Sickle Cell Disease Patient Registry (SCDIC-II) Protocol

Sponsor

National Heart, Lung, and Blood Institute

June 2022
Version 1.0

TABLE OF CONTENTS

APPENDICES (located in separate files)

1. BACKGROUND AND RATIONALE

1.1 Sickle Cell Disease

Sickle cell anemia (SCA) refers to the clinically similar disorders Hb SS or Hb $S\beta^0$-thalassemia. Sickle cell disease (SCD) refers to all disease genotypes, including SCA and compound heterozygous disorders, such as Hb SC, Hb $S\beta^{-II}$-thalassemia, and other less common variants. The carrier state for hemoglobin S (Hb AS or sickle cell trait) is not a form of SCD.

Hb A, the predominant hemoglobin after the newborn period, consists of two α and two β globin chains. Hemoglobin with two mutated β^s-globin chains is designated Hb S, sickle hemoglobin. The predominant hemoglobin during intrauterine development, Hb F (fetal hemoglobin), contains two γ-globin chains rather than two β-globin chains. Individuals with SCD-SS produce predominantly Hb S with some Hb F, and those with other forms of SCD produce both Hb S and another hemoglobin. When the Hb S is deoxygenated, it can aggregate into stiff rod-like structures. These structures make the RBCs less deformable and result in abnormally shaped RBCs that rupture easily, stick to other blood cells and vascular endothelial cells, occlude blood vessels, and interrupt blood flow in the tissues. These characteristics increase hemolysis and cause vaso-occlusion, which produce clinical symptoms of SCD.

The most common complications of SCD result from damaged blood vessels and entrapment of the sickled cells and other blood cells. Ischemia, tissue damage and hemolysis cause much of the morbidity and mortality associated with SCD, including episodic pain, splenic dysfunction and associated vulnerability to bacterial infections, strokes, renal dysfunction, acute and chronic lung injury, retinal damage, and skin ulcerations. The size of the U.S. SCD population is unknown but is estimated to be as high as 122,900 of which 99,888 (81%) have Hb SS. Minorities are disproportionately affected, occurring in roughly one of every 500 African American births and one of every 36,000 Hispanic births.

1.1.1 Natural History

Fifty years ago, it was rare for individuals with SCD to live beyond childhood. Advances in care delivery and treatment have more than doubled the life expectancy of individuals with sickle cell disease (SCD) since 1972. Beginning in the 1970s, measures such as newborn screening, prophylactic administration of penicillin and immunization against bacterial infections decreased complications and morbidity, increasing the length and quality of life of children with SCD. Blood transfusions are the only proven way to prevent some of the major complications of SCD, especially recurrent stroke. They are also used frequently to help manage some of the acute complications of SCD. More recently, the use of hydroxyurea as a therapeutic agent to increase fetal hemoglobin has been shown to further reduce the debilitating symptoms of and improve survival in SCD. L-glutamine and Crizanlizumab are additional treatments to reduce pain crises and Voxelotor is approved to lower the risk of anemia and improve blood flow. In the absence of a widely accessible cure, treatment for SCD is usually aimed at avoiding crises, relieving symptoms, and preventing complications. Many of the advances in treatment have not translated into an increase in longevity or quality of life for adolescents and adults because of disparities in access to routine primary health care.

Individuals with SCD experience a markedly increased mortality beginning in the second decade of life. The third and later decades of life are frequently associated with severe chronic pain, progressive organ damage and frequent hospitalizations. The provision of evidence-based and expert opinion-based care in SCD is complicated by the difficulties that many patients experience in obtaining access to the health care system and in receiving long-term care from knowledgeable providers.

1.1.2 Curative Therapies for Sickle Cell Disease

Bone marrow or stem cell transplantation is the only approved cure for SCD (Leonard et al., 2020). However, because of the risks associated with a bone marrow transplant, including death, the procedure is recommended

only for people, usually children, who have significant symptoms and complications of sickle cell anemia. Best outcomes are achieved in those with matched related donors which are available for a minority of the eligible population. Some patients and providers perceive the risks of the transplant-related mortality and morbidity to outweigh the potential for a cure.

New advancements are also being made in gene therapy as a promising cure for sickle cell disease. The NHLBI initiated the ***Cure Sickle Cell Initiative (CureSCi)*** in 2018 as a collaborative research effort designed to accelerate genetic therapies in SCD and move them safely into clinical trials. Phase I and Phase II trials are underway. As trials progress, a source of data from a control population is needed for comparison of quality of life and other outcomes in SCD patients that did not undergo gene therapy or bone marrow transplant. These comparison groups could come from contemporaneous natural history studies of patients with SCD.

