Received: 3 June 2020 | Revised: 27 August 2020 | Accepted: 28 August 2020

DOI: 10.1002/pbc.28703

Pediatric Blood & Cancer

REVIEW

Administrative data identify sickle cell disease: A critical review of approaches in U.S. health services research

Scott D. Grosse[1] | Nancy S. Green[2] | Sarah L. Reeves[3,4]

[1] Centers for Disease Control and Prevention, National Center on Birth Defects and Developmental Disabilities, Atlanta, Georgia

[2] Department of Pediatrics, Division of Pediatric Hematology, Oncology and Stem Cell Transplantation, Columbia University Medical Center, New York, New York

[3] Department of Pediatrics, Susan B. Meister Child Health Evaluation and Research Center, University of Michigan School of Medicine, Ann Arbor, Michigan

[4] School of Public Health, Department of Epidemiology, University of Michigan, Ann Arbor, Michigan

Correspondence
Nancy S. Green, Columbia University Medical Center, 630 West 168 St., Black Building, Box 168, New York, NY 10032.
Email: NSG11@cumc.columbia.edu212-305-0494

Funding information
National Heart, Lung, and Blood Institute, Grant/Award Number: 5K01HL132057

Abstract

To identify people living with sickle cell disease (SCD) and study their healthcare utilization, researchers can either use clinical records linked to administrative data or use billing diagnosis codes in stand-alone administrative databases. Correct identification of individuals clinically managed for SCD using diagnosis codes in claims databases is limited by the accuracy of billing codes in outpatient encounters. In this critical review, we assess the strengths and limitations of claims-based SCD case-finding algorithms in stand-alone administrative databases that contain both inpatient and outpatient records. Validation studies conducted using clinical records and newborn screening for confirmation of SCD case status have found that algorithms that require three or more nonpharmacy claims or one inpatient claim plus two or more outpatient claims with SCD codes show acceptable accuracy (positive predictive value and sensitivity) in children and adolescents. Future studies might seek to assess the accuracy of case-finding algorithms over the lifespan.

KEYWORDS
billing codes, diagnostic codes, healthcare use, Medicaid, sickle cell disease

1 | INTRODUCTION

Sickle cell disease (SCD) is an inherited blood condition that is most common in populations in sub-Saharan Africa, the Mediterranean basin, the Middle East, and India.[1] Cost and healthcare use analyses for SCD conducted in the United States commonly rely on administrative healthcare databases, notably hospital discharge and claimsdatabases.[2–5] Researchers have used multiple inpatient and outpatient encounters from both institutional and noninstitutional providers containing International Classification of Diseases (ICD) billing diagnosis codes for SCD to identify individuals living with SCD and track use of healthcare services such as inpatient and preventive services as well as expenditures. Most have utilized Medicaid data, because Medicaid is the leading payer of SCD-related hospitalizations, followed by Medicare and private insurance.[6] The methodological issues discussed in this focused review apply to claims databases for any or all payers.

Although this review is restricted to U.S. administrative claims and encounters databases, both public and private, that contain records on both inpatient and outpatient clinical encounters to identify people living with SCD, databases that include clinical data can yield more accurate case identification. For example, electronic health records (EHRs) combine encounter records with problem lists of patient diagnoses recorded by clinicians, and studies have demonstrated the accuracy of EHR diagnosis codes for the identification of people living with SCD.[7,8] Similarly, health system databases can identify those with

Abbreviations: CCW, Chronic Conditions Data Warehouse; CDC, Centers for Disease Control and Prevention; CMS, Centers for Medicare and Medicaid Services; ED, emergency department; EHR, electronic health record; HSR, health services research; ICD, International Classification of Diseases; ICD-10-CM, ICD Version 10 Clinical Modification; ICD-9-CM, ICD Version 9 Clinical Modification; NPV, negative predictive value; PPV, positive predictive value; SCD, sickle cell disease; SCDC, Sickle Cell Data Collection program; TCD, transcranial Doppler

clinical diagnoses of SCD.[9] This critical review is intended to help researchers choose among algorithms in stand-alone administrative databases lacking clinical or laboratory records to identify individuals, especially children and adolescents, living with SCD.

We focus on two types of case-finding algorithms, although other algorithms are also discussed. The first is a generic health services research (HSR) algorithm that requires either≥1 inpatient claim with a diagnosis code for a condition or ≥2 outpatient claims (including emergency department [ED] encounters not resulting in admission) on separate days or associated with distinct encounters during a reference period. This generic HSR approach is endorsed by the Centers for Medicare and Medicaid Services (CMS) Chronic Conditions Data Warehouse (CCW) for analyses of Medicare or Medicaid claims data for 16 common chronic conditions, and also 42 other chronic health, mental health, and potentially disabling conditions.[10] The rationale for the requirement of multiple outpatient claims with specified diagnosis codes is that outpatient claims are more subject to coding errors and "rule-out" codes for evaluation visits, laboratory tests, or imaging procedures, whereas hospitals have standardization and quality assurance procedures for coding.[11–18] A second, novel type of algorithm requires ≥3 claims with diagnosis codes in any setting within a reference period; in June 2019 this approach was endorsed by the CCW for one condition, SCD.

