i. Protocol overview

Title: REAL Answers (Registry Expansion Analyses to Learn) HL 167036

Sponsor(s): National Heart, Lung, and Blood Institute, National Institutes of Health

National Institutes of Health (NIH)
9000 Rockville Pike, Bethesda, MD, 20892

National Heart, Lung, and Blood Institute (NHLBI)
31 Center Drive, Bethesda, MD, 20892

Principal Investigator: Jeffrey Glassberg MD MA

Physical Address: 1 Gustave L. Levy Place, Box 1620, New York, New York 10029

Phone: 212-824-8056

ii. Purpose of the study and background

Purpose of the study: This study has three specific aims

AIM 1: Compare the effect of new SCD medications – crizanlizumab, voxelotor, and L-glutamine – on clinical outcomes in individuals with SCD.
AIM 2: Identify genetic and genomic predictors of response to crizanlizumab, voxelotor, and L-glutamine.
AIM 3: Integrate study data into the CureSCi metadata catalog (MDC) to enhance future cross-study analyses.

Background: Provide background material which supports the purpose of the research, and which is detailed enough to allow someone who is not an expert in the field to understand the context of the question, and the study design. References may be cited in the Background section.

Significance

Sickle Cell Disease (SCD) is a recessively inherited blood disorder that affects approximately 100,000 Americans and 3 million individuals worldwide.[1] There are 4 main genotypes of SCD, the most severe of which occurs in individuals who are homozygous for the hemoglobin S mutation (replacement of glutamic acid with valine at codon 6 of the β-globin gene). In SCD, hemoglobin forms long, rigid polymers upon deoxygenation, which impairs erythrocyte rheology, leads to vaso-occlusion and hemolysis. The disease manifests as chronic organ damage, recurrent vaso-occlusive pain, and premature death with a median life expectancy of 48 years.[2-4] Clinically, SCD is highly heterogeneous, and the 4 genotypes explain only a fraction of the clinical variation. Multiple genetic and epigenetic factors are known to contribute to variability in presentations, degree of organ injury, severity of symptoms, lifespan and response to treatment.[5]

List of Key Abbreviations	
SCD	Sickle Cell Disease
SCDIC	SCD Implementation Science Consortium
NT-proBNP	N-terminal pro brain natriuretic peptide
Urine ACR	Urine albumin to creatinine ratio
ASCQ-Me	Adult SCD Quality of Life Measurement System
scRNA-seq	Single cell RNA sequencing
WES	Whole exome sequencing
PBMC	Peripheral blood mononuclear cells
HbF	Fetal Hemoglobin
MPR	Medication possession ratio
EMR	Electronic Medical Record

New treatment options create knowledge gaps. As providers prescribe medications without comparative evidence, patients accumulate organ damage. The availability of 3 new medications to treat SCD has potential to reduce morbidity but is limited by a lack of comparative data regarding how new drugs affect outcomes (**Table 1**) and whether subgroups of patients benefit more from medications that target different mechanisms of disease. In the current state, providers incorporate a variety of non-evidence-based factors to their decision of which medication to start, including patient preference, insurance, and clinical gestalt. Our preliminary comparisons of the new agents suggest that their effects are not equal, and that some patients may benefit most from particular drugs. Crizanlizumab, a selectin inhibitor that prevents vaso-occlusion, provided the greatest protection from organ injury overall whereas voxelotor, an anti-sickling agent that reduces hemolysis and raises hemoglobin, demonstrated high variability in response (in both our data and published reports from others[6-9]), suggesting that small subgroups of patients derive most of the benefit. While providers are forced to guess which medication is best for their patient, our proposed study will fill these knowledge gaps and eliminate barriers to effective SCD treatment choices.

Table 1: FDA approved SCD medications

Medication	Year approved	Mechanism	Improves heart-lung injury?	Improves kidney injury?	Improves blood injury?	Improves symptoms-pain events?	Improves mortality?
Hydroxyurea	1998	HbF-multifactorial	Yes	Yes	Yes	Yes	Yes
L-glutamine	2017	Antioxidant	?	?	?	Yes	?
Crizanlizumab	2019	Selectin inhibition	?	Pending*	?	Yes	?
Voxelotor	2019	Anti-sickling	?	?	Yes	?	?

*The phase III STEADFAST trial will soon report the effect of crizanlizumab on kidney injury (urine ACR) (NCT04053764).

Target trial emulation with pseudo-randomization accelerates discovery. Hydroxyurea, a mainstay of SCD treatment, was initially approved in 1998, but its benefits were realized over the course of 30 years of clinical use. Without head-to-head comparisons, the comparative benefits of the 3 new SCD medications may take decades to uncover as observational research slowly accumulates. Target trial emulation, used widely in cancer epidemiology, involves the rigorous collection of observational data according to clinical trial standards and protocols. With carefully planned analyses, observational biases can be minimized, RCT-level data can be approximated, and the discovery timeline can be rapidly accelerated.[10-17]

Markers of heart-lung, kidney, and blood injury are outcomes that predict mortality benefit. A variety of outcomes have been used for phase III SCD trials including rate of acute pain episodes, patient-reported symptoms, and markers of organ injury. Rate of acute pain episodes is an important, patient-centered outcome which predicts mortality,[18] but is also problematic because not all patients have frequent pain; in fact, many patients at highest risk for death often have no pain at all.[19] Three widely-available, standard-care biomarkers are known to reflect mortality risk and response to treatment. While the effects of hydroxyurea on these organ injury markers are well-known, the effects of new medications on organ injury are unknown. This lack of evidence limits providers' ability to choose regimens that best improve organ function and survival.

Heart lung injury (NT-proBNP) – Cardiopulmonary complications of SCD, such as pulmonary hypertension and high-output cardiomyopathy, are common and a major cause of morbidity and mortality.[20,21] NT-proBNP reflects increased risk due to pulmonary hypertension, cardiomyopathy and other etiologies of myocardial injury and stress.[22] Elevations in NT-proBNP correlate linearly with the risk of death, and improvements with treatment reliably correlate to reductions in mortality risk. Treatment with hydroxyurea reduces NT-proBNP and relative-risk of death by 40%;[23]

Kidney injury (urine ACR [albumin to creatinine ratio]) – Renal injury is another organ-specific complication of SCD associated with premature mortality.[24,25] Urine ACR is the most sensitive, early clinical marker of ongoing kidney injury.[26] Even mild elevations in urine ACR can precede reduction in GFR and increased creatinine by decades, and are a strong predictor of future progression to renal failure and death.

Blood injury (hemolysis score) – Hemolysis is a signature manifestation of SCD, a predictor of premature death[27,28] and has been used as a phase III outcome,[6-9,29-36] The laboratory markers of hemolysis (bilirubin, reticulocyte percent, LDH, AST) are highly collinear.[19,32,37,38] We demonstrated that a combined score, created with principal components analysis, overcomes these issues and accurately predicts mortality and response to treatment.[28,37,39-41] The score has a mean of zero, with positive values indicating higher-than-average hemolytic burden.

Genetic and genomic analyses may allow providers to predict treatment failure before it occurs. In addition to the fact that genetics can explain a portion of the phenotypic variation in SCD, genes also explain variability of response to treatment with the well-established medication, hydroxyurea.[5,42-44] For voxelotor in particular (as preliminary data from our group and others indicate that benefits of this medication are driven by subgroups of 'responders'),[6-9,29] identification of genetic predictors of response will be highly beneficial. There is also potential to identify genetic predictors of response to L-glutamine, as it acts by mechanisms related to nitric oxide and arginine production, and arginine levels[45,46] vary widely among SCD patients driven by *eNOS* gene polymorphisms.[47,48] Especially because rare cell subpopulations play a key role SCD pathophysiology,[49-54] combining genomic data with single cell RNA sequencing (scRNA-seq) of peripheral blood mononuclear cells (PBMCs), neutrophils, and hematopoietic cells can provide key insights to the mechanisms that drive treatment response to tailor treatment to individual genotypic and phenotypic profiles.

Overall study design: The 3 aims will be accomplished within the context of a single study.

