Leveraging Clinical & Biologic "Big Data" to Advance Clinical, Quality, and Scientific Efforts for Pediatric Severe Sepsis

Rationale: Severe sepsis is life-threatening organ dysfunction caused by a systemic infection and is a leading cause of morbidity, mortality, and health care-associated costs. Globally, there are an estimated 2-5 million cases of pediatric severe sepsis with a mortality up to 24% for those treated in a pediatric intensive care unit (PICU). In the United States, 75,000 annual admissions and nearly 10,000 deaths in children are attributed to severe sepsis with a cost exceeding $5 billion. At CHOP, several hundred children are treated for severe sepsis in the emergency department or inpatient setting each year with a 12% mortality rate for those children requiring treatment in the PICU. In 2015, severe sepsis accounted for 41% of all PICU deaths at CHOP. A critical challenge to improving outcomes for children with severe sepsis is that accurate operationalization of non-specific consensus criteria has made identification of true cases of pediatric severe sepsis nearly impossible without comprehensive, time-consuming, and expensive manual chart reviews.

Two programs are focused on sepsis at CHOP—the hospital-wide Sepsis Governance Committee oversees clinical and quality improvement efforts and the Pediatric Sepsis Program bridges multidisciplinary research efforts. Both programs currently struggle with early and accurate case identification that would enable "real-time" feedback of clinical care, , facilitate enrollment into the CHOP Sepsis Registry, and enhance patient access to research opportunities (including reflex collection of residual biological samples). To date, sepsis leaders at CHOP have worked with clinical bioinformatics (i.e., EPIC-based analysts) to develop computerized criteria to identify patients with suspected sepsis using available data within the electronic health record. This approach avoids the substantial bias of using ICD-based administrative codes to identify patients with severe sepsis, but has been limited by available resources to iterate the optimal criteria and by a lack of a gold-standard comparison. New funding through the CHOP Department of Pediatrics Chairs' Initiative now supports a robust adjudication process to differentiate severe sepsis from alternative diagnoses. Moreover, recent publication of a validated computerized methodology that leverages data available within the electronic health record to identify cases of adult sepsis provides critical new data on which to further enhance existing efforts to "case-find" cases of pediatric severe sepsis at CHOP. The availability of an adjudicated "gold-standard" comparison and a validated adult model provides a new opportunity to develop a pediatric severe sepsis "case-finder" that would benefit numerous efforts within the CHOP Sepsis Governance Committee and Pediatric Sepsis Program, promote CHOP as a leader in the field of pediatric severe sepsis, and provide a paradigm that could be extended to other pediatric disorders.

Aim 1: Derive and validate a computerized algorithm that leverages data available within the electronic health record to identify cases of pediatric severe sepsis within 24 hours of clinical diagnosis

We will bring together a multidisciplinary group of bioinformatic and clinical experts to improve the existing computerized methodology used to identify cases of pediatric severe sepsis. Currently, cases are identified using non-specific criteria that suggested clinician concern for suspected infection, most of whom do not have confirmed infection and only a fraction of whom have organ dysfunction required for severe sepsis. For this aim, we will develop code to overlay an algorithm onto the EHR, similar to that used by Rhee et al., to incorporate organ dysfunction. Although Rhee et al. used criteria from the adult-specific Sequential Organ Failure Assessment (SOFA) score, we will substitute criteria from the recently validated pediatric SOFA score. We will validate the resulting computerized algorithm against adjudicated cases of pediatric severe sepsis with serial iterations as needed to optimize test characteristics.

Aim 2: Link the validated pediatric sepsis "case-finder" to automated clinical data extraction and available residual biological samples

We will work with a multidisciplinary group of bioinformatics and clinical experts to link pediatric severe sepsis case identification with automated data extraction into a revised version of our CHOP Sepsis Registry. We will evaluate existing code used to extract clinical data from the EHR for the 2012-2015 CHOP Sepsis Registry that was initially developed by the Department of Biomedical Health Informatics. For this proposal, we will improve accuracy of data collection and build new code to automate extraction of organ dysfunction measures. Finally, since all clinical specimens are electronically logged, we will work with CHOP clinical labs to identify residual biological specimens that could be used for research in lieu of additional specimen collection and/or collected for biorepository storage with appropriate ethical safeguards (including DNA samples to be banked at the CHOP Center for Applied Genomics).

"Pie in the Sky" potential future projects that could stem from this work:

1. Develop a clinical predictive model that can identify children at risk for sepsis upon hospital arrival that continues through their hospital stay. Could use the outcomes derived in Aim 1 as outcome.
    a. Initially would incorporate broad-swath of all clinical data available in EPIC (including medical history, prior admissions). Would be interesting to investigate and compare multiple possible machine learning algorithms for this.
    b. Could build upon this using more granular vital sign data (ie from monitors)
    c. Long term could consider using genetic predictors for patients in whom genomes are sequenced or other biomarkers that could be known prior to patient arrival at CHOP
    d. Clinical decision support could be built around all of this: many permutations possible particularly if alert can "learn" by incorporating increasing data during a patients ED/hospital day- alert could then be used throughout the hospital and throughout the patient's stay
    e. Could add bio-samples of at risk patients to bio repository above
    f. Could also utilize the reverse algorithm to identify children at low risk to develop antibiotic stewardship programming.
2. Expand "case-finding" algorithm to build a multicenter, prospective sepsis registry that could replace reliance on limited retrospective analysis of administrative databases to study the epidemiology of pediatric sepsis, ideally with linked bio-samples.