We also summarize the findings of studies that have used diagnoses in medical or laboratory records or newborn screening program data as references to validate SCD case-finding algorithms in children and adolescents using billing codes in administrative databases.[19–23] The most commonly used measures of the accuracy of case-finding algorithms in general are sensitivity and positive predictive value (PPV); fewer studies report specificity and negative predictive value.[16,24] Low sensitivity, the percentage of true cases detected by the algorithm in a validation sample, can adversely affect the representativeness of cases in a study. A high PPV, the proportion of cases identified in an algorithm confirmed to be true positive cases in a validation sample, minimizes misclassification of false positives.[18] However, PPV may be overstated if prevalence is markedly higher in the validation sample than in the administrative database.[24] Calculations of specificity and negative predictive value may be subject to overstatement if the validation cohort is enriched with cases, i.e., not representative of the administrative population with diagnosis codes.[24] Therefore, estimates of specificity and negative predictive in SCD validation analyses are not necessarily comparable. We report in Table 1 all measures using the information available in the original articles, but our focus in the text is on PPV and sensitivity.

2 | ALGORITHMS FOR DETECTION OF SCD

2.1 | Algorithms using ≥1 inpatient claim or ≥2 outpatient claims

The ICD Version 9 Clinical Modification (ICD-9-CM) diagnosis codes for SCD overall are 282.6× and, beginning in 2003, 282.41 and 282.41 for sickle cell-beta thalassemia. The corresponding ICD-10-CM codes, which have been used in U.S. healthcare databases since October 1, 2015, include all D57 codes except D57.3 for sickle cell trait. The ICD-10-CM diagnosis codes include new 6-digit codes for the presence of complications in combination with SCD subtypes.

Numerous peer-reviewed studies that analyzed public or private claims databases have required either 1 inpatient claim with a diagnosis code for SCD or ≥2 outpatient claims with a diagnosis code for SCD on separate days to identify cases of SCD.[2,3,25–45] One of these studies excluded individuals who also had a claim with a diagnosis code for sickle cell trait,[40] citing a Wisconsin study that reported that individuals who had diagnosis codes for both SCD (282.60) and sickle cell trait were confirmed to have trait.[8]

SCD studies have set minimum days between two outpatient claims from 1 to 30 days apart. The number of years of claims data searched to identify SCD cases, i.e., the reference period, has varied across this timespan among the cited studies. The influence of the length of the reference period on the number and characteristics of identified cases using the generic HSR algorithm has not been assessed in the literature. One study that used a 5-year reference period noted that 12% of children who met the SCD case algorithm had no SCD claim in the most recent year; those children had much lower healthcare expenditures.[2]

Three published reports have reported information on the accuracy of the generic HSR algorithm to identify cases of SCD in claims data using ICD-9-CM codes, two of which were designed as validation studies (Table 1). Reeves et al. linked newborn screening (NBS) program records of children with laboratory-confirmed SCD diagnoses to 12 months of Michigan Medicaid claims data in 2010 or 2011; in both years, an algorithm using ≥1 SCD inpatient claim or ≥2 outpatient claims on separate days had a PPV of 94.5%.[20] That is the only published study to date that compared the validity of a generic HSR algorithm to the approach of using ≥3 claims in any setting on separate dates to identify SCD cases.

Two studies used Tennessee Medicaid (TennCare) claims and encounters data to identify likely SCD cases with confirmation by either NBS or clinical records. First, Halasa et al. ascertained probable SCD cases based on the presence of SCD codes in 1 inpatient claim or 2 outpatient claims at least 30 days apart.[30] Among 363 children born during 1996 to 2003 who met the algorithm, 312 (PPV 86%) had SCD confirmed in NBS diagnoses. The authors also reported sensitivity of 91%. Supporting evidence comes from a nonvalidation study by Eckrich et al. that confirmed that 88.3% of 653 children who had Medicaid claims linked to medical records at one of two SCD treatment centers in Tennessee and had SCD diagnosis codes in ≥1 SCD inpatient claim or ≥2 outpatient claims at least 30 days apart had SCD.

2.2 | Algorithms using 3 or more claims with ICD diagnosis codes for SCD

Investigators at the University of Michigan and Michigan Department of Health and Human Services (2014) developed and validated an algorithm based on ≥3 claims on separate dates in any setting and position

TABLE 1 Assessments of the validity of ICD-9-CM–based algorithms for identifying pediatric SCD in U.S. healthcare databases