Study design: Prospective, target-trial-emulation cohort study of individuals with SCD. This is an observational study where, based on clinician discretion and local site preferences, patients may or may not receive treatment with a new SCD medication (crizanlizumab, L-glutamine or voxelotor) and patients may cycle on and off various regimens. While patients will not be randomly assigned treatment, they will be followed as though they are participating in a clinical trial with regular assessment of medication adherence and outcomes including clinical events, laboratory outcomes and patient-reported outcomes (PROs). All participants will be enrolled during year 1 of the award and followed for the 5-year duration. Study visits will occur at enrollment, before any SCD medication change and at 4-month intervals after medication changes.

iii. Criteria For Subject Selection

Number of subjects: State the total number of subjects expected to participate. For multi-center protocols, this should include both the overall total and the number of subjects to be enrolled at each site. Total subjects: up to 2400. 1200 will be enrolled in year 1. Additional subjects will be enrolled if subjects from year 1 drop out of the study. This will maintain the total enrollment at 1200 for the duration of the study.

Subjects per clinical site: 150-300. 150 will be enrolled in year 1. Additional subjects will be enrolled if subjects from year 1 drop out of the study. This will maintain the total enrollment at 150 per site for the duration of the study.

Gender of Subjects: Women and men are expected to participate equally in the study. The SCD population across the 8 SCDIC sites is approximately 56% female. There are no exclusions for child-bearing potential or pregnancy.

Age of Subjects: Age at enrollment will be recorded for all study participants. Children (under age 15) will be excluded from this study. The decision to exclude children was based on the following factors. 1) SCD medications are not all approved for use in children below the age of 15, 2) Issues of study design preclude direct applicability of hypotheses and/or interventions to both adults and children (specifically different stages of disease and age-related processes).

There is no older age limit for participation. As such, we expect the age distribution of participants in this study to accurately reflect the age distribution of adults living with SCD. The median age for individuals who participate in this study is expected to be 33.5 years. This is reflective of the adult SCD population across the 8 SCDIC sites.

Racial and Ethnic Origin:
This project focuses on individuals with SCD. Minority and disadvantaged individuals, specifically blacks, Latinos and the uninsured, are likely to be included however, no sex/gender or racial/ethnic groups will be excluded from participation if they meet the eligibility criteria.

Prevalence of individuals from the following ethnic categories is less than 1% thus the expected enrollment is zero for these groups (Asian 0.1%, Pacific Islander 0.1%, American Indian/Alaskan Native 0.1%). The prevalence of white, non-Latino individuals is 0.4%, thus we do not expect to enroll from this group.

I**nclusion Criteria:**

1. Individuals age 15 or older
2. Sickle cell disease (Hemoglobin SS, Sβthalassemia, SC, SHPFH, SJ, SO or other genotypes of sickle cell disease)

Exclusion Criteria:

1) Individuals who are incarcerated
2) cured of SCD by gene therapy or stem cell transplant,

Vulnerable Subjects: Individuals who meet inclusion criteria will be asked to come to participate at times separate from clinical care, thus institutionalized individuals and prisoners will not be able to participate. It is anticipated that individuals from economically or educationally disadvantaged backgrounds will be enrolled, however special care will be taken to ensure that such individuals understand all aspects of the study and all of their rights as research participants. As part of his training in the responsible conduct of research, the PI is trained in the recruiting and consenting of vulnerable populations.

iv. Methods and Procedures

Methods and Procedures: Summarize the research design and sequentially identify all procedures to be used to accomplish the specific aims of the project. Clearly identify and distinguish procedures that are considered experimental, procedures that are performed exclusively for research purposes (including “extra” routine tests), and procedures that would occur regardless of the research (i.e., standard of care). Point out any procedures, situations, or materials that may be hazardous, and the precautions to be exercised to maintain subject safety.

Study design: Prospective, target-trial-emulation cohort study of individuals with SCD. This is an observational study where, based on clinician discretion and local site preferences, patients may or may not receive treatment with a new SCD medication (crizanlizumab, L-glutamine or voxelotor) and patients may cycle on and off various regimens. While patients will not be randomly assigned treatment, they will be followed as though they are participating in a clinical trial with regular assessment of medication adherence and outcomes including clinical events, laboratory outcomes and patient-reported outcomes (PROs). All participants will be enrolled during year 1 of the award and followed for the 5-year duration. Study visits will occur at enrollment, before any SCD medication change and at 4-month intervals after medication changes (**Table 3**). All procedures described below are for research purposes unless specifically noted (e.g. "standard care labs" are for standard care, any other tests are for research).

Table 3: Timeline of Visits and Data Collection

Visit #	**V1**	**V2**	**V3**	**V4,5…**	**…Vn**	**Vn+1***	**Vn+2,3…**
Week #	0	Before med change	V2 + 4 months	Every 4 months	Before med change	Vn + 4 months	Every 4 months
Demographics, social determinants/health	X	X	X	X	X	X	X
Patient report: PROs, adverse events	X	X	X	X	X	X	X
Medical record extraction	X	X	X	X	X	X	X
Standard care labs†	X	X	X	X	X	X	X
Whole exome sequencing (WES)	X						
scRNA-seq (medication super- and non-responders)		One blood draw during years 2-4					
Adherence to key medications		Monitored monthly					
Financial incentive - $25	X	X	X	X	X	X	X

To emulate previous phase III trials, data collections will occur before initiation of a new medication regimen and then at 4-month intervals thereafter. † Standard labs CBC, Retic, CMP, LDH, NT-pro BNP, HbF%, bilirubin, urinalysis, urine ACR.

Biospecimens: In addition to clinical labs, blood will be collected at enrollment for whole exome sequencing (WES) and plasma preservation. EDTA tubes will be refrigerated and shipped weekly to Mount Sinai for DNA extraction and banking. DNA aliquots will be shipped to the BROAD Institute for WES. Additional aliquots will remain banked at Mount Sinai for future whole genome sequencing if funding can be obtained. During years 2-4, additional biospecimens will be collected on a small subset (n=60) of medication super- and non-responders. Biospecimens drawn on super- and non-responders will include a heparinized whole blood tube shipped overnight to Mount Sinai for PBMC isolation and whole-blood scRNA-seq analysis.

Predicted attrition: In previous interventional studies in our SCD population, the rate of study withdrawal experienced was 1.2% per month with a cumulative loss of 14% at one year.[104] We conservatively predict 20% study withdrawal, and plan to enroll new patients during years 2-4 to maintain the sample size at 1200 patients.

Outcomes: Table 4 lists primary and key-secondary outcomes emulated from phase III studies. Among a variety of outcomes used in previous trials, measures of organ function predominate as primary outcomes in the current study because they most strongly predict a drug's effect on long-term morbidity and mortality.

Table 4: REAL Answers primary and key-secondary SCD outcomes					
Primary outcomes	**Biomarker or score**	**Ascertainment**	**Rationale**	**Previous phase II or III trial outcome?**	**Clinical trial to emulate**
Heart-lung	NT-proBNP	Blood, urine and PROs measured at each data collection visit	Predicts mortality	Yes	Walk PHASST
Kidney	Urine ACR		Predicts mortality	Yes	STEADFAST
Blood	Hemolysis score		Predicts mortality	Yes	Walk PHASST
Patient-reported	ASCQ-Me symptom score		Patient centered	Yes	HOPE
Secondary outcomes - acute adverse events (other phase III trial outcomes)					
Utilization	Rate of painful episodes	Procedures modeled from SUSTAIN and HOPE trials	Patient centered	Yes	SUSTAIN/ HOPE
Safety	Rate of acute chest syndrome		Safety endpoint	Yes	SUSTAIN/ HOPE
Safety	Hospitalizations, ED visits		Safety endpoint	Yes	SUSTAIN/ HOPE
Safety	Stroke/TIA		Safety endpoint	Yes	SUSTAIN/ HOPE

Study Timeline							
Activity	**Months 0-3**	**Months 4-12**	**Yr 2**	**Yr 3**	**Yr 4**	**Yr 5**	**Post award Yrs 6 – 10**
Finalize case report forms, study documents	X						
Staff training	X						
IRB approval, site activation	X	X					
Recruitment and data collection		X	X	X	X	X	
Whole exome sequencing*			X	X	X	X	
scRNA-seq			X	X	X		
Interim analyses			X	X	X		
Genomic analyses			X	X	X	X	
Final analyses						X	
Make data publicly available						X	
Dissemination of results, additional funding						X	X
*Whole exome sequencing spread over 4 years due to cost constraints							

Data Analysis and Data Monitoring. Describe the statistical or analytical methods to be used. For all studies involving greater than minimal risk, describe how the data will be monitored to ensure the safety of the subjects. For research involving intervention that entails potential serious risk to subjects, compares blinded treatments over a long time period, or which may call for “stopping rules” for certain endpoints, a data monitoring committee may be required to protect the safety or welfare of subjects. A detailed description of its operation (such as, membership, function, frequency of review, stopping rules) should be included.