One concern with bone marrow transplant and gene therapy in patients with SCD is the risk of clonal hematopoiesis and development of acute myeloid leukemia (AML), myelodysplastic syndrome (MDS), and other blood malignancies related to the conditioning treatment (Ghannam et al., 2020).

1.1.3 Blood Cancers and Sickle Cell Disease

Individuals with SCD are known to have an increased risk of hematologic malignancies compared to the general population without SCD but of similar age, sex, race, and ethnicity. In a large SCD cohort in California, the risk of leukemia was more than 2-fold higher than the comparable general population, and AML had a 3.6-fold increase in incidence (Brunson et al., 2017).

A possible cause for the increased risk of blood cancers with SCD includes the use of hydroxyurea, although there is little evidence to support this hypothesis. Another cause is the general susceptibility of SCD patients based on common complications and their treatment. SCD is accompanied by chronic inflammation of the blood vessels, increased iron from blood transfusions, increased risk of infections and cellular turnover in the bone marrow. These risk factors have been shown to play a role in cancer development, including leukemia.

1.1.4 SCDIC-I Registry

National Heart, Lung, and Blood Institute funding for the SCDIC-I Registry has supported RTI International and 8 clinical centers across the United States since 2016. Over 2400 people who have sickle cell disease and were between the ages of 15-45 at the time of enrollment have been enrolled into the Registry. A rich resource of natural history data, the SCDIC-I Registry has longitudinal data from patient surveys (e.g. pain sleep, barriers to care, pregnancy), medical record abstraction (e.g. medications, transfusion history, comorbidities) and laboratory results. The SCDIC-II Registry will continue follow-up of this patient cohort and enroll new patients of all ages to enrich the data resource as a natural history study. All previously collected data collected under the SCDIC-I consents and protocol will transition to SCDIC-II for continued use by the researchers. A complete list of publications from the SCDIC-I Registry can be found here: https://scdic.rti.org/RESEARCH-DISSEMINATION/Publications.

2. REGISTRY OBJECTIVES

2.1 Objectives

The goal of the SCDIC-II Registry is to continue the longitudinal follow-up of the eligible patients that were previously enrolled in the SCDIC-I Registry and to enhance the cohort with additional patients to better understand the natural history of SCD.

To achieve these goals, the Registry will:

- continue collecting longitudinal data that will inform the natural history of sickle cell disease.
- evaluate the collection of standardized core data that can be combined across other SCD studies to form a Natural History Data Resource (NHDR) of de-identified data for use by qualified researchers.
- provide a cohort of well-characterized patients to serve as comparator groups for gene therapy or other clinical trials or comparative effectiveness studies.
- conduct pilot studies that inform the collection of common data elements in SCD, including extraction of structured data from the electronic health record.

2.2 Timeline

The overall timeline for the Registry is shown below. The protocol, patient facing data collection forms and other necessary materials will be submitted to the sIRB. Once sIRB approval has been obtained, each site will enroll subjects, the number of which will vary by site as it is based on eligibility for continuation from the SCDIC-I Registry. Enrollment of SCDIC-I subjects will be completed by June 2023; additional people may be enrolled after this time. Annual follow-up will begin in July 2023. At each phase of data collection, the Registry will collect patient reported outcomes and data abstracted from medical records.

	2022		**2023**				**2024**				**2025**			
SCDIC-II Registry	**Jul-Sep**	**Oct-Dec**	**Jan-Mar**	**Apr-Jun**	**Jul-Sep**	**Oct-Dec**	**Jan-Mar**	**Apr-Jun**	**Jul-Sep**	**Oct-Dec**	**Jan-Mar**	**Apr-Jun**	**Jul-Sep**	**Oct-Dec**
sIRB approval	X													
Reconsent, Year 1 survey completed, record abstraction	X	X	X	X										
Deposit Year 1 data into data repository						X								
Year 2 follow-up					X	X	X	X						
Deposit Year 2 data into data repository										X				
Year 3 follow-up									X	X	X	X		
Deposit Year 3 data into data repository														X
OSMB Meetings				X				X				X		
Data edits and QC	X	X	X	X	X	X	X	X	X	X	X	X	X	

3. REGISTRY ORGANIZATION

3.1 Overview

The following sections describe the organizations that are participating in the SCDIC-II Registry including the study sponsor, data coordinating center, and the eight SCDIC-II Clinical Centers.