Author	Case definition[a]	Study period and reference period for algorithm	Age group	Source of ICD-9-CM data	Validation data source	Validation results
(1) Algorithms using ≥1 inpatient or ≥2 outpatient claims or encounters						
Halasa et al. (2007)[30]	1 inpatient claim or ≥2 outpatient claims ≥30 days apart with a diagnosis code for SCD Note ICD-9-CM codes of 282.41 and 282.42 were not included	10 years, 1995-2004, for both	Children, born in state, 1996-2003	Tennessee Medicaid claims, 1995-2004	Tennessee NBS data, 1996-2003	PPV 86.0% (312 confirmed cases out of 363 who met SCD case-finding algorithm) Sensitivity 91.2% (312 true positives and 32 false negatives among 344 children with SCD confirmed by NBS records)
Reeves et al. (2014)[19]	1 inpatient claim or ≥2 outpatient claims ≥1 day apart with SCD codes during 12-month period	1 year, 2010 or 2011, for both	Children, ages 0-18 years, born in state	Michigan Medicaid claims, 2010 and 2011	Michigan NBS data, 1987-2010	2010 data PPV 94.5% Sensitivity 90.2% Specificity 90.4% 2011 data PPV 94.5% Sensitivity 90.2% Specificity 90.4%
(2) Algorithm using 3 or more claims with ICD codes for SCD						
Reeves et al. (2014)[19]	Presence of ≥3 claims (any setting) on separate dates with SCD codes during 12-month period	1 year, 2010 or 2011, for both	Children, ages 0-18 years	Michigan Medicaid claims, 2010 and 2011	Michigan NBS data, 1987-2010	2010 data: PPV 95.0% Sensitivity 90.7% Specificity 91.3% 2011 data: PPV 95.8% Sensitivity 89.7% Specificity 87.9%
Snyder et al. (2019)[23]	Presence of ≥3 claims any setting on separate dates with SCD codes during the 5-year period Additional ICD-9-CM code included: 282.6	5 years, 2004-2008, for both	Children and young adults, ages 0-21 years	Georgia Medicaid, Children's Health Insurance Program, State Health Benefit Plan, Georgia hospital discharge data, 2004-2008	Children's Healthcare of Atlanta medical and laboratory records, Georgia NBS records	PPV: 97.4% Sensitivity: 96.0% Specificity: 76.5% NPV: 68.2%
(3) Combinations of diagnosis codes with SCD-associated treatments, procedures, and complications						
Snyder et al. (2019)[23]	≥2 claims on separate dates with SCD ICD-9-CM codes and ≥1 claim with code for an SCD-associated treatment, procedure, or complication Additional ICD-9-CM code included: 282.6	5 years, 2004-2008, for both	Children and young adults, ages 0-21 years	Georgia Medicaid, Children's Health Insurance Program, State Health Benefit Plan, Georgia hospital discharge data, 2004-2008	Children's Healthcare of Atlanta medical and laboratory records, Georgia NBS records	PPV: 97.4%, Sensitivity 85.8% Specificity: 79.0% NPV: 38.2%

Abbreviations: ED, emergency department; ICD-9-CM, International Classification of Diseases Version 9 Clinical Modification; NBS, newborn screening; NPV, negative predictive value; PPV, positive predictive value; SCD, sickle cell disease

[a] All case definitions include the following ICD-9 codes unless otherwise indicated: 282.60-282.69 and 282.41,282.42.

in a single year of Medicaid claims to ascertain SCD in children.[20,46–48] The algorithm was derived from a data-driven process to assess the predictive power of 37 claims-based algorithms for SCD. The authors selected algorithms that reflected combinations of settings (inpatient, outpatient, home health care, ED, blood transfusion), medication categories (antibiotic prophylaxis, hydroxyurea), evaluation or consultation claims, and an overall count of SCD claims, irrespective of type of service.[20] The gold standard of NBS program records was linked to Michigan Medicaid claims data during 2010 to compare the accuracy of all 37 algorithms. The receiver-operating curve, which balances sensitivity and specificity, was high for four algorithms; an algorithm requiring ≥3 SCD claims in any setting and position in 12 months resulted

in a PPV of 95.0%, which was slightly higher than for the generic HSR algorithm, 94.5%. Reeves et al. validated both algorithms with 2011 Medicaid claims and found PPVs of 95.8% and 94.5%, respectively, although the HSR algorithm had slightly higher sensitivity, 90.2% vs 89.7%. Reeves et al. argued that the algorithm requiring ≥3 SCD claims in any setting had the advantages of simplicity and ease of calculation.

The Sickle Cell Data Collection (SCDC) program is a collaboration with the Centers for Disease Control and Prevention (CDC) in two states to date, California and Georgia. The SCDC has used the presence of SCD diagnosis codes in 3 or more claims or encounters in a 5-year reference period to ascertain SCD cases.[23,49] In California, Paulukonis et al. used ≥3 clinical encounters in administrative databases (Medicaid claims and state hospital and ED discharges) with SCD ICD-9-CM during a 5-year period, 2004 to 2008, to identify probable SCD cases.[49] In Georgia, Snyder et al. reviewed medical, laboratory, and newborn screening program records for individuals ≤21 years of age seen at Children's Healthcare of Atlanta during 2004 to 2008 with medical records linked to three administrative claims and encounters databases (Medicaid, Children's Health Insurance Program, and State Health Benefit Plan) for the same 5-year surveillance period. The authors assessed the PPV and sensitivity of 12 administrative case definitions based on from ≥1 up to 6 SCD-related encounters, defined as nonpharmacy medical claims (including laboratory and radiology claims) on separate service dates within a 5-year reference period (no restriction for continuous enrollment). In addition, they assessed a 13th algorithm, discussed in the next section. The PPV increased monotonically with the number of encounters, from 90.0% to 99.0%, and sensitivity decreased from 100.0% to 90.0%. Using ≥3 or more encounters as the criterion, sensitivity was 96.0% and PPV was 97.4%, compared with 98.4% and 94.8% for ≥2 or more encounters. The PPV was unchanged when the surveillance period was reduced from 5 years to a 12-month period within adjoining calendar years, although sensitivity was slightly reduced.