Data analysis AIM 1: We will analyze a subset of the collected data to achieve AIM1.

AIM 1: Compare the effect of new SCD medications – crizanlizumab, voxelotor, and L-glutamine – on clinical outcomes in individuals with SCD.

Hypothesis 1A: Crizanlizumab will provide superior protection from organ injury, symptom burden and disease exacerbation.

Hypothesis 1B: Patients receiving a medication with a mechanism of action matching their SCD phenotype will have greater response to treatment.

Rationale: While each of the 3 newly available medications were approved based on an ability to improve short-term clinical outcomes, the effects of each drug on organ injury and symptoms (outcomes most predictive of mortality) were not uniformly reported and the drugs were not compared to each other. Furthermore, clinical manifestations of SCD are highly variable, with patients clustering into phenotypes driven by different pathophysiologic mechanisms. Success in this aim would assist providers in choosing medications that are likely to provide the greatest benefit.

Hypotheses 1A analytical plan

Inclusion criteria: We will use criteria modified from the SUSTAIN (crizanlizumab NCT01895361), L-glutamine (NCT01179217) and HOPE (voxelotor NCT03036813) trials: 1) SCD, 2) Age > 15, 3) Hemoglobin > 4.0 mg/dL, 4) For participants taking hydroxyurea, the dose of hydroxyurea must be stable for 3 months.

Exclusion criteria: 1) receipt of chronic anticoagulation other than aspirin, 2) on a chronic transfusion regimen, 3) hepatic dysfunction with ALT > 4 x upper limit of normal, 3) severe renal dysfunction (GFR < 30 or chronic hemodialysis), 4) active participation in an experimental interventional trial at the time of enrollment (at the request of our patient stakeholders, if a patient enrolled in the REAL Answers study chooses at a later time to enroll in an interventional trial, they will be allowed to remain in the REAL Answers study).

As patients are followed during the 5-year award, clinical trial eligibility will fluctuate. We estimate that 90% of the 1200 patients (n=1080) will meet eligibility criteria for at least one follow-up interval during the award. Data during periods of ineligibility will be excluded from the primary analysis but included in a secondary analysis to meet our patient-stakeholder request to generate data for patients who are not eligible for clinical trials (**Fig. 3**).

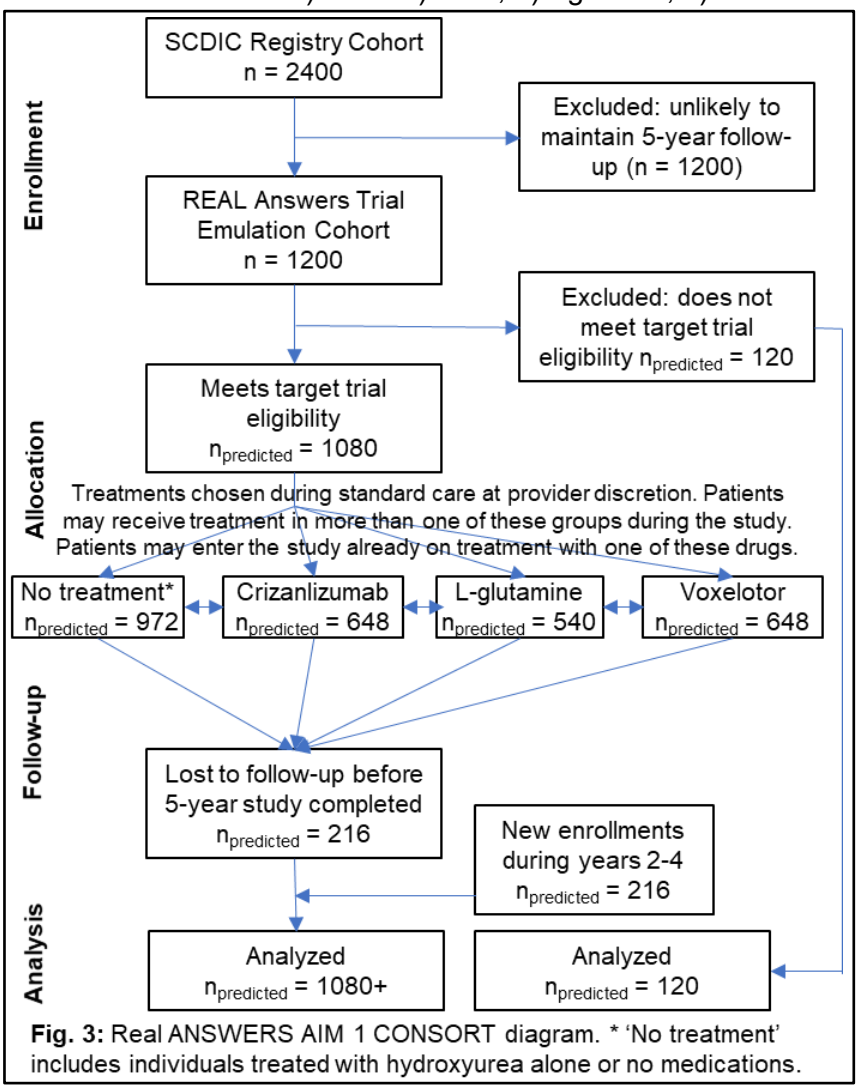


Fig. 3: Real ANSWERS AIM 1 CONSORT diagram. * 'No treatment' includes individuals treated with hydroxyurea alone or no medications.

Exposure: The primary exposure will be medication regimen: a single variable with four categories (no therapy, crizanlizumab, L-glutamine, voxelotor) with 'no therapy' as the referent value. This will be modeled as a time-varying exposure with assessment of organ injury and symptom outcomes every 4 months. While outcomes will be measured every 4 months, medication adherence will be assessed monthly. For oral medications (L-glutamine and voxelotor) adherence will be determined by medication possession ratio (MPR)[63] with direct confirmation with dispensing pharmacies and administering a validated adherence questionnaire.[64] For intravenous medication (crizanlizumab), which is administered in the hospital, adherence is determined by confirmation of drug administration in the electronic medical record (EMR). Hydroxyurea use will also be confirmed by MPR and patient survey. For the primary analysis, a patient will be considered 'on-treatment' with the medication if they received more than 80% of doses over the preceding 4 months. In the event that a patient is placed on more than one new medication, data collected during this period will be excluded from the primary analysis. We expect this to be exceedingly rare due to refusal of insurance to pay for such combination therapy.

Control group: The control group will be patients not on crizanlizumab, L-glutamine or voxelotor. Control patients may or may not be taking hydroxyurea.

Propensity score: To build a score for 4 possible treatment groups (no therapy, crizanlizumab, L-glutamine, voxelotor), we will use the Toolkit for Weighting and Analysis of Nonequivalent Groups (twang) package in R with the multinomial propensity score function. When creating the propensity score, accounting for the multilevel nature of the data produces less biased estimates and better accounts for unmeasured site-level confounders.[65-71] Thus, we will use multinomial logistic regression with site modeled as a fixed-effect to create a propensity score to predict the likelihood of receiving each therapy, calculated at the time of treatment assignment. We will include all patient-level variables in the propensity model, described below. The distribution of propensity scores will be compared across treatment groups to ensure that there is sufficient overlap before balancing methods are applied. We will then use inverse-probability treatment weighting (IPTW) to create a pseudo-population in which covariates are balanced across treatment groups. Covariates will be considered balanced if standardized mean differences and standardized differences in proportion are <0.1 in the weighted sample.[67,72-76] If there is imbalance, we will first consider variable transformation or modifying which variables are included. If balance is still not achieved, we will consider other approaches including doubly robust estimation,[77] entropy balancing[78] or alternative methods of propensity score creation such as generalized boosted models.[79] If a 4-level propensity score does not balance patients, we will fit three separate propensity score models for each new medication vs. 'no treatment'. IPTW methods will then be applied to each of these cohorts.