3.2 Organization and Participating Institutions

Sponsor

Following the Sickle Cell Disease Implementation Consortium (SCDIC), which was funded from 2016-2022, the SCDIC-II was established in 2022. Both studies are funded by the National Heart, Lung, and Blood Institute of the National Institutes of Health (NIH). The NHLBI's Division of Blood Diseases and Resources (DBDR) is responsible for organizing the overall program, providing oversight to the various components of the Consortium, overall monitoring of interim data and safety issues, and collaborating with the Data Coordinating Center (DCC) and the SCDIC investigators in the development and implementation of the Registry.

Data Coordinating Center

RTI International serves as the DCC for the SCDIC-II Registry. As the DCC, RTI is responsible for data capture and management, coordination of training, logistics and communications, study website, data quality, reporting, and statistical analysis. The DCC is responsible for development of the protocol, data collection forms, manual of operations, initial and annual submissions to the sIRB and preparation of progress reports to the Observational Study Monitoring Board (OSMB) and the NHLBI.

Clinical Centers

The eight Clinical Centers participating in the SCDIC-II Registry represent academic and medical institutions and research teams that are multi-disciplinary in sickle cell disease, pediatrics and adult care, hematology, and community-based research. The leadership at each Clinical Center is provided by the Principal Investigator (PI), who assumes responsibility for all aspects of the Center's participation in the Registry. Each Center provides care to patients from a wide geographic catchment area.

1. Augusta University
2. Duke University
3. Icahn School of Medicine at Mount Sinai
4. Medical University of South Carolina
5. St. Jude Children's Research Hospital
6. UCSF Benioff Children's Hospital Oakland
7. University of Illinois at Chicago and Sinai Health System
8. Washington University School of Medicine

Observational Study Monitoring Board

An NHLBI-appointed OSMB, comprised of a multi-disciplinary group of experts, will perform independent monitoring of overall progress of data collection and patient safety for the SCDIC-II Registry. The OSMB makes recommendations to NHLBI regarding appropriate protocol and operational changes. Any decision to modify the protocol or significantly change study operations may have a substantial effect upon the study. Thus, the OSMB plays an essential role in assuring quality research.

3.4 Project Website

The SCDIC-II website, managed by the DCC, serves as a source of information for the site research teams and the research public. It provides both historical information from the first Registry as well as content related to the SCDIC-II.

Public Website. The public side of the study website, http://scdic.rti.org, will be used most notably to disseminate information about the Registry to the research public. The public pages include 1) project description and rationale, 2) participating institutions and contact information, 3) links to project news, and (4) a list of publications and sources for accessing the data.

Private Website. The restricted area of the study website provides a secure channel for SCDIC-II study staff to access project materials (e.g. meeting agendas, training manuals), ensuring that all research staff have one place to access SCDIC-II information. The website is secure and requires a username and password to enter. Access is granted only to those whose role on the project warrants it.

4. ELIGIBILITY, RECRUITMENT AND INFORMED CONSENT

4.1 Inclusion and Exclusion Criteria

The SCDIC-II Registry will include as many as 2,200 patients who meet inclusion criteria across the eight participating sites. The numbers of patients enrolled at each site will vary based on eligibility for continuation from the SCDIC-I Registry and number of eligible adolescents. Women and children will be included in the SCDIC-II Registry to the extent that they exist in the population being studied and meet eligibility criteria set forth for study participation. In addition to parental/guardian consent, assent will be obtained from children 14-17 years of age.

Inclusion Criteria

- Previously enrolled in the SCDIC-I Registry or with permission from the DCC
- At least 14 years of age
- English speaking
- Confirmed SCD diagnosis. Confirmed is defined as supported by documentation in the medical record of a positive test for one of the following: Hb SS, Hb SC, Hb Sβ-thalassemia, Hb SO, Hb SD, Hb SG, Hb SE, Hb SF
- Available medical record for abstraction of clinical data (e.g. treatment, co-morbidities and complications, laboratory tests, imaging)
- Willing and cognitively able to give informed consent and complete the Patient Survey

Exclusion Criteria

- Unwilling or unable to give consent/assent or complete the Patient Survey
- Sickle cell trait (i.e., Hb AS)
- Successful bone marrow transplant
- No available medical record that documents SCD care

4.2 Recruitment

Recruitment for the SCDIC-II Registry will occur at each participating Clinical Center and consist of recontacting eligible patients that were enrolled in the SCDIC-I Registry. Enrollment of new patients (mostly adolescents) not in SCDIC-I may occur at the request of the DCC.

