Population Based Surveillance in Sickle Cell Disease: Methods, Findings and Implications From the California Registry and Surveillance System in Hemoglobinopathies Project (RuSH)

Susan T. Paulukonis, MA, MPH,[1]* William T. Harris, PhD,[2] Thomas D. Coates, MD,[3] Lynne Neumayr, MPH, MD,[4] Marsha Treadwell, PhD,[4] Elliott Vichinsky, MD,[4] and Lisa B. Feuchtbaum, MPH, DrPH[5]

Background. There are no population-based surveillance systems to determine prevalence, impact or outcomes in sickle cell disease (SCD). Estimates of the SCD population in California range broadly from 4,500 to 7,000, and little is known about their health status, health care utilization or health outcomes. A surveillance strategy was implemented using diverse data sources to develop a multi-dimensional, state-based surveillance system for SCD that includes adults and children and describes utilization, treatment and outcomes. ***Procedure.*** Data from California newborn screening, inpatient and emergency room records, Medi-Cal/Medicaid claims and two SCD special care centers were collected for 2004–2008. A multi-step, iterative linkage process was used to link and de-duplicate these data sources, and case definitions were used to categorize cases. ***Results.*** After linking and de-duplicating, there were 1,975 confirmed cases of SCD, 3,159 *probable* cases as well as 8,024 possible cases. Among individual data sources, newborn screening and data from clinics contributed the greatest number of unique cases to the total. Select analyses of utilization and treatments for the population are described. ***Conclusions.*** Using linked existing data sources, an estimate of the statewide count of the SCD population is possible. The approach can be used to create an in-depth health status profile of the affected population by aggregating utilization, treatment, and outcomes data including mortality and morbidity information. This effort sets the stage for development of an on-going, state-based surveillance system. Pediatr Blood Cancer

Key words: California; data linkage; public health surveillance; sickle cell disease

INTRODUCTION

Sickle cell disease (SCD) is a genetically determined blood disorder resulting from a single amino acid substitution in the beta chain of hemoglobin. SCD causes anemia, significant morbidity and decreased survival, and has a substantial adverse effect on employment, emotional status, productivity and quality of life [1–5]. Incidence at birth is well-monitored via universal newborn screening (NBS) in the US, and was recently determined to be 1:452 among African American and 1:5,644 among the approximately half million births each year in California [6,7]. However, SCD prevalence is not monitored in any state and is difficult to estimate due to a mobile population and immigration among affected racial and ethnic groups [8]. Life expectancy among SCD populations may be changing over time with better treatments [9–12]. SCD prevalence has been estimated using un-linked administrative data (e.g., Medicaid or hospital discharge data), census data combined with incidence data, and other methods with estimates of 80,000–100,000 nationally and 4,200–7,500 in California [13,14], but any single source may be biased toward a particular sub-population of cases [14–17]. For instance, SCD prevalence estimates based on hospital utilization are likely to miss healthier individuals. Additionally, these efforts do not constitute systematic, on-going surveillance, nor do they offer insight into complications, implementation of care standards, access to care and outcomes. Such surveillance is needed to inform policy makers and researchers, and is standard in similar health conditions [18–21]. To be effective, SCD surveillance must be truly population-based, assuring that as many cases as possible are counted and with complete information from all age groups and disease severities. Data collection and analysis should be cost effective and relatively simple to implement and maintain. Finally, information collected must be useful to stakeholders and policy makers, going beyond a simple count.

The Centers for Disease Control and Prevention, sponsored by the National Heart, Lung and Blood Institute, partnered with seven states (California, Georgia, North Carolina, Michigan, Pennsylvania, Florida and New York) to develop and test a multi-dimensional surveillance approach for adults and children with SCD and thalassemia in the included states [20,22,23]. These states' combined population of 16.2 million African American residents (2.3 million in California) and 24.4 million Hispanic residents (14.0 million in California) make up nearly half of the at-risk population for SCD in the US. The two-year (2010–2012) cooperative agreement was known as RuSH, or Registry and Surveillance System for Hemoglobinopathies. Methods used to collect and link California RuSH data, relative contributions and biases of each data source, descriptive aggregated profiles of the population and selected analyses of treatments and healthcare utilization are detailed herein.

