Supplemental Methods

The Patient Discharge Database (PDD) includes all patients admitted to non-Federal hospitals in California from 1991-2014 (there are 14 Federal facilities in California). Each admission includes demographic information, such as age, race/ethnicity, and gender. Clinical variables, coded using the International Classification of Diseases, 9th Revision, Clinical Modifications (ICD-9-CM), include a principal diagnosis and up to 24 other diagnoses and a principal procedure and up to 20 other procedures, including the corresponding procedure dates. The Emergency Department Utilization (EDU) database captures similar information to the PDD for all non-Federal California emergency department visits since 2005.

As described previously,[1,2] SCD cases had to have: 1) at least 2 separate admissions with an SCD code in the principal (first position) diagnosis, or 2) at least one admission with an SCD code in the principal position and a SCD code in a secondary position in at least 2 additional admissions. Specific SCD ICD-9-CM codes (282.41, 282.42, 282.60, 282.61, 282.62, 282.63, 282.64, 282.68, 282.69) were used. All patients had to be <65 years of age at first PDD or EDU encounter to be considered SCD cases. The algorithm emphasized specificity over sensitivity and was used in a previously published study.[1,2]

Person-years of follow-up were calculated for each SCD patient beginning at the estimated date of birth or January 1, 1988 (whichever occurred later). Follow-up was calculated to date of cancer diagnosis, date of death, or end of study -12/31/2014 (whichever occurred first). Person-years of follow-up and the numbers of cancers occurring among SCD patients (observed cancers) were calculated by 5-year age groups, 3-year time-periods, gender and race/ethnicity in order to allow for comparisons to the underlying California population. Expected numbers of first primary cancers were calculated by gender, race/ethnicity, 5-year age-categories, and 3-year time-periods using first primary cancers rates from the California Cancer Registry (CCR). California cancer rates were calculated using Surveillance, Epidemiology, and End Results (SEER) Stat version 8.3.2. Expected number of cancers were generated

assuming the number of observed cancers follows a Poisson distribution.[3] All other analyses were done using SAS® Version 9.4; macro %stratify[4,5] was used to calculate person years of follow-up among the SCD patients.

Supplemental References

1. Brunson A, Lei A, Rosenberg AS, White RH, Keegan T, Wun T. Increased incidence of VTE in sickle cell disease patients: risk factors, recurrence and impact on mortality. *Br J Haematol*. 2017.
2. Adesina O, Brunson A, Keegan T, Wun T. Osteonecrosis of the Femoral Head in Sickle Cell Disease: Prevalence, Comorbidities and Surgical Outcomes in California. *Blood Advances*. 2017;In Press.
3. Breslow NE, Day NE. Statistical methods in cancer research. Volume II--The design and analysis of cohort studies. *IARC Sci Publ*. 1987(82):1-406.
4. Rostgaard K. Methods for stratification of person-time and events - a prerequisite for Poisson regression and SIR estimation. *Epidemiol Perspect Innov*. 2008;5:7.
5. Macaluso M. Exact stratification of person-years. *Epidemiology*. 1992;3(5):441-448.
6. Charache S, Terrin ML, Moore RD, et al. Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. Investigators of the Multicenter Study of Hydroxyurea in Sickle Cell Anemia. *N Engl J Med*. 1995;332(20):1317-1322.

Table S1: Baseline Characteristics of California Sickle Cell Disease (SCD) Cohort, 1991-2014

Variables	All		Cancer		No Cancer	
	N	%	N	%	N	%
Total	6,423	100.0%	115	100.0%	6,308	100.0%
Gender						
Male	3,041	47.3%	50	43.5%	2,991	47.4%
Female	3,382	52.7%	65	56.5%	3,317	52.6%
Race/Ethnicity						
Non-Hispanic White	185	2.9%	7	6.1%	178	2.8%
African-American	5767	89.8%	99	86.1%	5668	89.9%
Hispanic	266	4.1%	6	5.2%	260	4.1%
Asian/Pacific Islander	35	0.5%	.	.	35	0.6%
Other/Unknown	170	2.6%	3	2.6%	167	2.6%
SCD Severity†						
Less Severe	3639	56.7%	61	53.0%	3578	56.7%
Severe	2,784	43.3%	54	47.0%	2,730	43.3%
Cancer-Age at Diagnosis						
Pediatric - age <15	6	0.1%	6	5.2%		
AYA- age 15-39	34	0.5%	34	29.6%		
Adult - age > 39	75	1.2%	75	65.2%		
Cancer-Year of Diagnosis						
1988-1994	6	0.1%	6	5.2%		
1995-2001	25	0.4%	25	21.7%		
2002-2008	36	0.6%	36	31.3%		
2009-2014	48	0.7%	48	41.7%		
Cancer Site‡						
Breast	16	0.2%	16	13.9%		
Digestive System	16	0.2%	16	13.9%		
Respiratory System	16	0.2%	16	13.9%		
Lymphoma	15	0.2%	15	13.0%		
Leukemia	12	0.2%	12	10.4%		
Miscellaneous	8	0.1%	8	7.0%		
Urinary System	8	0.1%	8	7.0%		
Male Genitals	6	0.1%	6	5.2%		
Female Genitals	5	0.1%	5	4.3%		
Brain and other CNS	4	0.1%	4	3.5%		
Myeloma	4	0.1%	4	3.5%		
Bone and Joint	1	0.0%	1	0.9%		
Endocrine System	1	0.0%	1	0.9%		
Kaposi Sarcoma	1	0.0%	1	0.9%		
Oral Cavity System	1	0.0%	1	0.9%		

Soft Tissue	1	0.0%	1	0.9%	

NH refers to Non-Hispanic;
AYA refers to adolescent and young adult, age 15-39 years, and
CNS refers to central nervous system.
†Patients with an averaging ≥3 visits per year were defined as severe SCD; patients averaging <3 visits per year were defined as less severe SCD.[6]
‡Cancer sites defined using SEER Site Recode ICD-O-3/WHO 2008 Definition.