A recruitment brochure has been created that targets patients that were enrolled in the SCDIC-I Registry (**Appendix A**). A small size version of this brochure (approximately 7”x5”) may be given to patients or made available to them to take in the clinic setting. A larger size will be available to the Clinical Centers for posting on a clinic or waiting room bulletin board.

SCDIC-II clinic staff will identify eligible patients using the eligibility criteria and the list of patients from SCDIC-I that are not Off Study. Eligible patients may be approached during clinic or inpatient visits, by phone, or through secure messaging as permitted by their institution. This flexibility on the part of the SCDIC-II will ensure inclusion of the greatest number of eligible patients for the Registry.

The SCDIC-II Centers will enroll up to 2,200 eligible participants who meet inclusion criteria for the Registry. A participant will be considered enrolled when consent is obtained, and inclusion criteria have been confirmed. The SCDIC-II enrolling Center will maintain a local log of consented participants and will also confirm enrollment status in the data management system (DMS) under the appropriate Subject ID.

Screening, approaching, consenting and verifying eligibility of potential study participants will be conducted by designated study staff. If the patient (and parent/guardian of minors) agrees, the designated study staff will meet with the patient for a more comprehensive explanation of the Registry and the requirements for participation. If there is continued agreement, the designated study staff will proceed with the consent and enrollment process using the approved site-specific consent form. A template of the consent form for use by the sites is shown in **Appendix B**.

4.3 Consenting and Enrollment

The process of obtaining signed informed consent will occur only in person or by phone/Zoom/Teams. When consenting is not done in person, the blank consent form will be sent to the patient by postal mail or email. If sent by email, the coordinator will confirm that the patient has access to a printer. Whether in person or by phone/Zoom/Teams, the patient will be given ample time to read the consent, or have it read to them, as well as ask questions at any time. If consent is done in person, the coordinator will make a copy of the signed consent and give it to the patient. If consent is obtained by phone/Zoom/Teams, the patient will either mail the signed consent form back to the coordinator or take a photo/scan of the signature page and send it back to the coordinator electronically (via portal or secure email).

Literacy in English will be required of the consenting patient or parent/guardian. Adolescent assent will be obtained from children 14-17 years of age. A signed HIPAA Research Authorization is also required of all participants and is incorporated into the template consent form.

Centers will provide Registry participants with a monetary incentive for their time and participation in the Registry. The appropriate amount and method of payment is determined by each SCDIC-II Clinical Center.

4.4 Human Subjects and Informed Consent

No data collection activities will begin at an individual SCDIC-II participating Clinical Center until approvals from RTI and the sIRB have been granted for the Center. RTI will submit the protocol, template consents, patient facing materials and data collection forms to the sIRB for review and approval. The sIRB will then allow each Center to add site specific language to the consent for further review and approval. Once the sIRB has approved the site-specific consent, they will notify the Center that they are approved to start consenting patients. All participating SCDIC Clinical Centers and RTI have a Federal Wide Assurance issued through the US Office of Human Research Protections which assures that the organizations are complying with all federal regulations to protect research subjects.

<u>Confidentiality of Data and Use of Identifying Information</u>

Full names and other identifying information, excluding date of birth, will be retained only by the Centers. Participants' data will be labeled and stored with coded identification numbers that can be linked to names only by the corresponding Center. All collected data will be kept confidential to the extent permitted by law. The DCC will not be able to link an individual to their identifying information. Protected Health Information (PHI) will be collected and transmitted to the DCC (RTI). In addition, RTI maintains the data from the SCDIC-I Registry which also contains PHI and will be used in conjunction with the SCDIC-II data. All data are stored in a secure protected environment. All SCDIC data shared with researchers outside the study will be de-identified and will not contain any PHI. Limited datasets may be shared with researchers within the Consortium under a Data Use Agreement.