For the propensity score, SCDIC data collection instruments contain a broad range of variables that reflect disease severity, sociodemographics, access and social determinants of health. Variables include:

Hospital site: Local practice patterns are expected to affect choice of medication.

Socio-demographics, access-to-care: Age, sex as a biological variable, insurance status, employment, smoking, zip code, race, SCDIC measures of access to care, barriers to care, social and emotional health.

Clinical adjustment variables: All comorbidities listed in the EMR, hemoglobin genotype, annualized rate of acute pain visits for the preceding 12 months, opioid consumption (average daily morphine mEq), transfusion history, prior medication regimens, SCD-related complications, steady-state vital signs and laboratory values, surgical history, pain phenotype (based on PROMIS and ASCQ-Me questions), echocardiography results.

Hydroxyurea: Both prior and current use of hydroxyurea will be included in the propensity score.

SCD phenotype: Based on the patient's steady-state lab values, we will use a Bayesian classification method to determine which phenotype of SCD (**Fig. 2**) the subject fits best.[80] In addition to other variables, this measure is useful to balance individuals' severity of illness.

Construction of the multivariable model for the primary hypothesis

Model: In the primary analysis, a model will be created for each of the four primary outcomes (**Table 4**). The outcome will be the change in biomarker or score over the preceding 4-month time interval (e.g. ΔNT-proBNP). This approach was chosen because it is sensitive to the ability of the drug to improve outcome

measures, and to prevent disease progression (i.e. control group outcomes are expected to worsen over time, effective drugs will slow this process and the model can compare the trajectory of this change between groups). This approach also accommodates patients who enter the study already on one of the three new medications because benefit can still be demonstrated by slowing of disease progression. We will assess the effect of each medication on outcomes using repeated-measures, mixed-effects linear regression with random intercepts to model patients' baseline values. Because some medication changes will be from 'no therapy' to starting a new drug, and in other situations, switching from one new drug to another, we will include a 'prior medication regimen' variable; site of participation and hydroxyurea use will also be included as fixed effects. Interaction terms between hydroxyurea and medication regimen will be included to assess whether medication effects differ based on a patient's use of hydroxyurea (**Table 5**).

Table 5: Variables for inclusion in multivariable models for H1A

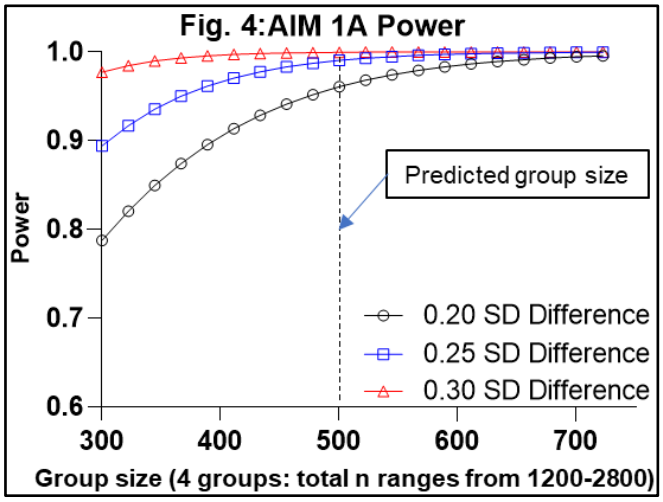
Fixed Effects
-Medication regimen
-Prior medication regimen
-Hydroxyurea use
-Site of participation
Random Effects
-Patient
Interaction Terms
-Hydroxyurea*medication regimen

In addition to comparing each drug to the referent value 'no therapy,' to compare the effect of each drug to one another, the model will include an analysis of type 3 effects.[81] We will check to determine if outcome variables meet assumptions of linearity and consider transformations or non-linear models, if necessary. Secondary acute adverse event outcomes (pain episodes, acute chest syndrome episodes) will be modeled with negative binomial distributions in the same manner as the primary linear outcomes. The α level of p=0.05 will be adjusted using the Holm's procedure to minimize false discovery rate.

Power and sample size: With 4 co-primary outcomes, we used a Bonferroni correction with α = 0.0125. With 1200 patients enrolled and followed for 4-5 years, the number and size of groups exposed to each medication is not known, however, a most conservative estimate of 300 patients per group would assume each patient is assigned only one treatment during the 5-year award and does not change therapy. We used an ANOVA to estimate minimum power because repeated measures will only increase power if within-subject measures are not highly correlated. To inform the sample size calculations, the standard deviation of change in each outcome was calculated using data from the SCDIC Registry. NT-proBNP and urine albumin/creatinine ratio were transformed to logarithms before changes were calculated to improve the fit to the assumptions of the linear model. Mean change and standard deviation are shown in **Table 6**, and **Fig. 4** depicts the change in power as group size increases beyond the minimum of 300. Overall, the study will be powered to detect extremely small differences to allow for additional subgroup analyses. For example, at the predicted group size of 500 patients, power to detect a 0.25 SD difference is 0.996.

Table 6: SDs for the 4 primary outcomes

Variable	Mean change	SD of change
Log(NT-proBNP)	-0.0158	0.2057
Log(urine ACR)	0.0264	0.5412
Hemolysis score	-0.0154	1.060
Pain score	0.0138	27.52



Sensitivity analyses: To confirm the robustness of our findings, we will vary the definition of 'on treatment' to include patients who received more than 50% of doses over the preceding 4 months. We will also perform an 'intent to treat' analysis to include all patients assigned to a therapy regardless of their adherence. While the primary approach will model outcomes as the changeover 4-month intervals, sensitivity analyses will consider outcomes as single values at each time point, with adjustment for baseline values as covariates.

Limitations and alternative approaches: To ensure that unexpected pitfalls are identified early with ample time to modify protocols and overcome challenges, we will conduct annual interim analyses with conservative α thresholds to minimize 'alpha spending.'[82] If the study is **underpowered**, we can enroll additional patients from the SCDIC registry up to, and beyond 2400 patients. If the **propensity score** fails to balance patients, we have several alternatives to achieve pseudo-randomization listed above.' Another concern is that **new medications** could gain FDA approval during the award. Most medications with possible approval during the award share mechanisms with existing therapy, and analyses can group these with older agents. Additionally, our power of >0.99 will allow for addition of new medication classes during the award. In the event that several new agents with novel mechanism of action are approved, we can enroll a significant number of new patients in a short period by leveraging the larger registry to achieve sufficient power for analysis.

Hypothesis 1B analysis: Our goal is to determine if treatment response improves when mechanism of action for a medication matches the patient's SCD phenotype (**Table 7**). Inclusion criteria are the same as H1A.

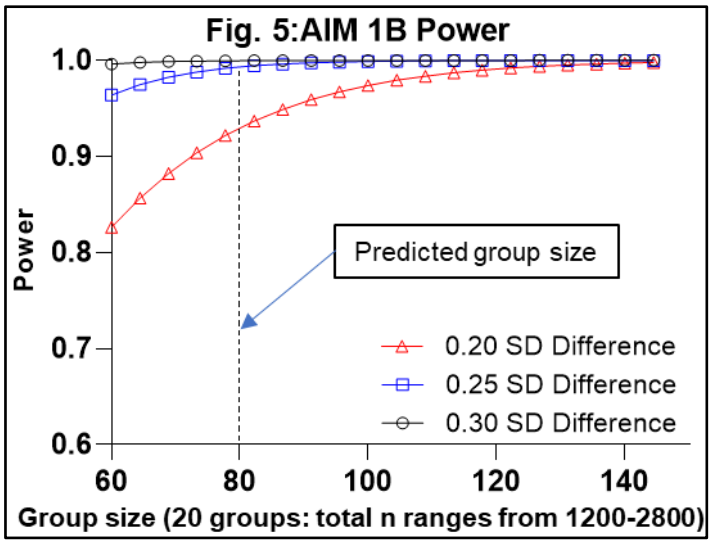
Table 7: Mechanism of action for each medication and the SCD phenotypes predicted to benefit most

Medication regimen	Mechanism of action	Predicted SCD phenotypes with greatest benefit
Crizanlizumab	Anti-leukocyte adhesion	Leuko-hemolytic, vaso-occlusive
L-Glutamine	Antioxidant, arginine repletion	Standard phenotype, hypo-hemolytic
Voxelotor	Anti-sickling	Hyper-hemolytic

Construction of the multivariable model for hypothesis 1B

Model: As for H1A, models will have a primary outcome of the change in organ injury or symptom score over the preceding 4 months in a repeated-measures, mixed-effects linear regression using IPTW and random intercepts to model patients' baseline values. Variables in the model will be the same as H1A (**Table 5**), however, to test the hypothesis that there is heterogeneity of treatment effect by phenotype and medication, an interaction term between medication exposure and SCD phenotype will be included. To assess for different patterns of biological interaction (additive, multiplicative), we will calculate relative excess risk, attributable portion, and synergy index for interaction terms, with bootstrapping to estimate confidence intervals.[83] To generate estimates of the effect of matching phenotype to drug, separate multivariable regressions will be run for each medication, with phenotype in the model to estimate the effect of phenotype on outcome.

