**Pediatric Hematology/Oncology/BMT: MACC Internal Funding Request FY23**

**PROPOSAL SECTIONS:**

**A. Abstract (250 words maximum)**

Much of our existing knowledge about the natural history and disease related complications in individuals with sickle cell disease (SCD) is based on the Cooperative Study of Sickle Cell Disease (CSSD) which was conducted in the 1980s. We know that individuals with SCD suffer from painful vaso-occlusive crises and are also at high risk of chronic end organ damage affecting heart, lung, kidney, eyes, bones, and brain. Since the CSSD concluded, multiple changes in disease management practices have occurred, drastically improving pediatric mortality rates. These changes warrant a re-evaluation to understand the epidemiology and natural history of the disease in the modern era. In the proposed project, we will leverage electronic health record data from a research network to determine the prevalence of co-occurrence of complications impacting the cardiopulmonary and renal system among children with SCD. Further, risk factors for the development of cardiopulmonary and renal complications among children with SCD in the modern era are unclear. We therefore propose to use using existing data from a research network, TriNetX to understand the prevalence, co-occurrence and risk factors for cardiopulmonary and renal complications among children with SCD. This research network includes electronic health record data for more than 25,000 patients with SCD from multiple sites across the nation.

**B. Specific Aims/Hypotheses (1/2 page maximum)**

There have been multiple changes in sickle cell disease (SCD) management practices in the recent decades, thereby drastically improving the pediatric mortality rate. However, children with SCD still suffer from chronic complications, which as they grow older, are shown to be associated with early mortality. A recent study in adults suggests renal and cardiopulmonary complications tend to cluster together, and multiple end-organ impairments are associated with poor survival. Whether this clustering occurs in children with SCD is unknown. In addition, knowledge of risk factors for these complications among children and adults with SCD is limited. Existing literature reflect mixed findings on the ability of laboratory parameters and history of acute complications to predict chronic end organ damage, and majority of data include the adult population. Thus, to address these knowledge gaps, we propose to leverage electronic health record data to help further understand the epidemiology of chronic complications among the pediatric SCD population.

**Aim 1: To determine the prevalence of co-occurrence of cardiopulmonary and renal end organ damage in children with sickle cell disease**

Hypothesis: We hypothesize that 15% of children with SCD will have co-occurrence of cardiopulmonary and renal complications.

**Aim 2: To determine risk factors for cardiopulmonary, renal complications and the co-occurrence of these in children with sickle cell disease.**

We hypothesize that a history of frequent acute SCD disease complications, low hemoglobin, low hemoglobin F and increased levels of proinflammatory markers will predict risk of end organ damage.

.

**C. Research Strategy**

1. **Significance**

**There is limited understanding of development of chronic complications in children with SCD.**

SCD, a genetic disorder, affects approximately 100,000 people in the United States. Much of our existing knowledge about the epidemiology of this disease in the United States primarily comes from the Cooperative Study of Sickle Cell Disease (CSSD) which was conducted in the 1980s1. This hallmark study has provided us with valuable insights regarding disease related complications experienced by individuals with SCD. However, there have been multiple changes in disease management practices since CSSD was concluded. Some changes that have transformed the disease landscape include universal newborn screening, broader antibiotic coverage to prevent pneumococcal disease2,3, use of hydroxyurea as a disease modifying therapeutic4-6 and use of Transcranial Doppler screening for prevention of stroke7. These changes warrant an updated epidemiological study for re-evaluation of the burden of disease related complications in the present times. However, traditional prospective epidemiologic studies to determine the prevalence of chronic complications among children with SCD are challenging to conduct due to slow recruitment rates, retention of participants and a smaller prevalence of chronic complications in this rare disease. Recent evidence shows that clustering of cardiopulmonary and renal complications is common in adults with SCD, and that multiple end-organ impairments are a significant predictor of mortality8. However, whether there is clustering of end-organ complications in children with SCD is unknown. *We therefore propose to use existing big data of electronic health records to identify a cohort of children with SCD and determine the co-occurrence of cardiopulmonary and renal complications among these children.*

**Understanding risk factors for chronic complications is essential to inform decisions related to therapeutic and preventative interventions.** Individuals with SCD experience anemia, hemolysis, inflammation and ischemia-reperfusion injury; and these events interplay with each other to result in the acute and chronic complications of SCD9,10. However, there is limited knowledge about the risk factors for chronic end organ damage. The predictive value of biomarkers reported for SCD across the literature is unclear11. In addition, existing literature on biomarkers for SCD reflect mixed findings12-15. For instance, the variables found to predict adverse outcomes in children in CSSD could not be replicated in another independent cohort of children with SCD12,13. Moreover, many studies evaluating biomarkers tend to focus on a single organ16-18. This knowledge gets harder to implement in clinical practice, especially when the expert guidelines and clinical care often combines the cardiopulmonary and renal complications together. *In this project, we propose to identify clinical and laboratory parameters that can predict chronic end organ damage in patients with SCD, leveraging the existing big data of EHRs.*