<u>Risks and Benefits</u>

The data collected for this Registry are from medical record abstraction and self-reported information. The patient surveys are not considered greater than minimal risk but may trigger uncomfortable feelings about

one's lifestyle, quality of life or personal history of disease. There are no direct benefits to the participants in this Registry. Some patients may benefit from knowing that they are helping to advance knowledge for future patients with their condition.

Unbiased Recruitment

All eligible participants will be recruited without bias. Adolescents, women and minorities will be included as they represent the patient population of each Center.

Inclusion of Children

Children aged 14-17 will be included in the SCDIC-II Registry to the extent that they exist in the population being studied and meet eligibility criteria set forth for study participation. In addition to parental/guardian consent, adolescent assent will be obtained from children 14-17 years of age by them signing a separate line on the same consent the parent/guardian signs. When the children turn 18 years of age during their participation in the Registry, they will be required to consent as an adult before any additional new data are collected.

Inclusion of Women and Minorities

This Registry's selection criteria include all individuals that fit the inclusion criteria, without regard to sex, race, or ethnic group. We recognize the importance of having all eligible persons participate in this study, and we are particularly conscious of the need to include women and minorities in the study population. Sickle cell disease occurs most often in people with African ancestry or identify themselves as Black, as well as people who come from Hispanic, southern European, Middle Eastern, or Asian Indian backgrounds. We expect the demographics of enrolled subjects to reflect the patient population seen by each Center. The DCC will monitor data collection to assure recruitment of women and minorities who meet eligibility criteria.

Rights of Refusal and Withdrawal

Patients will be free to refuse enrollment or withdraw from the study at any time. Participants may refuse to answer individual questions on the self-administered questionnaires.

5. DATA COLLECTION

5.1 Overview

Patients enrolled in the SCDIC-I Registry were assigned a unique identification number which was used to identify them throughout the first Registry. This number will carry forward for those patients that reconsent into the SCDIC-II Registry. Patients new to the SCDIC-II Registry will be assigned a unique site specific number starting after the last number assigned under SCIDC-I. The identification number will be assigned by the designated study staff after a patient provides consent. The link between the patient name and the associated identification number will remain at the enrolling Center under restricted access. Required forms for all enrolled participants to be completed annually include the Patient Survey, the Medical Record Abstraction Form, and the results from the most recent imaging and laboratory tests.

5.2 New Data Collection

After consent or assent is obtained, data collection can begin. The Patient Registration form is completed only at Enrollment. The other forms are completed at Enrollment and annual Follow-up.

Forms	Enrollment of SCDIC-I patients	Enrollment of new patients	Annual follow-up
Patient Registration	X	X	
Patient Survey	X	X	X
Enrollment Medical Record Abstraction		X	
Follow-up Medical Record Abstraction	X		X
Laboratory Reporting	X	X	X
Procedures Performed	X	X	X
Cardiac Procedures Results	X	X	X

5.2.1 Patient Registration Form

The Patient Registration Form (**Appendix C**) is completed by the study coordinator after signed informed consent or assent has been obtained and the eligibility criteria for enrollment has been confirmed. This form provides the date of consent and responses to the tiered components of the consent form that may have been added by the site (e.g. agreement to be contacted for future research studies). It is the first form entered in the DMS for all newly consented subjects and creates a record of all subjects in the SCDIC-II Registry.

5.2.2 Patient Survey

The self-administered Patient Survey (**Appendix D**) is to be completed by the participant, either as self-administered or interviewer administered. It provides information about the patient's health at the time of enrollment. Completion of this form will provide demographic and diagnosis information, pain experience, social and mental health information, other patient reported outcome domains of interest (e.g., sleep, fatigue), and alcohol and smoking history. Patients may refuse to answer some of the questions. The survey may be taken home for completion, but every effort should be made to have the patient complete it in the clinic/hospital setting. For those patients that do not attend routine in-person visits, a secure link to the survey may be sent by the site coordinator through their REDCap system. This allows the patient to complete the survey online and outside of the clinic setting.

The Patient Survey will not cause severe psychological distress, but some already fragile participants may become upset while thinking about or answering some of the questions. The instructions at the beginning of the survey include a statement about contacting the study coordinator if the patient feels a need to talk.