Secondary acute adverse event outcomes (pain episodes, acute chest syndrome episodes) will be modeled with negative binomial distributions in the same manner as the primary linear outcomes. A Holm's procedure will be performed to minimize false discovery rate.



Power and sample size: We used α = 0.0125. There are 20 possible combinations for the interaction term between the medication exposure variable ('no therapy,' crizanlizumab, L-glutamine, voxelotor) and the 5 SCD phenotypes. We used an ANOVA with 20 groups, with the assumption that the 5 groups where phenotype is appropriately matched to drug would have better outcomes than the other 15 groups. With the most conservative estimate, assuming that patients are assigned only one therapy during the entire award, the group size would be 60 (1200 divided by 20). We project a size of 80 patients per group which would yield power = 0.991 to detect a difference of 0.25 SD (**Fig. 5**).

Limitations and alternative approaches: As with H1A, if the study is underpowered, additional patients can be enrolled during years 2-4. Another concern is that novel phenotype-drug interactions may be

identified that were not specified *a priori*. If this happens, annual interim analyses will identify such trends. Additional data can be collected later in the award to validate any unexpected findings at interim analyses.

Approach to missing data: The DCC will monitor sites quarterly for presence of missing data and implement corrective action plans for any non-random patterns of missingness. Missingness will be assessed to determine if data are missing at random (MAR), or not missing at random (NMAR). We will investigate this possibility by taking advantage of the longitudinal nature of the study to determine whether the probability that observations of the outcome variables are missing is related to previous observations of the outcomes on the same subjects. If we do find evidence that missing data are NMAR, then we will consider two approaches. First, we will determine whether the statistical model can be adjusted to account for the missing data. This, however, can be difficult to accomplish without making unverifiable assumptions about the missing values. Second, we will undertake a sensitivity analysis to determine how much the missing values would have to depart from the observed values to alter the outcome of an analysis.[84]

Data analysis AIM 2: We will analyze a subset of the collected data to achieve AIM 2.

AIM 2: Identify genetic and genomic predictors of response to crizanlizumab, voxelotor, and L-glutamine.

Hypothesis 2A: There are genetic predictors of response to the 3 new SCD medications.

Hypothesis 2B: Super-response and non-response to treatment correlate with the distribution of cells in whole-blood and their transcriptional profiles

Rationale: Preliminary data from our group and others demonstrate that, in SCD, genetic variants can both increase the risk of organ-specific complications (such as renal disease) and decrease organ-specific response to treatment. In addition, genetic variants appear to correlate with the variability of fetal hemoglobin (HbF) level increase in response to hydroxyurea treatment,[85,86] suggesting that the response to treatment can be genetically regulated. Identification of such risk variant profiles for new SCD medications can aid in personalized medicine choices regarding whether to pursue combination therapies or higher-risk gene therapy cures.

Study design: We will include all 1200 study participants from Aim 1. Response to the 3 treatments will be approximated by change in NT-proBNP, urine ACR, hemolysis score and ASCQ-Me score after 4 months of treatment.

Genetic data: Whole exome sequencing (WES) will be performed at the Broad Institute with DNA extracted from blood samples drawn at the enrollment. The standard protocol includes paired ends reads of 150 base pair lengths, and a minimum coverage of 20X for > 85% targeted bases. We will use state-of-the-art strategies for the generation of variant calls. Briefly, we will use BWA-MEM for alignment of the reads to the genome reference build GRCh38, marking duplicates with Picard, base quality score recalibration with Genome Analysis Toolkit (GATK v4), and lossless conversion to CRAM format with Samtools. The discovery of robust structural variations is more challenging, and therefore we will use LUMPY[96] and Delly2[97] to generate a consensus of structural variants. We will perform sample and variant QC using the GVCF files generated in the previous steps. Sample QC steps will include removal of samples of poor quality based on evidence of contaminations, or insufficient haploid coverage, and removal of samples with sex chromosome aneuploidies or ambiguous sex assignment. Variant QC steps will include filtering based on the quality scores generated during the alignment and variant calling steps, depth of coverage > 20, lack of departure from Hardy Weinberg equilibrium, and call rate > 90%. We will annotate the variants using ANNOVAR[98] for variant annotation.

Genetic analysis: We have implemented a Nextflow pipeline for genetic association analysis that uses state-of the-art methods for (1) inferring genetic relations between study participants; (2) computing genome-wide principal components for adjustment by population admixture; (3) using advanced mixed effect regression models to account for genetic relations. The pipeline is described in and uses packages implemented in the free R software,[90] including PC-Relate[99] and PC-AIR[100] for the calculation of principal components and the genetic relations variance-covariance matrix, and GENESIS[101] and GMMAT[102] for association analyses using mixed effects models. Importantly, the pipeline generates logs of the analysis for easier reproducibility. We will employ the pipeline to conduct single variants analysis using mixed effect

linear regression of changes of each biomarker, adjusted by age, gender, hemoglobin genotype, as well as genetic confounders such as principal components and HbF-mediated risk alleles such as BCL11A, HBS1L-MYB and γ-globin loci.[103] Statistical significance of each variant effect will be assessed using Wald tests and score tests for rare variants. Exome-wide significant results will be identified based on p-value < 5xE-07, and a less conservative level of significance will be used for less rare variants as suggested by Fadista et al.[104] In addition to single variant analyses, we will also consider gene-wide analyses by using burden tests,[105] with weights determined by the single variant associations. We will use $2.5x10^{-6}$ as the gene-based level of significance.

Complex genetic regulations: To estimate the overall joint effects of multiple variants on the response to treatment, we will generate polygenic risk scores (PRS) that include variants not in linkage disequilibrium that are significantly associated with response to treatment with various level of significance. We will establish an optimal threshold for the selection of the genetic variants based on the predictive accuracy of each PRS, using cross validation and ROC analysis.

Hypothesis-driven analysis: Many variants known to modify risk in SCD are not tissue specific and affect overall disease severity. Therefore, in addition to unbiased exome-wide analyses, we will also examine the association of a subset of genes that are known to be involved with tissue specific injuries as summarized in **Table 8**. The advantage of focusing on the specific gene list is to reduce the correction for multiple testing.

Table 8: Genes that are known to be involved with tissue-specific injuries will be used for secondary, hypothesis-driven analyses.

Heart-lung injury: Genes associated with increased risk of pulmonary hypertension, vasculopathy, cardiovascular disease and death (*TGFβ* superfamily: *ACVRL1, BNPR2, BMP6;* β-1 adrenergic receptor. Genes associated with nitric oxide biology (*NOS2, NOS3*), *TNF, IL4R, VCAM1, GALNT13, PRELP*, adenosine-A2B receptor and others.[103,106] Genes known to increase risk of hypertension in African Americans and SCD (e.g. *DRD2, MIR4301*, general African American alleles) are associated with cardiovascular complications.[103]	***Blood injury:*** Genes associated with the degree of blood injury in SCD (*NPRL3, OR51I2, OR51I1, OR51L1*)[107] and hemolysis following blood storage;[108] α-globin gene regulatory elements.[109] ***Kidney injury:*** Genes associated with kidney disease in African Americans and SCD include heme-oxygenase 1, *MYH9 and APOL1.*[103] ***Symptom burden:*** Genes associated with frequency and severity of SCD pain include catechol-*O*-methyltransferase, *OPRM1, GCH1, COMMD7, GSTM1, MTHFR, FVL, VEGFA, CYP2D6, CPY3A, UGTB7,* and *ABCB1.*[110]

Power analysis: We used PASS 18.09 to assess the power of the single variant analysis in various scenarios. We assumed a level of significance of 5xE-07 and a genetic heritability of approximately 30%. The power to detect a 0.7 standard deviation change in response ranged between 70% with 300 subjects and > 90% with 400 or more subjects. The power to detect a 0.6 standard deviation change in response ranged between 68% with 400 subjects and > 88% with 500 or more subjects. The power to detect a 0.5 standard deviation change in response ranged between 51% with 500 subjects and > 86% with 700 or more subjects. The analysis suggests that moderate to smaller effects will be easier to detect with larger numbers of treated patients. However, in general, the study will provide sufficient statistical power to detect clinically relevant effects.