**Knowledge of the complications of SCD is necessary to inform decisions related to therapeutic and preventative interventions.** The knowledge of which individuals with SCD are at high risk of developing chronic end organ complications early in life is essential to aid in tailoring care for maximal benefit for these patients. Identification of SCD patients who are at high risk of a severe disease phenotype would allow early initiation of treatments to prevent the downstream complications or initiation of screen strategies to identify early end-organ decline. Our work will also inform who could potentially benefit the most from recently approved drugs19-21, bone marrow transplant or other potential treatments under development including gene therapy22, and thus help in clinical decision making. Finally, our work will also provide directions for future research to further understand biological mechanisms of complex SCD pathophysiology.

1. **Innovation**

* **The proposed study would be the largest longitudinal cohort study ever conducted for SCD.** We will use existing EHR data from multiple sites across the United States that contribute data to a research network (TriNetX). These data include more than 25% of the estimated number of 100,000 people living with SCD in the United States. Further, these data include long periods of follow-up for thousands of patients with this disease. A traditional prospective cohort study to examine the longitudinal course of the disease would be highly expensive to conduct and would not be feasible given how long it would take for patients to age over time.
* **These data offer the advantage of having laboratory results which are generally not available through other big data sources such as administrative data.** These data allow for inclusion of a combination of laboratory parameters and clinical information for a large cohort of patients with SCD to predict the risk of chronic end organ damage.
* **The PI and her research team will use validated algorithms23,24 to identify individuals with SCD and conduct sophisticated statistical analyses for the proposed work**. The multidisciplinary project team, consisting of the principal investigator for the project who is an epidemiologist (Singh) who has expertise in big data and outcomes research for SCD patients, a hematologist (Brandow) who provides content expertise and has health services research experience, and a biostatistical expert (Brazauskas) who has expertise in longitudinal data analysis and risk prediction. This team is well positioned to conduct the proposed study and infer clinical meaning and implications of the findings.

1. **Approach**

**Preliminary data**

**TriNetX research network data for individuals with SCD**:The research network includes 27,592 patients with a diagnosis code for SCD from more than 30 healthcare organizations across the United States, of which 57% are females. This cohort includes children (N = 7,536) and adults (N = 20,056), the mean (standard deviation) age being 32 (sd = 19) years. Of the patients included in the cohort, 74% are Black, 9% are White and for 16% race is unknown. This cohort includes a small proportion of Hispanics/Latinos (3%), the remaining are Non-Hispanics (63%) or have missing ethnicity (34%).

**Generalizable algorithms for identification of SCD patient cohorts within the EHR**23,24: My prior research expanded previous work done to accurately identify a SCD cohort within the HER. More recent work includes the use of International Classification of Diseases (ICD) codes, version 10. The revised algorithm identifies children with SCD with a sensitivity of 93.3% and a positive predictive value of 97.9%. Our research team also created an algorithm to determine the genotype of SCD. Using our algorithm, we can identify patients with hemoglobin SS/Sβ0 thalassemia disease (commonly referred to as sickle cell anemia) with a sensitivity of 90% and a specificity of 97% within the cohort of SCD patients. In addition, we created the computable algorithm to identify vaso-occlusive pain crises encounters for patients with SCD. This algorithm, using a composite definition that includes information based on ICD codes and the administration of intravenous pain medication, can identify acute pain crises encounters with a sensitivity and specificity of 95.1% and 96.1% respectively. These algorithms were applied across patient-Centered Outcomes Research Institute’s Greater Plains Collaborative (GPC) network which contains EHR data from twelve leading medical centers in various states. The distribution of patients with sickle cell anemia observed within the network was consistent with prior studies across the nation and hence provides face validity to our algorithms. This shows our ability to work with multisite EHR data and experience with standardized coding systems, positioning us to conduct the proposed research.

**Study design and population:** The proposed study is a retrospective cohort study leveraging the longitudinal data available during the study period (2010 – 2020). The study population will include individuals with SCD of <19 years of age and all genders. This population will be identified within the existing EHR data using previously developed algorithms described in the preliminary data section.