Additional Supporting Information may be found in the online version of this article at the publisher's web site.

[1]Public Health Institute, Oakland, CA; [2]Anthem Blue Cross, Walnut Creek, CA; [3]Children's Hospital Los Angeles, Los Angeles, CA; [4]Children's Hospital & Research Center Oakland, Oakland, CA; [5]Genetic Disease Screening Program, California Department of Public Health, Richmond, CA

Grant sponsor: US Centers for Disease Control and Prevention and the National Heart, Lung and Blood Institute, cooperative agreement numbers; Grant numbers: U50DD000568; U50DD001008

Conflict of interest: Nothing to declare.

*Correspondence to: Susan T. Paulukonis, Public Health Institute, PHRESH Project, Genetic Disease Screening Program, 850 Marina Bay Parkway, Bldg F, Richmond, CA 94704 E-mail: susan.paulukonis@cdph.ca.gov

Received 20 March 2014; Accepted 10 July 2014

DOI 10.1002/pbc.25208
Published online in Wiley Online Library (wileyonlinelibrary.com).

METHODS

Data with identifying variables (described below) were requested from state agencies and two of California's largest SCD special care centers, Children's Hospital Los Angeles (CHLA) and Children's Hospital & Research Center Oakland (CHRCO). SCD special care centers provide timely, comprehensive, and appropriate care using multidisciplinary teams; these two centers are located in areas with greater concentration of African American residents and are well known for their high quality of care in SCD. Access to state data was approved as public health surveillance and exempt from review by the California Committee for the Protection of Human Subjects. Use of state data was approved by each agency's data use committee or privacy officer, and use of data from care centers was approved by their respective IRBs [24]. Following the completion of the linkage process, identifiable variables were removed, resulting in a de-identified dataset compiled from seven data sources representing a five-year cohort (2004–2008) of unique individuals.

Clinic Data

CHLA and CHRCO contributed data on cases seen in their clinics or hospitals between 2004 and 2008. CHLA contributed 422 California resident cases seen during 2004–2008, while CHRCO contributed 614 with the same criteria. Data for linking included birth date, sex and zip code; approximately 80% of the individuals in these data also had a valid SSN.

Newborn Screening Data (NBS)

The California Genetic Disease Screening Program NBS program obtains blood specimens from over 98% of live births. NBS data do not contain social security numbers (SSN), but do contain infants' date of birth, sex and race/ethnicity information, mother's zip code and birth hospital; all were used in linking. These data elements were available in useable digital form for birth years 2000 through 2008 for NBS-identified SCD cases.

Medi-Cal Data

The California Department of Health Care Services (DHCS) contributed claims data from state-sponsored health plans—Medi-Cal (Medicaid), California Children's Services and Genetically Handicapped Persons Program (together abbreviated as Medi-Cal). DHCS first searched their 2004–2008 claims for SCD ICD-9 diagnostic codes (see Supplemental Appendix I); all records for any claimant with these codes were retrieved. RuSH staff then compared the list of those cases identified in the Medi-Cal data to those identified in other RuSH data sources and sent DHCS a list of SSNs not seen in the first Medi-Cal data set. DHCS searched for these SSNs and sent matching claims records. Over 95% of Medi-Cal claims included SSNs, and all included birth date, sex and race.

Office of Statewide Health Planning and Development (OSHPD)

OSHPD collects data on hospital inpatient stays and emergency room (ER) treat-and-release visits for all non-military facilities. Records for 2004–2008 inpatient stays and 2005–2008 for ER visits (2005 was the first year data were available) were received. The inpatient data contain over 17,000,000 records; ER data contain 33,000,000 visits. These data included up to 24 ICD-9 diagnostic codes as well as up to 24 procedure codes, and over 90% of records had a valid SSN. All contained birth date, sex, race and patient zip code.

Vital Records Death Certificates

All (>1,144,000) unique state death records for the period 2004–2008 were obtained. Of these, over 99% contained SSNs. These data did not add to case finding efforts and no unique cases were found, but did provide additional information on cause of deaths among the cases that died during the study time frame.

Data Linkage and De-Duplication

Using techniques demonstrated to provide accurate matches across data sources and minimize bias [25–29], a multi-step linkage and de-duplication process was undertaken.

