

Discussion

- This project will lead to re-usable resources and processes that will enable researchers in pediatric oncology to develop and evaluate predictive models, perform outcomes research and to complete health systems research.

Preliminary Data/Results

- Preliminary queries indicate that there are 18,785 patients with the diagnosis of leukemia in the HF database (Table 2).
- Beginning with individual patients, we have demonstrated data rich visualizations showing the spacing of clinical activities and trends in the values for lab tests and vitals (Figure 3).
- We are developing the data science processes focused on pediatric cancers, beginning with acute lymphoblastic leukemia (ALL) and acute myelogenous leukemia (AML), from HF that are optimized for "trajectory analysis."

Diagnosis	Number
ALL	16,120
AML	1,695
ALL/AML	970
Total	18,785

Table 2. Number of patients based on leukemia type in HF database. ALL (acute lymphoblastic leukemia); AML (acute myelogenous leukemia)

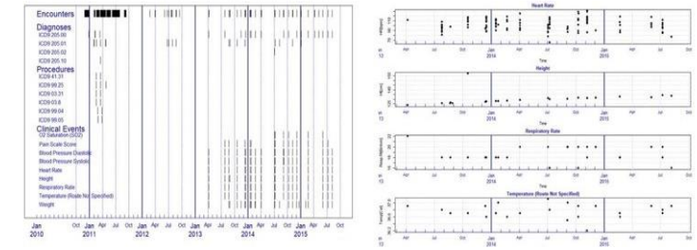

Figure 3. Example patient trajectory generated from HF data

Children's Mercy KANSAS CITY

KANSAS STATE UNIVERSITY

UMKC School of Medicine