


Validity of International Classification of Diseases Codes for Sickle Cell Trait and Sickle Cell Disease



Kabir O. Olaniran, MD, MPH^{1,2} , Harish Seethapathy, MBBS¹, Sophia H. Zhao, MD, PhD¹, Andrew S. Allegretti, MD, MSc¹, Sahir Kalim, MD, MMSc¹, and Sagar U. Nigwekar, MD, MMSc¹

¹Division of Nephrology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; ²Division of Nephrology, University of Texas Southwestern Medical Center, Dallas, TX, USA.

KEY WORDS: sickle cell trait; sickle cell disease; ICD; validity.

J Gen Intern Med 35(4):1323–4

DOI: 10.1007/s11606-019-05555-w

© Society of General Internal Medicine 2019

INTRODUCTION

Sickle cell trait (SCT) is increasingly being studied as a risk factor for diseases disproportionately affecting African Americans.¹ Research into sickle cell disease (SCD) is also increasing due to poor outcomes in this understudied condition.² Most sickle cell research uses hemoglobin electrophoresis or genetic data to identify patients. However, such information is not always collected or available in clinical care records, limiting adequate sample sizes in large databases with useful outcomes. The utility of International Classification of Diseases (ICD) codes to determine sickle cell status has not been examined in detail. We sought to determine sickle cell prevalence and the validity of sickle cell ICD codes in African American adults in a large multi-hospital healthcare system.

METHODS

We reviewed all adult African American patients with a hemoglobin electrophoresis in the patients' data registry of Partners Healthcare, Boston Massachusetts. Hemoglobin electrophoresis reports were used as the "gold standard" for the diagnosis of SCT and SCD. ICD codes input any time after January 1, 2005, were used as the test. All analyses were conducted using STATA 14.2 (StataCorp.).

SCT ICD codes used were 282.5 or D57.3. SCD ICD codes used were 282.4x (1, 2), 282.6x (0–4, 8, 9), 289.52, 517.3, D57.0x (0–2), D57.1, D57.20, D57.21x (1, 2, 9), D57.40, D57.41x (1, 2, 9), D57.80, and D57.81x (1, 2, 9). ICD code algorithms evaluated were (i) at least one, (ii) at least two, and (iii) at least five of the same or different codes. Anemia (average hemoglobin < 12 g/dL) and low average urine

specific gravity (USG < 1.015), both of which may occur in sickle cell,³ were subsequently added. We determined SCT and SCD prevalence, and the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the area under the curve (AUC) for each algorithm. This study was approved by the Institutional Review Board at Partners Healthcare, Boston Massachusetts and the need for informed consent was waived.

RESULTS

We identified 10,877 African American patients who had undergone a hemoglobin electrophoresis (Table 1). The prevalence of SCT and SCD was 12% and 2%, respectively. Results are shown in Table 2. InSCT, using at least one ICD code had the highest AUC (0.784). Adding anemia and low USG improved the SCTAUC (0.866) but diminished the PPV. In SCD, at least two ICD codes achieved an AUC of 0.956. The addition of anemia and low USG further increased SCD AUC to 0.977 (Table 2).

DISCUSSION

Our findings suggest that ICD codes can adequately adjudicate sickle cell status for observational studies.

SCT rarely has overt clinical manifestations¹; therefore, physicians may not code for sickle cell trait unless prompted. Hence, SCT PPV reached 78% despite the low prevalence in our cohort. Although adding complications increased SCT AUC to 0.866 from 0.784, the PPV fell to 45.5% due to an increase in false positives. Not having a SCT code remained highly predictive (95.8%) for SCT absence. In contrast, SCD ICD code algorithms achieved high AUCs although the PPV was diminished (27%) due to low prevalence. The absence of a SCD ICD code essentially excluded SCD. Adding complications improved SCD AUC; however, the PPV was moderate (52.2%).

The influence of hemoglobin electrophoresis on ICD coding by physicians is unclear; therefore, these results need to be confirmed in a validation cohort where the gold standard is not clinically indicated. Using hemoglobin electrophoresis may have falsely increased the prevalence of sickle cell in our

Received May 23, 2019

Accepted November 8, 2019

Published online November 19, 2019

Table 1. Characteristics of All Patients as of January 1, 2005, Based on Hemoglobin Electrophoresis

Characteristics	Sickle cell trait	Sickle cell disease	Other trait/disease	Normal phenotype
N (%)	1275 (12%)	214 (2%)	674 (6%)	8714 (80%)
Mean age (SD), years	38 (± 15)	34 (± 12)	39 (± 15)	34 (± 13)
Female	78%	55%	75%	88%
Mean hemoglobin < 12 g/dL	53%	93%	57%	57%
Mean hemoglobin (SD), g/dL	11.9 (± 1.5)	9.3 (± 1.7)	11.8 (± 1.5)	11.7 (± 1.3)
Mean urine specific gravity < 1.015	57%	91%	26%	22%
Mean urine specific gravity (SD)	1.0149 (± 0.0038)	1.0135 (± 0.0199)	1.0188 (± 0.0058)	1.0191 (± 0.0058)

Other trait/disease consisted primarily of hemoglobin C trait, beta-thalassemia trait, and homozygous or heterozygous rare variants of hemoglobin

cohort (which was higher than the national average^{4, 5}) and biased this population towards anemia.