Step 1 (Fig. 1): Data were cleaned to standardize formats. Step 2: From inpatient, ER and Medi-Cal, all records with SSN and ICD 9 codes related to SCD were selected, as well as all cases with a valid SSN from clinic data. Step 3: These lists were then merged and de-duplicated with a combination of SSN, sex and birth date. Step 4: For those cases without SSNs (from NBS or clinics), a search among all data sources to determine SSN was conducted using the following algorithm: Step A: For a given no-SSN case A_i, determine all possible matches B_j in each data source. Score/weight matches based on available variables such as birth date, sex, diagnosis, zip code, hospital visited and/or case name. $|A| < |B|$; Step B: Select match with best score, assign SSN to case, and remove case A_i and selected match B_x for next iteration of matching; Step C: Steps A and B are repeated until all cases A_k are matched. Manually review matches and scores to assure appropriateness of scoring system. Cases remaining with no SSN were assigned a unique identifier. Step 5: Utilization data were merged to the case list using SSN, birth date and sex. This was a one-to-many link, so that one SSN might have hundreds of records. These encounters and claims were collapsed to form one record with counts of complications, treatments and procedures by year for each case. Step 6: Using the utilization data and/or source of data, case definitions were applied to reflect three levels of diagnostic certainty: confirmed, probable or possible cases.

Case Definition

The CDC's RuSH Surveillance Design team, made up of clinicians and others knowledgeable in the clinical course of SCD, developed case definitions to determine the level of certainty of diagnoses from the data. Complications were culled from the diagnosis fields and treatments and procedures were found in the procedure fields of the inpatient, ER and Medi-Cal data. *Confirmed Case:* Laboratory-confirmed diagnosis of SCD includes cases identified through the state NBS program where confirmatory testing is routine, or clinic data containing physician diagnosis with documented confirmatory testing. *Probable Case:* Laboratory screening result of SCD reported by results of state NBS program without report of confirmatory testing (not applicable in California), or appropriate SCD-related ICD-9 code (Supplemental Appendix I) in two or more separate healthcare encounters plus one or more

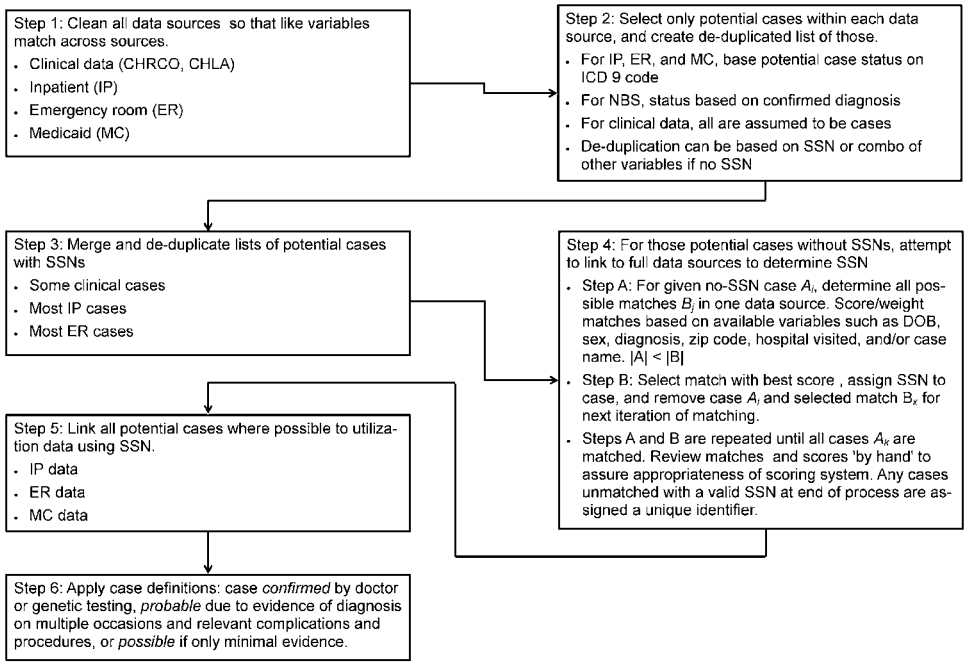

Fig. 1. California RuSH Data Linkage and De-Duplication Process Flow Chart.