Data source: We propose to use existing EHR data from more than 30 sites across the nation that are part of a research network, TriNetX25. This research network optimizes clinical research and enables discoveries through the generation of real-world evidence, by providing real-time access to longitudinal clinical data. This research network already has invested in data infrastructure and maintains high-quality deidentified patient level data, which is refreshed on a regular basis, weekly – monthly depending on the site. Specifically, this network provides an anonymized dataset of electronic medical records (diagnoses, procedures, medications, laboratory values) for approximately 28,000 patients with a diagnosis of SCD. We will request the de-identified data online via the network specific platform, which can be accessed by the research team.

**Statistical analysis**

***Aim 1:* To determine the prevalence of co-occurrence of cardiopulmonary and renal end organ damage in children with sickle cell disease.**

Primary outcomes (defined): Cardiopulmonary and renal end organ damage will be identified using ICD codes and clinical laboratory markers8. Specifically, FEV1% predicted values of ≤70% would be considered as end-organ lung disease. A tricuspid regurgitant jet velocity (TRJV) cutoff of ≥2.5 m/second will be considered as abnormally elevated and indicative of pulmonary hypertension based on the study by Gladwin et al.16 and the American Thoracic Society clinical practice guideline on pulmonary hypertension in SCD. Nephropathy will be defined as a serum creatinine of ≥2 mg/dL or 1.5 times the baseline creatinine (when this creatinine value is ≥1 mg/dL) or an estimated glomerular filtration rate of <60 mL/min/1.73 m[2](https://onlinelibrary.wiley.com/doi/10.1002/ajh.25202#ajh25202-bib-0002) body surface area persistent over at least 3 months based on the National Kidney Foundation Kidney Disease Outcomes Quality Initiative guidelines.

Analyses: We will describe the prevalence of chronic complications as percentage of patients who meet criteria for cardiopulmonary or renal end organ damage. We will also assess the prevalence of children who have the co-occurrence of complications impacting both cardiopulmonary and renal systems.

***Aim 2:* To determine risk factors for cardiopulmonary and renal complications in children with sickle cell disease.**

Primary outcome: Cardiopulmonary and renal end organ damage will be identified as defined above.

Variables of interest: These include sex, age, genotype, disease modifying therapy and laboratory parameters (i.e., Hemoglobin F/Hemoglobin in total blood, Hemoglobin (g/dl), makers of hemolysis26, platelet count, total leukocyte count, percent lymphocytes, percent monocytes, percent eosinophils, percent basophils).

Analyses: We will perform a series of landmark analyses designed to estimate end organ damage probabilities of patients who are of a certain age at a landmark time point. For example, the first landmark analysis group will consist of all 10-year-old patients who have not developed end organ damage by that age. Cumulative incidence curves will be used to estimate the probability of end organ damage in subsequent years among these patients. Death will be treated as a competing risk. Patients will be considered as lost to follow up if they have no data for a continuous period of two or more years and be then censored in our analyses. In addition, Cox proportional hazards model will be used to identify risk factors significantly associated with end organ damage. Covariates which change over time and/or whose values are measured repeatedly will be included in the model as time varying covariates. We will ensure that Cox model assumptions are satisfied and will take appropriate steps to remedy the situation should we find any departures from proportional hazards assumption. That may include stratified Cox models or introduction of time dependent covariates where needed. Eventually, one model will be fitted to combine all of the individual landmark analyses. All landmark analyses will be done following the methodology established by van Houwelingen et al.27,28

**D. Funding Trajectory [*How will these funds be leveraged to generate external grant applications?*] (1/2 page maximum)**

The proposed work will lay the foundation for the ability to understand the prevalence of overlap of cardiopulmonary and renal complications among children with SCD. These data will enable the ability of the primary investigator to apply for external funding to understand the time to development of these complications which could provide epidemiological data to facilitate effective management strategies to prevent these complications. Collectively, identification of risk factors for cardiopulmonary and renal complications will allow the primary investigator to prepare for a Mentored Research Scientist Career Development Award (NIH Parent K01)to further incorporate advanced prediction methodologies such as machine learning and compare predictive abilities of various models.

**E**

**References**

1. Farber MD, Koshy M, Kinney TR. Cooperative Study of Sickle Cell Disease: Demographic and socioeconomic characteristics of patients and families with sickle cell disease. *J Chronic Dis.* 1985;38(6):495-505.

2. Gaston MH, Verter JI, Woods G, et al. Prophylaxis with oral penicillin in children with sickle cell anemia. A randomized trial. *N Engl J Med.* 1986;314(25):1593-1599.

3. Cober MP, Phelps SJ. Penicillin prophylaxis in children with sickle cell disease. *J Pediatr Pharmacol Ther.* 2010;15(3):152-159.

4. Yawn BP, Buchanan GR, Afenyi-Annan AN, et al. Management of sickle cell disease: summary of the 2014 evidence-based report by expert panel members. *Jama.* 2014;312(10):1033-1048.