If the participant contacts the coordinator because of distress, they will be offered a list of local mental health resources. The participant may also be reminded that s/he can follow up to discuss any concerns with the sickle cell provider or team psychologist

5.2.3 Medical Record Abstraction Form

The Medical Record Abstraction Form (**Appendix E**) will be completed by research staff for all participants using the patient's electronic health record. Completion of this form will provide information about curative therapies, anthropometric measurements, medication usage, organ systems review, co-morbidities and complications, and treatment. If a new diagnosis of chronic kidney disease or end stage renal disease is indicated on the form, the coordinator will also complete the **Renal Form** (**Appendix F**). At enrollment, this form will look as far back as when the last abstraction was completed under SCDIC-I or when the SCD diagnosis occurred (for new patients). At follow-up, it will look only as far back as when the last Medical Record Abstraction Form was completed.

5.2.4 Laboratory Reporting Form

The Laboratory Reporting Form (**Appendix G**) will be completed by research staff at enrollment and follow-up for all laboratory test results that were completed in the previous 12 months.

5.2.5 Procedures Form

The Procedures Form (**Appendix H**) is completed at enrollment and follow-up to obtain information about specific procedures that were done such as liver biopsy, cardiac or brain MRI, and other SCD-related procedures. At enrollment, the look-back period will include the patient's entire medical history. At follow-up, it will look only as far back as when the last Procedures Form was completed.

5.2.6 Cardiac Procedures Form

The Cardiac Procedures Form (**Appendix I**) is completed at enrollment and follow-up to document the results of any cardiac ECHO, EKG or heart catheterization procedures that were identified as having been done on the corresponding Procedures Form. It is completed using the patient's medical records.

5.2.7 Off Study Form

The Off Study Form (**Appendix J**) is completed whenever a patient is no longer available for follow-up. Reasons include death, withdrawal, lost-to-followup, and ineligibility after enrollment such as when the patient has undergone a curative therapy.

5.3 Coordinator Training

To ensure consistent and standardized data collection across the participating Centers, the DCC will develop and implement a training program for the research coordinators based on the training manual and manual of operations. This training will be web-based and will review the detailed instructions on subject enrollment, study procedures, data collection, and data management procedures. Specifically, the training will include:

- Purpose of study
- Enrollment criteria (inclusion/exclusion)
- Strategies for outreach and enrollment
- Abstraction of information from medical records
- Dealing with distressed participants
- Question by question specifications
- Review of the data entry and data management systems
- Privacy and data security policies
- Communications with the DCC
- Data entry and access to monitoring reports

If there is attrition, the DCC will train new study coordinators. In addition to training, each site PI is responsible for providing supervision and oversight of the local Registry coordinators and assuring quality control of data collected including review and sign-off of record abstractions and any reabstraction, as needed.

5.4 Extant Data from the SCDIC-I Registry

All data previously collected on study participants enrolled in the SCDIC-I Registry will transition to the SCDIC-II Registry. These data are stored at the DCC and include patient surveys and clinical data from the medical record. PHI from SCDIC-I includes zip code, date of birth, and dates of health events from the medical record. SCDIC-I data from participants that consent to SCDIC-II will be linked under the same Subject ID. The SCDIC-II consent also includes a statement about the linking of data from SCDIC-I to the SCDIC-II for patients that have these extant data. These data will be incorporated into the SCDIC-II data resource for use by Registry investigators for sickle cell related analyses.

6. DATA MANAGEMENT

6.1 Overview

There are complexities in handling data from multiple centers. Data management and quality control systems will be implemented to ensure timely receipt of clinic data and resolutions of discrepancies and duplications.

6.2 Data Management System

Data will be entered into a password protected, secure web-based data management system (DMS) using RTI's Research Electronic Data Capture (REDCap). The REDCap website is protected by a login and password and Transport Layer Security (TLS) technology. This technology scrambles the data as it is being keyed and is the industry standard for protecting personal information during transfer. Within the DMS, the DCC will build in edit, range and validity checks on the data as they are being entered. In addition to data entry, the REDCap DMS will allow SCDIC-II and Center staff to produce data management reports to monitor their performance and to view the data entered on patients. The DCC will train Center staff in data collection and management in accordance with the protocol and manual of operations.

6.3 Data Edits

Quality control checks will be programmed into the web-based REDCap data entry system developed for collection of data from Centers. Checks for internal consistency with respect to dates, acceptable ranges, required items, and skip patterns will be set up as validation at the time of data entry. In batch, the data will undergo additional automated, electronic edits that could not efficiently be included in the data entry screens, such as cross-form editing. Edit checks that are performed in real time will be replicated by the DCC in batch. In addition, complex within-form and across-form consistency and logic checks will be applied at the DCC. These checks will be based upon the specifications appropriate to each of the data collection forms and will not include missing data on the Patient Survey. Any failures will be reported to the Center as error resolution reports. Center personnel will enter the corrections for the keyed data into the electronic file, and an audit trail of corrections will be maintained. The error resolution will be done online via the REDCap data entry system.