Hypothesis 2B preliminary data:

Preliminary definition of medication super-response through a retrospective cohort – We analyzed data from the Mount Sinai SCD clinic where 85 patients have been treated with crizanlizumab, 91 patients treated with voxelotor and 52 patients were treated with L-glutamine. Before-and-after data regarding heart-lung, kidney and blood injury (Δ NT-proBNP, Δ urine ACR, Δ hemolysis score) was available for all patients. We identified the following patterns of super-response. For crizanlizumab, a small portion (10%) of patients

exhibited greater than 300% improvement in urine ACR and 100% improvement in NT-proBNP, where the median improvement 8% and 6% respectively. For voxelotor 10% of patients exhibited greater than 100% improvements in all three measures of organ injury where the median improvement was less than 20% for all three measures. For L-glutamine, the top performing 10% of patients exhibited improvements in urine ACR of greater than 60% where the median change was 4% worsening. Based on this, we chose a preliminary definition of 'super response' as an improvement of 100% or more in all three markers of organ injury. This definition will be refined during the award as larger amounts of data accumulate. The data suggest that we will be able to identify super- and non-responders for the proposed analyses.

Analysis of single cell RNA isolated from PBMC – We have experience with analysis of RNA sequence data of cells isolated from whole-blood that will be generated in this project. **Fig. 7** shows the UMAP plot of cell types identified from the analysis of single cell RNA sequence data. Cells were isolated from PBMCs of centenarians[111] from the New England Centenarian Study and younger controls, and RNA from individual cells was sequenced using Illumina NextSeq 2000. Raw sequencing files were processed using CellRanger pipeline and Seurat v.3 for filtering, normalization, and principal component analysis. The top 20 significant principal components were corrected for batch effects using the Harmony algorithm.[112] We clustered cells using Louvain community detection method and used the UMAP algorithm to visualize clusters. We identified main cell types using 10 protein markers in our dataset and used publicly available immune cell profiles and singleR[113] to produce a finer classification. We then compared the distributions of cell types isolated from PBMCs of centenarians with those of younger age groups using additional data and showed an increase in the proportion of monocytes and decrease of other cell types. We used a new statistics developed in our group to measure cell type distribution to show that the differences were statistically significant.[91]

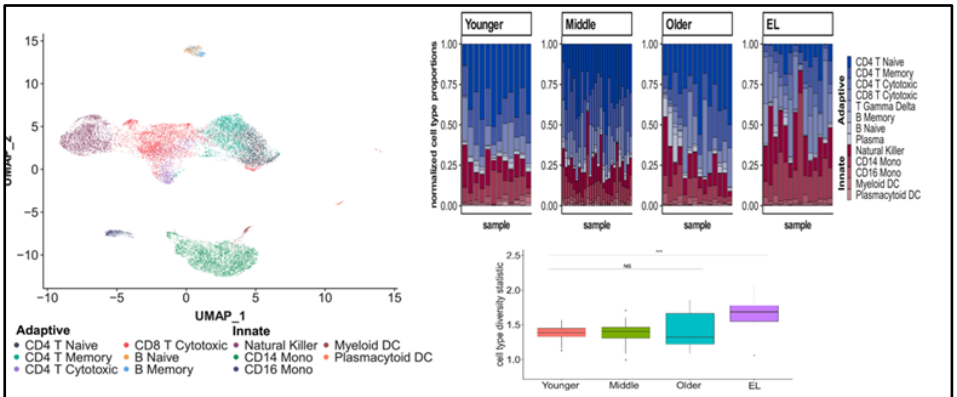


Fig. 7: UMAP plot of cell types detected by scRNA-seq of single cells isolated from PBMCs of centenarians and younger controls. The bar plots show the proportions of cell types in different age groups and suggest major changes of the immune system with age. The differences summarized by a novel statistics are statistically significant (boxplots on the bottom right).

Study design: A case-control study to compare super- and non-responders to medication. Preliminary definitions for super- and non-responders (i.e. cases and controls) are described above. Final definitions will be determined during the award with input from OSMB and study investigators after at least one interim analysis. To reduce variability, controls will be matched to cases by age, gender, and additional confounders. We will include 10 super- and 10 non-responders for each of 3 SCD medications, for a total of 60 participants (**Fig. 8**).

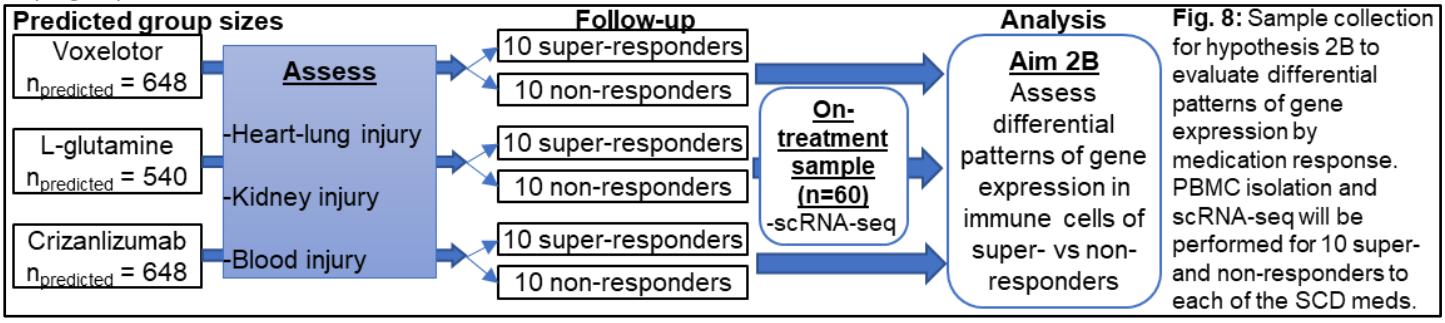


Fig. 8: Sample collection for hypothesis 2B to evaluate differential patterns of gene expression by medication response. PBMC isolation and scRNA-seq will be performed for 10 super- and non-responders to each of the SCD meds.

Biospecimen collection: A total of 20 mL of blood will be collected from each participant: 10 mL in a heparinized tube for PBMC isolation (methods published previously[114]) and scRNA-seq, 10 mL for whole blood scRNA-seq.

scRNA-seq protocol: ScRNA-seq of PBMC will be conducted according to our group's previously published protocol.[115] To generate a comprehensive single-cell transcriptomic profile of whole blood that includes neutrophils, erythrocyte and megakaryocyte lineages we will use methods described by Xie et al.[116]

Single cell transcriptomic analysis: We will convert raw sequencing data to fastq files and demultiplex using bcl2fastq v.2.20 and Cellranger v.3.0.2. We will derive counts for the expression and antibody capture libraries by mapping reads to the human genome (GRCh38) and to the feature reference of the TotalSeq-C antibodies. We will use Seurat v.3[117] for filtering, normalization, and principal component analysis and perform batch correction using the Harmony algorithm.[111] We will cluster cells using graph-based methods (SNN and Louvain community detection method) of the top 20 Harmony components and use the UMAP algorithm to visualize clusters.[112] For identification of cell types represented by the various clusters, will use the cell surface markers and agreement with established gene-expression signatures using the singleR algorithm.[113] We will analyze variations in cell type compositions between cases and controls for each treatment using Bayesian Poisson models, conditionally on the total number of cells for each subject. We will analyze the variations of cell type specific gene expression data using mixed effect regression models of the log-transformed normalized data, with random effects that capture within-subjects variability. We will use the program hypeR[118] for functional annotation of the cell specific gene expression profiles that differ between super- and non-responders.