5. Thornburg CD, Files BA, Luo Z, et al. Impact of hydroxyurea on clinical events in the BABY HUG trial. *Blood.* 2012;120(22):4304-4310; quiz 4448.

6. Lanzkron S, Strouse JJ, Wilson R, et al. Systematic Review: Hydroxyurea for the Treatment of Adults with Sickle Cell Disease. *Annals of Internal Medicine.* 2008;148(12):939-955.

7. Adams RJ, McKie VC, Hsu L, et al. Prevention of a First Stroke by Transfusions in Children with Sickle Cell Anemia and Abnormal Results on Transcranial Doppler Ultrasonography. *New England Journal of Medicine.* 1998;339(1):5-11.

8. Chaturvedi S, Ghafuri DL, Jordan N, Kassim A, Rodeghier M, DeBaun MR. Clustering of end-organ disease and earlier mortality in adults with sickle cell disease: A retrospective-prospective cohort study. *American Journal of Hematology.* 2018;93(9):1153-1160.

9. Piccin A, Murphy C, Eakins E, et al. Insight into the complex pathophysiology of sickle cell anaemia and possible treatment. *Eur J Haematol.* 2019;102(4):319-330.

10. Elliott P Vichinsky. Overview of the clinical manifestations of sickle cell disease. UpToDate. Waltham, MA Web site. Published Accessed January 2020. Accessed.

11. Rees DC, Gibson JS. Biomarkers in sickle cell disease. *British Journal of Haematology.* 2012;156(4):433-445.

12. Quinn CT, Lee NJ, Shull EP, Ahmad N, Rogers ZR, Buchanan GR. Prediction of adverse outcomes in children with sickle cell anemia: a study of the Dallas Newborn Cohort. *Blood.* 2008;111(2):544-548.

13. Miller ST, Sleeper LA, Pegelow CH, et al. Prediction of Adverse Outcomes in Children with Sickle Cell Disease. *New England Journal of Medicine.* 2000;342(2):83-89.

14. Powars DR. Sickle Cell Anemia and Major Organ Failure. *Hemoglobin.* 1990;14(6):573-598.

15. van Beers EJ, van Tuijn CF, Mac Gillavry MR, van der Giessen A, Schnog JJ, Biemond BJ. Sickle cell disease-related organ damage occurs irrespective of pain rate: implications for clinical practice. *Haematologica.* 2008;93(5):757-760.

16. Gladwin MT. Cardiovascular complications in patients with sickle cell disease. *Hematology Am Soc Hematol Educ Program.* 2017;2017(1):423-430.

17. Saraf SL, Zhang X, Kanias T, et al. Haemoglobinuria is associated with chronic kidney disease and its progression in patients with sickle cell anaemia. *Br J Haematol.* 2014;164(5):729-739.

18. Kinney TR, Sleeper LA, Wang WC, et al. Silent Cerebral Infarcts in Sickle Cell Anemia: A Risk Factor Analysis. *Pediatrics.* 1999;103(3):640-645.

19. Administration UFD. FDA approves voxelotor for sickle cell disease. <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-voxelotor-sickle-cell-disease>. Published 2019. Accessed.

20. Administration UFD. FDA approves crizanlizumab-tmca for sickle cell disease. 2019.

21. Administration UFD. FDA approved L-glutamine powder for the treatment of sickle cell disease. Published 2017. Accessed.

22. Ribeil J-A, Hacein-Bey-Abina S, Payen E, et al. Gene Therapy in a Patient with Sickle Cell Disease. *New England Journal of Medicine.* 2017;376(9):848-855.

23. Michalik DE, Taylor BW, Panepinto JA. Identification and Validation of a Sickle Cell Disease Cohort Within Electronic Health Records. *Acad Pediatr.* 2017;17(3):283-287.

24. Singh A, Mora J, Panepinto JA. Identification of patients with hemoglobin SS/Sβ0 thalassemia disease and pain crises within electronic health records. *Blood Advances.* 2018;2(11):1172-1179.

25. TriNetX. <https://www.trinetx.com/>. Accessed2020.

26. Novelli EM, Hildesheim M, Rosano C, et al. Elevated Pulse Pressure is Associated with Hemolysis, Proteinuria and Chronic Kidney Disease in Sickle Cell Disease. *PLOS ONE.* 2014;9(12):e114309.

27. VAN HOUWELINGEN HC. Dynamic Prediction by Landmarking in Event History Analysis. *Scandinavian Journal of Statistics.* 2007;34(1):70-85.

28. van Houwelingen HC, Putter H. Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. *Lifetime Data Anal.* 2008;14(4):447-463.