6.4 Monitoring Reports

To monitor enrollment, data flow, delinquent data, and data quality, the DCC project team will run reports that monitor the performance of the individual centers. These reports will be distributed and reviewed regularly by the Center leadership and staff. Reports will show the number of patients enrolled at each Center, as well as the number and type of forms received, the number of overdue forms, and the number of unresolved data edits. The DCC will collaborate with the Center coordinators and investigators to design reports that are meaningful and assist in monitoring the conduct of the study and producing high quality data for analysis.

6.5 Data Security

Every effort has been made to maximize the security of the computer systems used by SCDIC. The Data Management System (DMS) is managed by RTI's Clinical Research Informatics group, combining clinical informatics, data management, and system security expertise.

The DMS system can only be accessed by secure web connection. Users require a username and password to log onto the DMS. Each user has a unique account for data entry. This account will only allow access to areas of the DMS necessary to perform data entry and query study data related to that site. The center coordinator or designee is the only individual authorized to request user accounts. Authorized requests will then be filled by designated RTI staff. Any direct, administrative access to the system and database is both limited by location (IP address) and to specific computers using cryptographic key pairs. The REDCap system used for secure data capture has an option for sites to be able to export their own data that they entered in to the DMS. All hard-copy forms will remain at the Centers; RTI will only receive data electronically through the DMS.

7. STATISTICAL ANALYSIS

7.1 Overview

The SCDIC-II Registry cohort will consist of approximately 1,800 patients with confirmed sickle cell disease. In general, the analyses of the Registry data will be descriptive and exploratory in nature and are for hypothesis-generating rather than hypothesis testing. With a sample size of 1,800 there is sufficient power to detect differences among treatment or outcome groups. In addition to analyses that may be conducted by the SCDIC-II investigators, de-identified Registry data will be placed in NHLBI data repositories for access by other researchers.

7.2 Primary Outcomes

This is a natural history study and there will be several outcomes of interest for the Registry depending on the analysis and the interest of the investigator. At this time, there are no specific research questions, but our approach to future analyses that would be done by the DCC are presented in the following sections.

7.3 Approach to Analysis

Patients enrolled in the Registry will be characterized by their demographics (e.g. age, gender), clinical conditions, treatment procedures, and self-reported health status. Categorical or ordinal variables such as conditions and treatments will be summarized by frequency distribution. Continuous variables such as blood chemistry results will be summarized by mean, median, standard deviation, minimum and maximum value. Outliers and possible data errors will be detected for further formal statistical analysis.

Regression methods may be used to examine the relationship between specific factors and an outcome of interest while controlling for potential confounding factors. According to the nature or measurement of outcome variables, linear regression, logistic regression, multinomial logistic regression, or time-dependent covariate analysis, such as Cox proportional hazards will be performed. Random effects models may also be used because the Registry is conducted through multiple Clinical Centers. Outcome variables such as disease severity or use of hydroxyurea may vary by site because of recruitment procedures or physician practice. Ignoring this effect in the regression analysis may result in underestimated standard errors of parameter estimates.

There is a likelihood of missing data which can seriously affect the results. Ignoring missing data, or assuming that excluding missing data is sufficient, will increase the risk of reaching invalid and insignificant results. Also, missing data may reduce the precision of calculated statistics because there is less information than originally planned. Another concern is that the assumptions behind many statistical procedures are based on complete cases, and missing values can complicate the theory required. We will review mean, standard deviation, frequencies, number of missing and non-missing values, number of extreme values for all variables to understand the missing data. We will conduct missing value analysis to find if the cases with missing values are systematically different from cases without missing values. Based on our initial missing value analysis, we will employ appropriate ad-hoc imputation methods (such as, last observation carried forward, minimum value replacement, maximum value replacement, estimates based on different regressing models) or multiple imputations to impute the missing values.

7.4 Statistical Methods

Statistical analysis of the Registry data will require a variety of methods. Different statistical techniques will be employed to answer specific research questions. The selection of statistical methods will be based on the research question and the type of outcome variables used (categorical or continuous). The following are a few examples of statistical methods we will use to explore the specific type of scientific relationship between variables.