SCD-associated complications, treatments or procedures documented between 2004–2008, without laboratory-based confirmation. *Possible Case:* Appearance of sickle cell trait ICD-9 code (282.5) on two or more separate healthcare encounters without SCD codes, plus one or more SCD-associated complications, treatments or procedures documented between 2004–2008, without laboratory-based confirmation, or a single encounter with SCD-related ICD code in administrative database, without lab-based disease confirmation.

RESULTS

Potential Cases Among Unlinked Data Sources

Table I contains counts and demographic information from individual data sources and the contribution of each to the combined dataset. In the 2000–2008 NBS data, there were 1,012 SCD cases. There were 1,036 unique cases identified in the combined clinic data; 1,027 were confirmed via lab testing and nine were presumed SCD but not confirmed.

There were 7,085 individuals with one or more SCD-related stays in the inpatient data, and 7,688 with one or more ER visits. There were 10,111 individuals with one or more SCD-related encounters (an encounter was defined as all activities during one day or one inpatient stay) identified in the Medi-Cal data.

Linked Data

Table I shows the unique contributions of each data source to the confirmed and probable cases combined. While nearly half of NBS cases were unique, in clinical and administrative data there was more overlap among data sources; just over 20% of the combined confirmed/possible cases were unique to one of these data sources.

Table II shows details on confirmed, probable and possible cases identified after linking and de-duplication of all data sources. There were 1,975 unique confirmed SCD cases, 3,159 probable cases and an additional 8,024 possible cases. Confirmed cases, whose status was identified via NBS and/or clinical data, were young (mean age 11.8 years, median 7.7 years), and showed an expected demographic pattern: 87% Black, 7% mixed race, and 7% Hispanic [7,30]. Probable cases came primarily from the administrative data, with the case definition algorithm applied. Their mean age is higher (most younger cases are in the confirmed column), but race and ethnic patterns are similar: 83% Black and 9% Hispanic. Possible cases, those with minimal evidence of being a true case of SCD, were 42% Black, 45% White and 33% Hispanic (ethnicity overlaps race categories); 44% were male. There were a total of 712 death records linked to all cases; including just 152 that stated primary cause of death was SCD.

Among the 5,134 confirmed/probable cases combined, average age (as of 2008) was 23.9 (SD 17.6, median 21.2) years. Cases in this combined group were 85% Black, 7% White, 8% mixed race, 8% Hispanic and 46% male. Among this combined group, 352 had death records linked via SSN. Table III shows a sample of available data on hospital utilization for any diagnosis, mortality (by any cause) and treatments from this combined group by age. Among these cases, 1,496 had no ER visits and 1,385 had no inpatient stays. Data on outpatient and pharmaceutical treatments were available only for the subset of this group included in Medi-Cal data, (n = 3,640). Treatment columns describe the proportion of Medi-Cal cases that had one or more claims for treatments (transfusion, penicillin and hydroxyurea) during the 2004–2008 period. The proportion of children under six years of age with penicillin prescriptions was 62.4%.

Sensitivity of the probable group case definition was calculated using confirmed cases as the true diagnosis, determining whether

TABLE I. Contributions of Data Sources to De-Duplicated Total of Confirmed and Probable SCD Cases