In June 2019, CMS endorsed a CCW case-finding algorithm for SCD requiring ≥3 claims on separate dates with ICD-9-CM or ICD-10-CM codes for SCD within 5 years of look-back data.[50,51] In place of a single 5-year look-back period ending with the year for which services or expenditures are assessed, the current SCDC approach uses a rolling 5-year period to assess SCD case status.[52,53] No restriction is placed on minimum length of continuous enrollment.

2.3 | Combinations of diagnosis codes with SCD-associated treatments, procedures, and complications

Some investigators have considered using procedure or drug codes in addition to SCD diagnosis codes to identify probable SCD cases in claims data. Paulukonis et al. proposed the combination of an ICD-9-CM code for SCD in at least two healthcare encounters, independent of setting, and at least one code for an associated treatment, procedure, or complication of SCD,[54] an approach recently adapted by other researchers.[55,56] However, Snyder et al. reported that algorithm had a false-negative rate (1 minus sensitivity) of 14.2%, compared with 4.0% for one requiring 3 diagnostic claims.[23]

3 | DISCUSSION

Administrative data are often used for population-level assessments of utilization of care or expenditures, especially for conditions with low prevalence. For example, U.S. insurance claims databases, both Medicaid and private insurance, have been used to estimate medical costs,[2,3,32,33] uptake of antibiotic prophylaxis,[19,26,28,31,34,48] documented receipt of immunizations,[31,34,47,57] use of hydroxyurea,[35,37,55,58–63] and receipt of transcranial doppler (TCD) screening among individuals with SCD.[27,31,38,46,47,64–66] Some of those studies, especially ones published in the past 5 years, merged NBS or clinical databases, which were used to identify cases of SCD, with linked claims data that were used to track utilization of services.[57,58,64–66] Some of those studies used clinical or NBS records to identify cases of sickle cell anemia associated with homozygous sickle disease (HbSS) or hemoglobin S-beta thalassemia^0 to calculate quality indicators for preventive services with recommendations specific to sickle cell anemia, e.g., TCD screening and hydroxyurea.[58,60,64–66] It is challenging to identify cases of SCD subtypes using administrative data.

Use of 2 years of data to ascertain cases of certain chronic conditions has been reported to improve performance relative to a single year of data,[67,68] which is consistent with the default CCW algorithm for chronic conditions. Snyder et al. found similar predictive value for SCD requiring multiple claims to occur within either 5 years or a 12-month period within adjoining calendar years, not limited to continuous enrollment, although the authors cautioned that their results might not be generalizable.[23] Use of multiple years of claims data may lead to improved identification of individuals with milder disease phenotypes who have fewer healthcare encounters. Amendah et al. reported that among children with SCD identified using 5 years of either Medicaid or Commercial claims data, 12% had no SCD claims during the fifth year despite continuous enrollment during that year.[2]

Algorithms using ≥3 claims in any setting to identify children with SCD were found in the Michigan study to have a slightly higher PPV than an algorithm that used ≥1 inpatient claims or ≥2 outpatient (including ED) claims on separate days, although the sensitivity of each algorithm was comparable (Table 1).[20] In the validation study using Georgia pediatric data, an algorithm requiring ≥3 claims in any setting with a SCD code had a higher PPV than an algorithm requiring ≥2 claims with SCD codes. However, the advantage of requiring 3 claims in both studies was modest in magnitude; other algorithms may have similar performance.

SCD diagnosis codes in inpatient claims were more predictive of SCD case status in the Michigan study than were SCD codes in outpatient claims.[20] A study using data from a children's hospital found that only a minority of individuals who had a single admission with a SCD code and no outpatient SCD encounters had SCD.[23] However, that finding cannot be generalized to claims data because false-positive

SCD diagnoses in inpatient settings are more common in statewide claims data than in records from children's hospitals.[22]

The two published studies that sought to compare and validate multiple claims-based SCD case-finding algorithms each have limits to generalizability. One limitation is that the PPV in Medicaid claims data, as used in the Michigan study, may be higher than in employer-sponsored insurance. The Georgia study also included CHIP data and claims data for state government employees and their dependents but did not separately evaluate PPV by plan type.