Power Analysis: We used PASS 16.09 to estimate that with at least 200 cells, this study will have > 80% power to detect difference of at least 1.8-fold change between super-and non-responders. Cell types that are more common will provide greater power to detect smaller fold changes (for example, 1000 cells will provide > 80% to detect differences of at least 1.3-fold change).

Integration of genetics and genomic predictors: The transcriptional changes will suggest possible biological mechanisms that may be activated by the various medications. Although the numbers of super-responders and non-responders is limited by the costs of the data generation, we will correlate the genetic variants and PRS discovered in the previous analyses with cell type abundance and cell transcriptional signatures to test the hypothesis that the genomic data mediate the genetic effects of these variants. This exploratory analysis will help prioritize the findings from the genetic and genomic analysis for future work.

Limitations and alternative approaches: Preliminary data on the distribution of small numbers of super-responders to medication suggest that there will be rare variants associated with the response to treatment and that the effect of these rare variants will be substantial. If, rather than individual rare variants, the genetic effect is made by the combination of common variants with less statistical power, we will refocus our analysis on the use of PRS. The field of scRNA is rapidly evolving, and better methods and technology continue to emerge. We will continue to assess our analysis plans to ensure that we use the most advanced and robust procedures.

Data analysis AIM 3: This aim involves integration of study data into the CureSCi metadata catalog.

AIM 3: Integrate study data into the CureSCi metadata catalog to enhance future cross-study analyses

Biospecimens: DNA extraction on all 1200 participants will generate a minimum of 8 aliquots of DNA and plasma. One aliquot will be used for WES and we will make additional aliquots available through NIH and BioDataCatalyst (BDC) so that additional sequencing (whole genome sequencing, next-generation sequencing, proteomics) can be performed by other investigators and linked to the REAL Answers dataset.

Data curation: CureSCi metadata catalog (MDC) procedures have been described previously.[119] Briefly, key study-level metadata are curated, including 25 data fields organized into five groups as "General," "Research," "Access," "Study Population," and "Documentation." Finally, data dictionaries, protocols, and individual data collection forms are collected, and shared via the online MDC portal (curesicklecell.rti.org). The CureSCi MDC employs a three-tiered conceptual framework to organize and curate study variables. This hierarchical classification system starts with the concept category and is followed by the subcategory

and data element. Variables in each study are first cross-referenced with the existing data elements in the MDC to see if they can be categorized into existing categories. If none of the existing data elements are appropriate for the variable at hand, a new data element is created. PRO Measures are identified and accessioned in the MDC with the measure name and source. To increase interoperability, controlled vocabularies and ontologies such as Medical Subject Headings (MeSH)[120,121,122] and the Sickle Cell Disease Ontology[123] were referenced and adopted for initial category, subcategory, and data element creation. Through the web portal, study-level data or data element level data can be browsed and searched with multiple options for filtering results. The study dataset will be accessible through the MDC but stored at BDC. Genomic raw sequence data of all types will be deposited in the Sequence Read Archive (SRA) or dbGAP, as appropriate, within 90 days after quality control. De-identified study data will also be released publicly and shared following the ENCODE directions. A linking file will be made available in BDC so that users can combine patient-level genetic and clinical data. Genomic data sharing is described in greater detail in the data sharing section.

Data analysis for additional exploratory analyses: Pre-specified analyses are described above. The data collected for this study will also be used for future, hypothesis generating work. REAL Answers data may be combined with other research datasets to improve power.

Data monitoring: A separate document (DSMP REAL Answers) details the study monitoring plan. Briefly, an OSMB will be formed to monitor the study. Annual interim analyses will be performed to evaluate the need for protocol modifications.

Data Storage and Confidentiality: Describe where the research data will be stored during and after the study and how it will be secured. The investigator must take necessary steps to maintain confidentiality of data. This includes coding data and choosing an appropriate and secure data storage mechanism preventing unauthorized access to data. State who will have access to the data and how the data will be used. If data with subject identifiers will be released, specify the person(s) or agency to whom the information will be released and the purpose of the release (such as, routine verification of case report forms).

Transition from Research Participation: If applicable, describe how subjects terminating their participation in the research will be returned to their usual care (such as, taper study medication and resume usual medication, return to primary care provider).

v. Risk/Benefit Assessment

Risk Category. This is a minimal risk study.

Confidentiality breach:
There is a small risk of accidental disclosure of the participants' responses, identifiable information, or study data. The likelihood of these risks is low because the investigators will exercise all due caution in protecting the participants' identity and safeguarding the security of study data to prevent accidental disclosure. The seriousness of these risks is low because the information collected is unlikely to threaten the respondent's financial standing, employability, or reputation.

Phlebotomy:
At the initial baseline visit and at some follow ups, from 15-30 ml of blood (1-2 tablespoons) will be drawn by trained personnel with experience in phlebotomy through a peripheral vein for clinical tests and study assays. Phlebotomy is often associated with mild pain and occasionally with a small bruise at the site. Rarely, patients feel lightheaded or have syncope. There is also a very small risk of infection at the site of the needle puncture.

Questionnaires:
Subjects may feel uncomfortable about answering questions about the impact of their SCD, and may become agitated or upset during the survey. We will do everything within our power to minimize these emotional risks – including training all research staff on these topics and how to handle emotionally-laden situations. Patients will be informed of all procedural risks and of the alternative to participation, which is non-participation. Participants will be assured that their ability to receive care at Mount Sinai will not be affected in any way by their decision about initiating or continuing participation in the study.

Subjects will be repeatedly assured that our only interest is to assess the impact of SCD on their lives. Any subject who is in any way hesitant to disclose information will not be pressured to provide information. Patient identifiable information will not be disclosed to other healthcare providers or any other individual.

At the beginning of all surveys, participants will be informed that all responses are strictly confidential. Participants can refuse to answer any question without any adverse consequences. If a participant becomes uncomfortable at any point during the survey and does not want to continue, he/she will be informed that this can be done with no adverse consequences.

The PI and research assistants will not discuss individual responses or names of individuals who respond to the survey with people who are not project staff.

Protection Against Risk:
All possible steps will be taken to protect the participants against the unlikely and minimal risks described above. In order to ensure patient confidentiality, the following measures are proposed. Electronic data will be stored in password-encrypted files and transmitted by the research assistant via secure, encrypted email to the PI. All paper data forms will be housed in a locked file cabinet within DCT offices until they are batched and brought in person to the research assistant for data entry. The research assistant will check the forms for completeness and enter all data into a REDCap database. The study database will be password-protected and encrypted and data files will be backed up routinely to guard against losses. Access to the study database will be limited to the study personnel. All additional protections of participant confidentiality mandated by the Health Insurance Portability and Accountability Act (HIPAA) will be strictly followed. Additionally, the following data safeguarding procedures will be observed: 1) training staff on data sensitivity and data safeguards being employed to assure confidentiality, 2) storing and processing a hard copy in a centralized location, 3) securing a sensitive hard copy in a locked file, 4) destroying all identifiable linkages to data after data accuracy has been verified and final analyses have been computed, and 5) protecting the database by encrypted password. An encrypted master list of name to participant ID number will be kept by the PI separately in a locked location.

As discussed above, there are risks associated with participation in the proposed project. As detailed in 1.c, risks associated with participation are minimal and all efforts will be made to avoid these risks. Staff will be trained to minimize risks associated with phlebotomy and other collection procedures. Participants will be followed very closely to assess for adverse events (see Data Safety and Monitoring Plan).

If at any time, a subject expresses verbal or nonverbal unwillingness to participate, the subject will be withdrawn from the study.

Protections Against Risk for Children: Not applicable as children are not included in this study. Thus in accordance with CFR 46.404, no additional protections to minimize risks to children are necessary.

Data and Safety Monitoring Plan: To ensure adequate oversight and protection from risks, a data and safety monitoring plan is proposed (see DSMP).

Potential Benefits to the Subjects: Participants are unlikely to directly benefit from the proposed research. Others may benefit from the research in the following ways. If a SCD medication is identified as superior to other medications, subjects may benefit because their provider can place them on that medication sooner than would have been possible without this research. Participants may benefit from the satisfaction of knowing that they are helping other people who might suffer from complications of SCD in the future.