	Data Source				
	NBS[1]	Clinics[2]	Inpatient[2]	Medi-Cal[2]	ER[3]
Total Potential Cases	1,012 (100%)	1,036 (100%)	7,085 (100%)	10,111 (100%)	7,688 (100%)
Sex					
Male	528 (52%)	517 (50%)	2,751 (39%)	4,466 (46%)	3,150 (41%)
Race					
Asian	15 (2%)	5 (<1%)	179 (2%)	454 (7%)	153 (2%)
Black	865 (85%)	916 (88%)	5,303 (75%)	5,320 (49%)	5,263 (69%)
White	74 (7%)	31 (3%)	1,116 (16%)	3,582 (37%)	1,726 (22%)
Mixed/Other	58 (6%)	84 (8%)	487 (7%)	755 (8%)	546 (7%)
Ethnicity					
Hispanic	92 (9%)	50 (5%)	811 (11%)	2,893 (30%)	1,315 (17%)
Age in 2008					
0 to 10 y	1,012 (100%)	371 (36%)	867 (12%)	4,134 (42%)	1,691 (22%)
11 to 20 y	—	336 (32%)	1,061 (15%)	1,572 (18%)	1,114 (15%)
21 to 30 y	—	200 (19%)	1,579 (22%)	1,443 (14%)	1,607 (21%)
31 to 40 y	—	57 (6%)	1,209 (17%)	1,036 (10%)	1,149 (15%)
41 to 50 y	—	37 (4%)	1,064 (15%)	874 (8%)	1,022 (13%)
51 to 60 y	—	30 (3%)	693 (10%)	577 (5%)	645 (8%)
Over 60 y	—	5 (<1%)	612 (9%)	475 (4%)	460 (6%)
Mean Age in 2008 (SD)	4.3y (2.5 y)	17.6y (12.8 y)	33.7y (18.9 y)	22.6y (19.6 y)	29.4 (19.0 y)
Unique Cases in Confirmed/Probable Categories (% of Total)	498 (10%)	194 (4%)	114 (2%)	155 (3%)	12 (<1%)

[1]Newborn screening variables used for matching: Birth date, sex, zip code, race and diagnosis (SCD or thalassemia). Cases include screening years 2000–2008, inclusive. [2]Clinic, inpatient and Medi-Calvariables used for matching: SSN, birth date, sex, zip code (provider zip from Medi-Cal, patient zip from the rest), race and diagnosis. [3]ER variables used for matching: SSN, birth date, sex, zip code, race and diagnosis; only data 2005–2008 available.

they would have fallen into the probable (test positive) or possible case categories (test negative) if lab confirmation was not available. Using 807 confirmed cases that included SSN, sensitivity was 82% overall, with 18% (147) falling into the possible category (i.e., there was not enough evidence in the administrative data to move them into the probable category). Of these 147 test negative cases, 80 (54%) had multiple encounters with SCD ICD-9 codes, but no SCD-related complications, treatments or procedures, while the remainder had one encounter with SCD codes. None of the 807 was missing/unmatched.

TABLE II. Linked and De-Duplicated Counts, SCD Cases by Type

	Confirmed Cases	Probable Cases	Possible Cases
Total	1,975 (100%)	3,159 (100%)	8,024 (100%)
Sex			
Male	985 (50%)	1,352 (43%)	3,504 (44%)
Race			
Asian	18 (1%)	33 (5%)	471 (6%)
Black	1,712 (87%)	2,627 (83%)	3,395 (42%)
White	105 (5%)	251 (8%)	3,604 (45%)
Mixed/Other	140 (7%)	248 (8%)	554 (7%)
Ethnicity			
Hispanic	135 (7%)	299 (9%)	2,659 (33%)
Age			
0 to 10 y	1,284 (65%)	241(8%)	3,313 (41%)
11 to 20 y	334 (17%)	554 (18%)	997 (12%)
21 to 30 y	213 (11%)	697 (22%)	1,104 (14%)
31 to 40 y	70 (4%)	603 (19%)	809 (10%)
41 to 50 y	39 (2%)	583 (18%)	730 (9%)
51 to 60 y	30 (1%)	319 (10%)	522 (7%)
Over 60 y	5 (<1%)	162 (5%)	549 (7%)
Mean Age (SD)	11.8y (11.7 y)	33.5y (16.5 y)	24.2y (21.3 y)
Median Age	7.65 y	32.3 y	18.3 y

TABLE III. Linked and De-duplicated Confirmed and Probable Cases–Utilization, Treatments and Mortality 2004–2008