An important limitation is that the SCD validation studies only included pediatric subjects. The Michigan study was restricted to children up to 18 years of age, and the Georgia study was restricted to individuals up to 21 years of age. It is unknown whether billing codes are equally predictive of SCD case status in adults. One study reported that ICD-9-CM codes for SCD were more predictive of case status in pediatric hospital EHRs than in adult hospital EHRs,[8] but similar comparisons have not been made with claims data.

Validation data sets can be either population-based or provider-based. Findings may be more generalizable from the Michigan study, which encompassed Medicaid enrollees managed by all types of providers in the state, than the Georgia study, whose validation data were restricted to individuals seen at a children's hospital with an SCD clinic and affiliated facilities.

A final consideration is the number of years of data used to validate claims. As already mentioned, health services researchers have frequently found that use of more than 1 year of claims data, specifically at least 2 years, improves the accuracy of case-finding algorithms. The Georgia study analyzed 5 years of claims data and compared the accuracy of algorithms using 2 calendar years (12 months from first claim) versus 5 calendar years. In contrast, the Michigan study used a single calendar year of data to identify SCD cases, 2010, which it replicated with a separate analysis of 2011 claims data. It did not assess the accuracy of algorithms pooling the 2 years of available claims data.

In June 2019, CMS endorsed a CCW case-finding algorithm for SCD requiring ≥3 claims on separate dates with ICD-9-CM or ICD-10-CM codes for SCD within 5 years of look-back data and used it to estimate the prevalence of SCD in Medicaid and Medicare claims data.[50,51] The CCW condition algorithm is the same as the SCDC algorithm used by Snyder et al., including use of the most recent 5 years of claims data as the reference period and exclusion of pharmacy claims, even though CMS pharmacy claims do not contain diagnosis codes. In place of a single 5-year look-back period ending with the year for which services or expenditures are assessed, the current SCDC approach uses a rolling 5-year period to assess SCD case status.[52] No restriction is placed on minimum length of continuous enrollment.

Some researchers use the presence of ≥1 claim in any setting with an SCD diagnosis code to identify putative SCD cases.[61–63,69–73] However, owing to the frequency of false-positive diagnoses in outpatient claims, that approach can lead to misclassification and result in underestimation of the use of services or costs among individuals with SCD.[20]

4 | CONCLUSION

Researchers can use stand-alone administrative databases for research on healthcare utilization among persons living with SCD. Evidence from validation studies conducted using data for children and adolescents indicates that algorithms that require multiple SCD codes, particularly in records of outpatient claims or encounters, can yield acceptable accuracy of SCD case ascertainment, although the accuracy may vary across the lifespan. Researchers can decide which algorithm best suits their study purposes, e.g., the assessment of uptake of services recommended for those with specific SCD subtypes, taking into account the number of years of available data.

CONFLICTS OF INTEREST

The authors have no conflict of interest to disclose.

DISCLAIMER

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

ACKNOWLEDGMENTS

This work received support from NHLBI grant funding 5K01HL132057 (Reeves). Helpful comments on previous versions were provided by Thomas Adamkiewicz, MD, and Amanda B. Payne, MPH.

ORCID

Scott D. Grosse https://orcid.org/0000-0003-1078-6855
Nancy S. Green https://orcid.org/0000-0002-9877-1561

REFERENCES

1. Piel FB, Steinberg MH, Rees DC. Sickle cell disease. *N Engl J Med*. 2017;376(16):1561-1573.
2. Amendah DD, Mvundura M, Kavanagh PL, Sprinz PG, Grosse SD. Sickle cell disease-related pediatric medical expenditures in the U.S. *Am J Prev Med*. 2010;38(4 Suppl):S550-556.
3. Mvundura M, Amendah D, Kavanagh PL, Sprinz PG, Grosse SD. Health care utilization and expenditures for privately and publicly insured children with sickle cell disease in the United States. *Pediatr Blood Cancer*. 2009;53(4):642-646.
4. Brousseau DC, Owens PL, Mosso AL, Panepinto JA, Steiner CA. Acute care utilization and rehospitalizations for sickle cell disease. *JAMA*. 2010;303(13):1288-1294.
5. Grosse SD, Boulet SL, Amendah DD, Oyeku SO. Administrative data sets and health services research on hemoglobinopathies: a review of the literature. *Am J Prev Med*. 2010;38(4 Suppl):S557-567.
6. Fingar KR, Owens PL, Reid LD, Mistry K, Barrett ML. *Characteristics of Inpatient Hospital Stays Involving Sickle Cell Disease, 2000–2016: HCUP Statistical Brief #251*. Rockville: Agency for Healthcare Research and Quality; 2019.
7. Singh A, Mora J, Panepinto JA. Identification of patients with hemoglobin SS/Sbeta(0) thalassemia disease and pain crises within electronic health records. *Blood Adv*. 2018;2(11):1172-1179.
8. Michalik DE, Taylor BW, Panepinto JA. Identification and validation of a sickle cell disease cohort within electronic health records. *Acad Pediatr*. 2017;17(3):283-287.