Alternatives to Participation: Participants may choose not to participate. This will have no effect on the care they receive.

vi. Subject Identification, Recruitment And Consent/Assent

Method of Subject Identification and Recruitment

Setting: Hematology clinics of Mount Sinai Hospital (MSH), Augusta University, Children's Hospital of Oakland Research Institute, Duke University, Medical University of South Carolina, University of Illinois at Chicago and Washington University in St. Louis.

Subject identification and recruitment: Participants will be recruited from the hematology clinics of the 8 clinical sites mentioned above. The clinical sites have patient censuses that range from 350 to 900 individuals. Each site has already enrolled over 300 patients into a prospective clinical data registry which demonstrates the feasibility of recruiting and enrolling 150 patients per site into the current study. The initial enrollment, (including consultation, consenting, baseline demographic and HRQOL surveys, blood draw) takes 45-60 minutes.

Process of Consent: Describe or list everyone who is authorized to obtain consent and how the process of informed consent will be structured to be conducive to rational and thoughtful decision making by the subject (or subject's legally authorized representative) without any element of coercion or undue influence. If used, 'Auditor/Witness' roles would be described in this section.

Recruitment and informed consent:
Each site will rely on WCG IRB. WCG will be the single IRB for this study in accordance with CFR 46.114.b.1 NIH common rule for single IRBs with multi-site research.

During preliminary recruitment, the site PI or research associate will obtain permission from the treating physician to approach the subject. At enrollment, all elements of informed consent will be followed and research information participant forms and consent documents will be reviewed and approved by WCG IRB. Individuals who are approached to participate and indicate interest will follow up in a private consultation area. The consent form will fully disclose the nature of the study, and the procedures involved. The PI or research associate will provide the potential participant with any other information the individual needs to make a decision about whether to participate. The individual will have an unlimited amount of time to make a decision. Patients who elect not to enroll in the study will continue to receive standard therapy for the management of SCD. As per WCG IRB regulations, participants will have all information regarding the purpose and description, costs/reimbursements, potential risks, discomforts and/or benefits, alternatives to participation, confidentiality, compensation/treatment, voluntary nature of participation/right to terminate, and contact information for the PI and WCG IRB. Subjects will be informed of the data sharing plan and the protections against identification of individuals associated with it.

Subject Capacity: Subjects with limited capacity will not be enrolled, however a subject may lose capacity during participation in the study. In this case, a legally authorized representative can provide consent according to the following policy. If this policy cannot be followed, the subject will be withdrawn from the study.

POLICY

1.1 Unless the IRB has waived the requirement to obtain consent, when research involves adults unable to consent, permission must be obtained from a Legally Authorized Representative. A "legally authorized representative" means an individual or judicial or other body authorized under applicable law to consent on behalf of a prospective subject to the subject's participation in the procedure(s) involved in the research. If you are unclear about who can act under the law to consent on behalf of a prospective subject, contact the ISMMS Legal Department. Under DHHS and FDA regulations a "legally authorized representative" means an individual or judicial or other body authorized under applicable law to consent on behalf of a prospective subject to the

subject's participation in the procedure(s) involved in the research. Unless the IRB has waived the requirement to obtain consent, when research involves adults unable to consent, permission must be obtained from a legally authorized representative.

1.1.1 When research is conducted in New York, the following individuals meet this definition:

1.1.1.1 A court appointed guardian who is specifically given authorization to consent to participation in research.

1.1.1.2 In the absence of a court appointed guardian who is specifically given authorization to consent to participation in research, any of the following individuals, as defined in the New York Family Health Care Decisions Act, Mental Hygiene Law, and applicable Mount Sinai policies:

1.1.1.2.1 A court appointed guardian who is specifically given authorization to consent to health care.

1.1.1.2.2 A previously designated health care proxy

1.1.1.2.3 Spouse (if not legally separated) or domestic partner[1]

1.1.1.2.4 Children > 18 years of age

1.1.1.2.5 Parents

1.1.1.2.6 Siblings > 18 years of age

1.1.2 For research outside New York, a determination of who is a Legally Authorized Representative is to be made with consultation from legal counsel. This consultation will be facilitated by the IRB staff.

1.1.3 One person from the list of individuals meeting the definition of surrogate from the highest class in priority when persons in prior classes are not reasonably available, willing, and competent to act, shall be the surrogate for an adult patient who lacks decision-making capacity. However, such person may designate any other person on the list to be surrogate, provided no one in a class higher in priority than the person designated objects.

1.1.4 Mount Sinai policies should be followed concerning determination of capacity to consent. Particular attention should be paid regarding who can make these determinations when incapacity may be due to mental illness or intellectual disabilities.

1.2 DHHS and FDA's Subpart D applies to all research involving children. When research is conducted in New York, all individuals under the age of 18 years meet this definition with the following exceptions:

1.2.1 Minors, defined as individuals who meet one of the following criteria, do not meet the DHHS and FDA definition of "children": (Article 24-A §2504)

1.2.1.1 Married/widowed/divorced;

1.2.1.2 A parent;

1.2.1.3 In the case of medical, dental, health and hospital services relating to prenatal care a female who is pregnant.

1.2.2 Individuals under the age of 18 when the research procedures are limited to:

---

[1] ARTICLE 29-CC FAMILY HEALTH CARE DECISIONS ACT defines "Domestic partner" as a person who, with respect to another person:

(a) is formally a party in a domestic partnership or similar relationship with the other person, entered into pursuant to the laws of the United States or of any state, local or foreign jurisdiction, or registered as the domestic partner of the other person with any registry maintained by the employer of either party or any state, municipality, or foreign jurisdiction; or

(b) is formally recognized as a beneficiary or covered person under the other person's employment benefits or health insurance; or

(c) is dependent or mutually interdependent on the other person for support, as evidenced by the totality of the circumstances indicating a mutual intent to be domestic partners including but not limited to: common ownership or joint leasing of real or personal property; common householding, shared income or shared expenses; children in common; signs of intent to marry or become domestic partners under paragraph (a) or (b) of this subdivision; or the length of the personal relationship of the persons.

Each party to a domestic partnership shall be considered to be the domestic partner of the other party. "Domestic partner" shall not include a person who is related to the other person by blood in a manner that would bar marriage to the other person in New York State. "Domestic partner" also shall not include any person who is less than eighteen years of age or who is the adopted child of the other person or who is related by blood in a manner that would bar marriage in New York State to a person who is the lawful spouse of the other person.

1.2.2.1 Diseases dangerous to the public health;

1.2.2.2 Chemical dependency (Mental Hygiene Law §22.11) if, in the judgment of a physician, parental or guardian involvement and consent would have a detrimental effect on the course of treatment of a minor who is voluntarily seeking treatment for chemical dependence or if a parent or guardian refuses to consent to such treatment and the physician believes that such treatment is necessary for the best interests of the child.

1.2.2.3 Prenatal care in the case of pregnant children.

1.2.2.4 Certain outpatient mental health services as described in Mental Hygiene Law §32.21(c) through §32.21(e).

1.3 For research outside New York, a determination of who meets the DHHS and FDA definitions of "children" is to be made with consultation from legal counsel.

1.4 Under DHHS and FDA regulations a "guardian" means an individual who is authorized under applicable State or local law to consent on behalf of a child to general medical care. When research involves children and parental permission is required, consent may only be obtained from parents (biologic or adoptive) or a guardian as defined by DHHS and FDA regulations. When research is conducted in any jurisdiction and permission for a child to participate in research is to be obtained from an individual other than biological or adoptive parents, the individual providing such permission must provide written documentation of the legal ability to consent to the child's general medical care. A copy of this documentation is to be kept with the consent document in the investigator's files.

Subject/Representative Comprehension: Subjects will have an unlimited amount of time to ask questions during the consent process. Comprehension will be assessed by asking the subject if they understand and asking the subject to describe the study in their own words. If study staff have any concerns about comprehension, the PI will be notified prior to completion of the consent process.

Debriefing Procedures: N/A

Consent Forms: See attached consent form.

Documentation of Consent: All consents will be documented in writing. Each site will keep consent forms on file locally. Consent forms will be kept in locked offices where only approved study staff can access them.

Costs to the Subject. N/A

Payment for Participation: Subjects will receive $25 for each study visit which includes a blood draw for research purposes. In most cases this will occur only once. For a small subset of patients who are identified as super-responders or non-responders to therapy (n=60) a second blood draw will occur.