	Confirmed and Probable Cases Total	Mean # of ER Visits per Year per Case	Mean # of Hospital Admissions per Year per Case	Deaths by any Cause (Count)	Proportion of Medi-Cal Cases Prescribed Penicillin[1]	Proportion of Medi-Cal Cases Prescribed Hydroxyurea[1]	Proportion of Medi-Cal Cases One or More Transfusions[1]
Age							
0 to 10 y	1,525 (30%)	0.59	0.45	16	56.6%	9.7%	48.8%
11 to 20 y	888 (17%)	1.09	1.15	23	41.6%	28.8%	57.7%
21 to 30 y	910 (18%)	3.47	2.12	54	21.7%	30.9%	77.5%
31 to 40 y	673 (13%)	3.08	1.98	64	13.6%	22.7%	72.4%
41 to 50 y	622 (12%)	3.37	2.12	87	12.5%	24.5%	67.3%
51 to 60 y	349 (7%)	3.07	1.50	62	11.0%	11.0%	63.5%
Over 60 y	167 (3%)	1.70	1.32	46	7.3%	8.1%	57.7%
All Ages	5,134	2.05	1.37	352	30.1%	21.4%	63.3%

[1]Among 3,640 cases with linked Medi-Cal claims data.

DISCUSSION

Data sources such as inpatient and ER records, NBS data, and death records have been used individually in the past to estimate SCD prevalence, mortality, complications and healthcare utilization [12–17,31]. In contrast, the effort described in this document uses multiple data sources to improve count estimates and develop a multi-dimensional surveillance system for SCD in California.

While NBS provides genotype information on confirmed cases, complete data are available only from 2000; children over age eight in 2008 and adults were not included. Additionally, resident children with SCD who were born outside of California are not counted in the NBS dataset (and may be missing from the statewide estimate if they did not appear in other included datasets). There were limited variables to link NBS cases to other data and no health care utilization information about these cases after diagnosis. (In 2011 the NBS program implemented a long-term follow-up system to monitor access to care, utilization and outcomes through age five.) Nearly half of NBS cases in the final data set were unique and not linkable to other data sources, perhaps due to the lower likelihood of hospital utilization and SCD complications in the <10 years age range.

Clinical data were more difficult to obtain than state data. Separate IRB approval or exemption for each center was obtained. Both care centers became contractual partners on the RuSH project and staff time was supported through the cooperative agreement. This arrangement was not pursued with other California hemoglobin specialty care centers because of limited resources. Clinical data include cases of all ages, cases born or diagnosed outside of California, and some who are privately insured and/or healthier (i.e., those who were not found in the inpatient or ER datasets).

Medi-Cal, inpatient and ER data, all with a high proportion of SSNs included, were a significant source of case finding (all probable cases came from these combined sources). These three datasets were the only sources for utilization and treatment data and were the primary sources for outcomes data, including disease complications and procedures performed. While each of these sources provides valuable information about the SCD population, there are biases and gaps in using these sources for case finding. ER and inpatient data do not include the healthiest of SCD cases. Medi-Cal data include only those without private insurance. These data do not provide confirmation of diagnosis, and ICD-9 codes pointing to SCD may be 'rule out' codes or data entry errors. While techniques such as counting only individuals with multiple appearances of relevant ICD-9 codes in the data may improve the accuracy of case counts, they may also intensify a bias toward sicker cases.

Using combinations of SSN, birth date, sex, diagnostic information, race/ethnicity, zip code and treating facility, cases were linked across data sources. Using case definitions based on data source and utilization data, cases were placed into categories indicating strength of evidence for diagnosis. Using the confirmed clinical cases with valid SSNs, we found that the probable case definition had a sensitivity of 82%.

The case count of confirmed/probable cases of 5,134 represents 79% of one recent estimation of the state's incidence [13] and is within the range predicted by another [14]. Among 10 year age groups, there are nearly double the number of confirmed/probable cases in the youngest age group compared to the 11–20 y and 21–30 y age groups. This suggests that there may be young cases from multiple data sources unlinked with each other (resulting in duplicates). Thus, the final dataset may have an over-count of younger children. Proportions in all other age groups for the confirmed/probable cases appear as expected given the estimates of population and life expectancy for this disorder [14].