9. Schopper HK, D'Esposito CF, Muus JS, Kanter J, Meyer TA. Childhood hearing loss in patients with sickle cell disease in the United States. *J Pediatr Hematol Oncol*. 2019;41(2):124-128.
10. Centers for Medicare and Medicaid Services. Chronic conditions data warehouse condition categories. 2019; https://www.ccwdata.org/web/guest/condition-categories. Accessed February 18, 2019.
11. Ronald LA, Ling DI, FitzGerald JM, et al. Validated methods for identifying tuberculosis patients in health administrative databases: systematic review. *Int J Tuberc Lung Dis*. 2017;21(5):517-522.
12. Worth RM, Mytinger RE. Medical insurance claims as a source of data for research: accuracy of diagnostic coding. *Hawaii Med J*. 1996;55(1):9-11.
13. Andrade SE, Scott PE, Davis RL, et al. Validity of health plan and birth certificate data for pregnancy research. *Pharmacoepidemiol Drug Saf*. 2013;22(1):7-15.
14. Metcalfe A, Sibbald B, Lowry RB, Tough S, Bernier FP. Validation of congenital anomaly coding in Canada's administrative databases compared with a congenital anomaly registry. *Birth Defects Res A Clin Mol Teratol*. 2014;100(2):59-66.
15. Severinsen MT, Kristensen SR, Overvad K, Dethlefsen C, Tjonneland A, Johnsen SP. Venous thromboembolism discharge diagnoses in the Danish National Patient Registry should be used with caution. *J Clin Epidemiol*. 2010;63(2):223-228.
16. McPheeters ML, Sathe NA, Jerome RN, Carnahan RM. Methods for systematic reviews of administrative database studies capturing health outcomes of interest. *Vaccine*. 2013;31(Suppl 10):K2-6.
17. Sickbert-Bennett EE, Weber DJ, Poole C, MacDonald PD, Maillard JM. Completeness of communicable disease reporting, North Carolina, USA, 1995–1997 and 2000–2006. *Emerg Infect Dis*. 2011;17(1):23-29.
18. Mullooly JP, Donahue JG, DeStefano F, Baggs J, Eriksen E. V. S. D. Data Quality Working Group. Predictive value of ICD-9-CM codes used in vaccine safety research. *Methods Inf Med*. 2008;47(4):328-335.
19. Reeves SL MB, Shevrin CA, McCormick J, Freed GL, Dombkowski KJ, Antibiotic prophylaxis among children with sickle cell anemia. 2017; http://chear.org/sites/default/files/stories/pdfs/sca_antibiotic_measure_testing.pdf. Accessed May 2, 2019.
20. Reeves S, Garcia E, Kleyn M, et al. Identifying sickle cell disease cases using administrative claims. *Acad Pediatr*. 2014;14(5 Suppl):S61-67.
21. Eisenbrown K, Nimmer M, Brousseau DC. The accuracy of using ICD-9-CM codes to determine genotype and fever status of patients with sickle cell disease. *Pediatr Blood Cancer*. 2015;62(5):924-925.
22. Snyder AB, Lane PA, Zhou M, Paulukonis ST, Hulihan MM. The accuracy of hospital ICD-9-CM codes for determining sickle cell disease genotype. *J Rare Dis Res Treat*. 2017;2(4):39-45.
23. Snyder AB, Zhou M, Theodore R, Quarmyne MO, Eckman J, Lane PA. Improving an administrative case definition for longitudinal surveillance of sickle cell disease. *Public Health Rep*. 2019;134(3):274-281.
24. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol*. 2011;64(8):821-829.
25. Kuhlthau K, Ferris TG, Beal AC, Gortmaker SL, Perrin JM. Who cares for Medicaid-enrolled children with chronic conditions. *Pediatrics*. 2001;108(4):906-912.
26. Sox CM, Cooper WO, Koepsell TD, DiGiuseppe DL, Christakis DA. Provision of pneumococcal prophylaxis for publicly insured children with sickle cell disease. *JAMA*. 2003;290(8):1057-1061.
27. Eckrich MJ, Wang WC, Yang E, et al. Adherence to transcranial Doppler screening guidelines among children with sickle cell disease. *Pediatr Blood Cancer*. 2013;60(2):270-274.
28. Warren MD, Arbogast PG, Dudley JA, et al. Adherence to prophylactic antibiotic guidelines among Medicaid infants with sickle cell disease. *Arch Pediatr Adolesc Med*. 2010;164(3):298-299.