In conclusion, sickle cell ICD codes are valuable tools for identifying SCT and SCD patients for much needed large epidemiological studies. Due to moderate PPVs, the use of sickle cell ICD codes in epidemiological studies should be accompanied by a sensitivity analysis using only sickle cell cases confirmed by the available gold standard to verify the direction of observed estimates. Given the risk for misclassification associated with moderate PPVs, we would not recom-

mend using sickle cell ICD codes to create prediction models based on sickle cell status. Rather, with the sensitivity analysis caveat, sickle cell ICD codes would be best suited for investigating clinical associations in retrospective observational data which can subsequently be evaluated prospectively.

Table 2. Sensitivity, Specificity, Predictive Values, and Areas Under the Curve (AUC) of ICD Code-Based Algorithms for the Determination of Sickle Cell Trait and Sickle Cell Disease Status

Algorithm	Sensitivity	Specificity	PPV	NPV	AUC
Sickle cell trait					
Same or different ICD codes ≥ 1	59.1%	97.8%	78.0%	94.7%	0.784
Same or different ICD codes ≥ 2	40.6%	97.7%	70.0%	92.5%	0.692
Same or different ICD codes ≥ 5	17.9%	98.5%	61.6%	90.0%	0.582
ICD codes ≥ 1 + anemia	59.1%	97.8%	78.0%	94.7%	0.805
ICD codes ≥ 1 + low USG	82.8%	74.7%	30.3%	97%	0.854
Sickle cell disease					
Same or different ICD codes ≥ 1	95.3%	94.8%	27%	99.9%	0.951
Same or different ICD codes ≥ 2	93.5%	97.7%	44.4%	99.9%	0.956
Same or different ICD codes ≥ 5	89.7%	98.6%	55.8%	99.8%	0.942
ICD codes ≥ 2 + anemia	93.5%	97.7%	44.4%	99.9%	0.967
ICD codes ≥ 2 + low USG	93.5%	97.7%	44.4%	99.9%	0.973
ICD codes ≥ 2 + anemia + low USG	93.5%	98.3%	52.2%	99.9%	0.977

ICD, International Classification of Diseases; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve

Acknowledgments: K.O.O. is supported by the Ben J. Lipps Research Fellowship Award of the American Society of Nephrology. A.S.A is supported by the American Heart Association Career Development Award 18CDA34110131. S.U.N. is supported by the National Center for Research Program Winter 2015 Fellow-to-Faculty Transition Award 15FTF25980003 from the American Heart Association and by the KL2/Catalyst Medical Research Investigator Training award TR001100 (an appointed KL2 award) from Harvard Catalyst, the Harvard Clinical and Translational Science Center (National Center for Research Resources, and the National Center for Advancing Translational Sciences, National Institutes of Health). S.K. is supported by National Institutes of Health award K23 DK 106479.

Corresponding Author: Kabir O. Olaniran, MD, MPH; Division of Nephrology University of Texas Southwestern Medical Center, 5939 Harry Hines Boulevard, POB 2, Suite HQ7.711C, Dallas, TX 75390-8523, USA (e-mail: kolaniran@mgh.harvard.edu).

Compliance with Ethical Standards:

This study was approved by the Institutional Review Board at Partners Healthcare, Boston Massachusetts and the need for informed consent was waived.

Conflict of Interest: The authors have no conflicts of interest to declare.

REFERENCES

- Naik RP, Haywood C. Sickle cell trait diagnosis: clinical and social implications. *Hematol Am Soc Hematol Educ Progr* 2015;2015(1):160-167. doi:<https://doi.org/10.1182/asheducation-2015.1.160>
- Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. *Lancet* (London, England). 2010;376(9757):2018-2031. doi:[https://doi.org/10.1016/S0140-6736\(10\)61029-X](https://doi.org/10.1016/S0140-6736(10)61029-X)
- Nath KA, Heibel RP. Sickle cell disease: renal manifestations and mechanisms. *Nat Rev Nephrol* 2015;11(3):161-171. doi:<https://doi.org/10.1038/nrneph.2015.8>
- Centers for Disease Control and Prevention. Incidence of Sickle Cell Trait — United States, 2010. *Morbidity and Mortality Weekly Report*. <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6349a3.htm>. Published 2014. Accessed June 2, 2017.
- Centers for Disease Control and Prevention. Data and Statistics | Sickle Cell Disease | NCBDDD | CDC. Sickle Cell Disease Homepage. <https://www.cdc.gov/ncbddd/sicklecell/data.html>. Published 2016. Accessed June 8, 2017.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.