Descriptive and Bivariate Analyses

Descriptive analysis will be performed to examine the distribution of data and detect possible error. To explore the crude association of an outcome variable with a single factor we will use bivariate analysis. Selection of statistical methods will depend on the measurements of two variables. Cross-tabulation will be used when both variables are categorical. Chi-square, or Fisher's exact test in the case of sparse cells, will be performed to test the significance of the associations. T-test and analysis of variance (ANOVA) will be used to compare continuous variables such as age, number of transfusions and lab values among different outcomes. Simple correlation coefficient may be used to describe the association between two continuous or interval variables such as age and pain severity.

Regression Analysis

To examine the relationship between a continuous outcome of interest and a group of factors that may be significantly associated with that outcome variable, multiple linear regressions will be used. Multiple regression will also be used to test the relationship between outcome variables while controlling for potential confounding factors. Logistic regression will be used to analyze binary outcomes, such as whether an individual patient was using hydroxyurea and whether the patient experienced certain comorbidities. Logistic regression will be used to analyze those binary outcomes in relation to predictors such as insurance status, gender, barriers to care, age, pain severity and measures of quality of life. Multinomial logistic regression will be used to analyze categorical outcomes such as type of stroke and transfusion.

Time to Event Analysis

To analyze the survival type data, also known as time to event(s) data, we will use Kaplan-Meier survival analysis. To examine the relative risk for specific factors while adjusting for other covariates and potential confounding factors we will use Cox-proportional hazard model. If the proportional hazard assumption isn't valid for the Cox model, we will use Exponential regression and Weibull regression. If the survival or failure data have multiple failure events, then we will use Multiple decrement life table analysis to calculate the cumulative incidence rate.

Other advance statistical analysis methods such as Mixed models, GEE, competing risk survival model, factor analysis, principal component analysis, and discriminant analysis may be also employed to answer scientific questions utilizing the Registry data. Analyses will mainly be implemented using SAS and other statistical packages such as R, Stata, and MLwin.

8. DATA USE AND DISSEMINATION

8.1 Overview

Sharing of research data expedites the translation of research results into knowledge and procedures to improve human health. Data from well-characterized population samples constitute an important scientific resource. It is the view of the NIH that their full value can only be realized if they are made available, under appropriate terms and conditions consistent with the informed consent provided by individual participants, in a timely manner to the largest possible number of qualified investigators. Under funding from the NHLBI's CureSC Initiative, data will be de-identified per NIH/NHLBI guidelines and released approximately annually, after each phase is complete.

8.2 Data Sharing Policy

The primary goal of the SCDIC-II Registry is to establish a data repository that will be shared with Registry investigators and qualified researchers outside the Registry interested in studying additional aspects of SCD that are not being addressed by this protocol. De-identified patient-level data will be made available to researchers outside the Registry through two NHLBI data repositories – BioLINCC and Biodata Catalyst. Both repositories require an application and approval process facilitated by the NHLBI. To protect the privacy and confidentiality of the study subjects, investigators granted access to the de-identified data must adhere to strict requirements incorporated into a standard Data Use Agreement. In accordance with NHLBI policy, outside researchers will also be required to submit an approval from their Institutional Review Board (IRB).

8.3 Data Sharing with SCDIC Researchers

SCDIC Registry Researchers (i.e., researchers inside the Consortium) may request a limited dataset from RTI for analysis. Prior to secure data transfer, a Data Use Agreement must be in place.

9.0 REFERENCES

Leonard A, Tisdale J, Abraham A. Curative options for sickle cell disease: haploidentical stem cell transplantation or gene therapy? Br J Haematol. 2020 May;189(3):408-423. https://doi.org/10.1111/bjh.16437.

Ghannam JY, Xu X, Maric I, Dillon L, Li Y, Hsieh MM, Hourigan CS, Fitzhugh CD. Baseline TP53 mutations in adults with SCD developing myeloid malignancy following hematopoietic cell transplantation. Blood. 2020 Apr 2;135(14):1185-1188. https://doi.org/10.1182/blood.2019004001.

Brunson, A., Keegan, T., Bang, H., Mahajan, A., Paulukonis, S., & Wun, T. (2017). Increased risk of leukemia among sickle cell disease patients in California. Blood. 2017;130(13):1597–1599. https://doi.org/10.1182/blood-2017-05-783233.