29. Shankar SM, Arbogast PG, Mitchel E, Cooper WO, Wang WC, Griffin MR. Medical care utilization and mortality in sickle cell disease: a population-based study. *Am J Hematol*. 2005;80(4):262-270.
30. Halasa NB, Shankar SM, Talbot TR, et al. Incidence of invasive pneumococcal disease among individuals with sickle cell disease before and after the introduction of the pneumococcal conjugate vaccine. *Clin Infect Dis*. 2007;44(11):1428-1433.
31. Bundy DG, Muschelli J, Clemens GD, et al. Preventive care delivery to young children with sickle cell disease. *J Pediatr Hematol Oncol*. 2016;38(4):294-300.
32. Delea TE, Hagiwara M, Thomas SK, Baladi JF, Phatak PD, Coates TD. Outcomes, utilization, and costs among thalassemia and sickle cell disease patients receiving deferoxamine therapy in the United States. *Am J Hematol*. 2008;83(4):263-270.
33. Kauf TL, Coates TD, Huazhi L, Mody-Patel N, Hartzema AG. The cost of health care for children and adults with sickle cell disease. *Am J Hematol*. 2009;84(6):323-327.
34. Beverung LM, Brousseau D, Hoffmann RG, Yan K, Panepinto JA. Ambulatory quality indicators to prevent infection in sickle cell disease. *Am J Hematol*. 2014;89(3):256-260.
35. Brousseau DC, Richardson T, Hall M, et al. Hydroxyurea use for sickle cell disease among Medicaid-enrolled children. *Pediatrics*. 2019;144(1):e20183285.
36. Leschke J, Panepinto JA, Nimmer M, Hoffmann RG, Yan K, Brousseau DC. Outpatient follow-up and rehospitalizations for sickle cell disease patients. *Pediatr Blood Cancer*. 2012;58(3):406-409.
37. Ritho J, Liu H, Hartzema AG, Lottenberg R. Hydroxyurea use in patients with sickle cell disease in a Medicaid population. *Am J Hematol*. 2011;86(10):888-890.
38. Bundy DG, Abrams MT, Strouse JJ, Mueller CH, Miller MR, Casella JF. Transcranial Doppler screening of Medicaid-insured children with sickle cell disease. *J Pediatr*. 2015;166(1):188-190.
39. Bundy DG, Muschelli J, Clemens GD, et al. Ambulatory care connections of Medicaid-insured children with sickle cell disease. *Pediatr Blood Cancer*. 2012;59(5):888-894.
40. Han J, Zhou J, Kondragunta V, et al. Erythropoiesis-stimulating agents in sickle cell anaemia. *Br J Haematol*. 2018;182(4):602-605.
41. Ballas SK, Kanter J, Agodoa I, et al. Opioid utilization patterns in United States individuals with sickle cell disease. *Am J Hematol*. 2018;93(10):E345-E347.
42. Badawy SM, Payne AB. Association between clinical outcomes and metformin use in adults with sickle cell disease and diabetes mellitus. *Blood Adv*. 2019;3(21):3297-3306.
43. Boulet SL, Okoroh EM, Azonobi I, Grant A, Craig Hooper W. Sickle cell disease in pregnancy: maternal complications in a Medicaid-enrolled population. *Matern Child Health J*. 2013;17(2):200-207.
44. Singh A, Yan K, Brandow AM, Panepinto JA. Longitudinal trend in emergency department reliance for pain among sickle cell disease patients in Wisconsin. *J Pediatr Hematol Oncol*. 2019;41(7):e438-e442.
45. Kang HA, Barner JC, Richards KM, Bhor M, Paulose J, Kutlar A. Association between vaso-occlusive crises and opioid prescriptions among patients with sickle cell disease: a retrospective claims-based study. *J Health Econ Outcomes Res*. 2020;7(1):94-101.
46. Reeves SL, Fullerton HJ, Dombkowski KJ, Boulton ML, Braun TM, Lisabeth LD. Physician attitude, awareness, and knowledge regarding guidelines for transcranial Doppler screening in sickle cell disease. *Clin Pediatr (Phila)*. 2015;54(4):336-345.
47. Nero AC, Akuete K, Leasure Reeves S, Dombkowski KJ. Pneumococcal vaccination rates in children with sickle cell disease. *J Public Health Manag Pract*. 2014;20(6):587-590.
48. Reeves SL, Tribble AC, Madden B, Freed GL, Dombkowski KJ. Antibiotic prophylaxis for children with sickle cell anemia. *Pediatrics*. 2018;141(3). pii: e20172182.

