

Building a Natural Language Processing Tool to Identify Acute Chest Syndrome in Children with Sickle Cell Disease

Presenting Author: Allan Simpao, MD, MBI, The Children's Hospital of Philadelphia

Co-Authors: Brian Park, MS3, Drexel University College of Medicine; Luis Ahumada, PhD, The Children's Hospital of Philadelphia; Jorge Galvez, MD, MBI, The Children's Hospital of Philadelphia

Introduction: Acute chest syndrome (ACS) is a debilitating complication of sickle cell disease (SCD). Studies have reported the prevalence of perioperative ACS based on billing codes, but not on actual clinical reports. Natural language processing (NLP) is a method by which computers use algorithms to analyze and understand the human language. NLP tools can be used to analyze clinical documentation to identify disease presence and can potentially reduce time and error in processing manual clinical reports. One example of a clinical NLP application is in identifying venous thromboembolism from ultrasound reports. However, in order to apply and validate the NLP tool, a curated data set must be built and key words and phrases must be identified. Thus, we performed a review of anesthesia information management system (AIMS) and electronic health record (EHR) data to build a curated data set of ACS in SCD patients undergoing general anesthesia for surgery.

Methods: The AIMS and EHR databases were queried to identify patients with sickle cell disease who underwent general anesthesia at The Children's Hospital of Philadelphia between Jan 1, 2009 and May 1, 2014. Two individuals independently reviewed patients' charts manually for the ACS clinical criteria, and then recorded keywords in the clinical documentation that were associated with the diagnosis.

Results: A total of 327 SCD patients met the study criteria; the study cohort was 54% male, median weight was 32.9 kg (IQR 18-48), and median age was 8.8 years (IQR 4-14). Sixteen cases of ACS were identified. The most frequently occurring terms in the ACS patients' documentation belonged to three categories: signs/symptoms ("fever", "febrile"), radiographic findings ("infiltrate"), and medications ("ampicillin", "azithromycin").

Conclusion: Reviewing the AIMS and EHR databases manually for clinical documentation of ACS was feasible yet laborious and time consuming. The curated data set will be used to validate the NLP tool and apply it to other clinical documentation. Successful application of the NLP tool will facilitate mining the AIMS and EHR for future data-driven outcomes and quality assurance projects involving sickle cell patients and ACS. The same methodology can be applied to other clinical diagnoses in future studies.