49. Paulukonis ST, Feuchtbaum LB, Coates TD, et al. Emergency department utilization by Californians with sickle cell disease, 2005–2014. *Pediatr Blood Cancer*. 2017;64(6):e26390.
50. Wilson-Frederick SM, Hulihan M, Anderson KK. *Prevalence of Sickle Cell Disease among Medicaid Beneficiaries in 2012*. Baltimore: CMS Office of Minority Health Data Highlight:2019.
51. Wilson-Frederick SM, Hulihan M, Blaz J, Young BM. *Prevalence of Sickle Cell Disease among Medicare Fee-for-Service Beneficiaries*. Baltimore: CMS Office of Minority Health Data Highlight; 2019.
52. California SCDC Data, 2016: Map and Figures. 2019; https://www.cdc.gov/ncbddd/hemoglobinopathies/scdc-state-data/california-2016.html. Accessed October 2, 2019.
53. Kayle M, Valle J, Paulukonis S, et al. Impact of Medicaid expansion on access and healthcare among individuals with sickle cell disease. *Pediatr Blood Cancer*. 2020;67(5):e28152.
54. Paulukonis ST, Harris WT, Coates TD, et al. Population based surveillance in sickle cell disease: methods, findings and implications from the California registry and surveillance system in hemoglobinopathies project (RuSH). *Pediatr Blood Cancer*. 2014;61(12):2271-2276.
55. Zhou J, Han J, Nutescu EA, Gordeuk VR, Saraf SL, Calip GS. Hydroxycarbamide adherence and cumulative dose associated with hospital readmission in sickle cell disease: a 6-year population-based cohort study. *Br J Haematol*. 2018;182(2):259-270.
56. Han J, Zhou J, Saraf SL, Gordeuk VR, Calip GS. Characterization of opioid use in sickle cell disease. *Pharmacoepidemiol Drug Saf*. 2018;27(5):479-486.
57. Reeves SL, Jary HK, Gondhi JP, Kleyn M, Wagner AL, Dombkowski KJ. Pneumococcal vaccination coverage among children with sickle cell anemia, sickle cell trait, and normal hemoglobin. *Pediatr Blood Cancer*. 2018;65(10):e27282.
58. Anders DG, Tang F, Ledneva T, et al. Hydroxyurea use in young children with sckle cell anemia in New York state. *Am J Prev Med*. 2016;51(1 Suppl 1):S31-38.
59. Candrilli SD, O'Brien SH, Ware RE, Nahata MC, Seiber EE, Balkrishnan R. Hydroxyurea adherence and associated outcomes among Medicaid enrollees with sickle cell disease. *Am J Hematol*. 2011;86(3):273-277.
60. Stettler N, McKiernan CM, Adejoro OO, Walczak NB. Proportion of adults with sickle cell anemia and pain crises receiving hydroxyurea. *JAMA*. 2015;313(16):1671-1672.
61. Lanzkron S, Haywood C, Jr, Fagan PJ, Rand CS. Examining the effectiveness of hydroxyurea in people with sickle cell disease. *J Health Care Poor Underserved*. 2010;21(1):277-286.
62. Stallworth JR, Jerrell JM, Tripathi A. Cost-effectiveness of hydroxyurea in reducing the frequency of pain episodes and hospitalization in pediatric sickle cell disease. *Am J Hematol*. 2010;85(10):795-797.
63. Castro O, Nouraie M, Oneal P. Hydroxycarbamide treatment in sickle cell disease: estimates of possible leukaemia risk and of hospitalization survival benefit. *Br J Haematol*. 2014;167(5):687-691.
64. Reeves SL, Madden B, Freed GL, Dombkowski KJ. Transcranial Doppler screening among children and adolescents with sickle cell anemia. *JAMA Pediatr*. 2016;170(6):550-556.
65. Reeves SL, Fullerton HJ, Cohn LM, et al. Missed opportunities for transcranial Doppler screening among children with sickle cell disease. *Clin Pediatr (Phila)*. 2016;55(12):1093-1099.
66. Reeves SL, Braun TM, Dombkowski KJ, Fullerton HJ, Boulton ML, Lisabeth LD. The role of neighborhoods in the receipt of transcranial Doppler screening among children with sickle cell disease. *J Pediatr Hematol Oncol*. 2015;37(4):269-273.
67. Hebert PL, Geiss LS, Tierney EF, Engelgau MM, Yawn BP, McBean AM. Identifying persons with diabetes using Medicare claims data. *Am J Med Qual*. 1999;14(6):270-277.
68. Rector TS, Wickstrom SL, Shah M, et al. Specificity and sensitivity of claims-based algorithms for identifying members of Medicare+Choice health plans that have chronic medical conditions. *Health Serv Res*. 2004;39(6 Pt 1):1839-1857.
69. Crego N, Douglas C, Bonnabeau E, et al. Sickle-cell disease comanagement, health care utilization, and hydroxyurea use. *J Am Board Fam Med*. 2020;33(1):91-105.
70. Raphael JL, Dietrich CL, Whitmire D, Mahoney DH, Mueller BU, Giardino AP. Healthcare utilization and expenditures for low income children with sickle cell disease. *Pediatr Blood Cancer*. 2009;52(2):263-267.
71. Shah N, Bhor M, Xie L, Paulose J, Yuce H. Sickle cell disease complications: prevalence and resource utilization. *PLoS One*. 2019;14(7):e0214355.
72. Shah N, Bhor M, Xie L, et al. Evaluation of vaso-occlusive crises in United States sickle cell disease patients: a retrospective claims-based study. *J Health Econ Outcomes Rese*. 2019;6(3):106-117.
73. Shah N, Bhor M, Xie L, et al. Treatment patterns and economic burden of sickle-cell disease patients prescribed hydroxyurea: a retrospective claims-based study. *Health Qual Life Outcomes*. 2019;17(1):155.

How to cite this article: Grosse SD, Green NS, Reeves SL. Administrative data identify sickle cell disease: A critical review of approaches in U.S. health services research. *Pediatr Blood Cancer*. 2020;67:e28703.
https://doi.org/10.1002/pbc.28703