Obtaining, cleaning and linking these data were a complex and time consuming task, and not all RuSH states were successful in obtaining and reporting surveillance data for SCD within the two year funding period. California's clinical partners and state agencies were highly supportive of this effort, streamlining development of data sharing/use agreements and transfer of data as well as providing input on development and linking of the surveillance system. The California RuSH data set is an improvement over single source surveillance efforts, but one with limitations. None of these data sources were designed for disease surveillance, and even when combined there are gaps. Cases not seen in the inpatient or ER data, born prior to 2000 or who moved to the state after birth, not covered by Medi-Cal and not receiving care from CHLA or CHRCO are not included. Cases not seen at CHLA or CHRCO and seen only once for SCD in the other data sources or without expected complications, treatments or procedures will be included in the possible cases. Despite these limitations, this is the most complete picture to date of the SCD population in California

and establishes a baseline for surveillance. This effort demonstrates that linkages across multiple data sources can be established, and the resulting de-duplicated confirmed/probable cases likely represent the majority of SCD cases in the state. Ongoing surveillance would improve the case count and increase the value of these data. With additional funding, on-going data collection and linkage with clinical data from all of the California hemoglobinopathy special care centers, combined with incorporation of electronic health record data would further improve the utility of this surveillance effort and allow for monitoring trends in care and outcomes over time at a population level.

It is possible to develop a SCD surveillance system using diverse data sources and containing diagnostic, treatment and outcome data. RuSH establishes a foundation for discourse on SCD as a serious public health problem in California. Efforts are underway to better validate these methods and case definitions with plans to further describe patient utilization of treatments and services. Unmet needs, outcomes for different care settings, and trends in morbidity and mortality will be described. With continued data collection, a long-term surveillance system can be implemented to follow trends in California's SCD population.

ACKNOWLEDGEMENTS

This work was supported by the US Centers for Disease Control and Prevention and the National Heart, Lung and Blood Institute, cooperative agreement numbers U50DD000568 and U50DD001008. The authors wish to thank Mary Hulihan and Dr. Althea Grant from the Division of Blood Disorders of the National Center on Birth Defects and Developmental Disabilities at the CDC for their leadership and guidance throughout the project, as well as Ms. Hulihan's comments on versions of this manuscript.

REFERENCES

1. Dampier C, Lieff S, LeBeau P, et al. Health-related quality of life in children with sickle cell disease: A report from the Comprehensive Sickle Cell Centers Clinical Trial Consortium. Pediatr Blood Cancer 2010;55:485–494.
2. Prabhakar H, Haywood C Jr, Molokie R. Sickle cell disease in the United States: Looking back and forward at 100 years of progress in management and survival. Am J Hematol 2010;85:346–353.
3. Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. Lancet 2010; 11; 376:2018–2031.
4. Swanson ME, Grosse SD, Kulkarni R. Disability among individuals with sickle cell disease: Literature review from a public health perspective. Am J Prev Med 2011;41:S390–397.
5. Siddiqi AE, Jordan LB, Parker CS. Sickle cell disease–the American saga. Ethn Dis 2013;23:245–248.
6. Benson JM, Therrell BL Jr. History and current status of newborn screening for hemoglobinopathies. Semin Perinatol 2010;34:134–144.
7. Feuchtbaum L, Carter J, Dowray S, et al. Birth prevalence of disorders detectable through newborn screening by race/ethnicity. Genet Med 2012;14:937–945.
8. Buseh A, Kelber S, Millon-Underwood S, et al. Knowledge, Group-Based Medical Mistrust, Future Expectations, and Perceived Disadvantages of Medical Genetic Testing: Perspectives of Black African Immigrants/Refugees. Public Health Genomics 2013; 24.
9. Lanzkron S, Carroll CP, Haywood C Jr. Mortality rates and age at death from sickle cell disease: U.S., 1979–2005. Public Health Rep 2013;128:110–116.
10. Platt OS, Brambilla DJ, Rosse WF, et al. Mortality in sickle cell disease. Life expectancy and risk factors for early death. N Engl J Med 1994;330:1639–1644.
11. Powars DR, Chan LS, Hiti A, et al. Outcome of sickle cell anemia: A 4-decade observational study of 1056 patients. Medicine 2005;84:363–376.
12. Hamideh D, Alvarez O. Sickle Cell Disease Related Mortality in the United States (1999–2009). Pediatr Blood Cancer 2013;60:1482–1486.
13. Brousseau DC, Panepinto JA, Nimmer M, et al. The number of people with sickle-cell disease in the United States: National and state estimates. Am J Hematol 2010;85:77–78.
14. Hassell KL. Population estimates of sickle cell disease in the U.S. Am J Prev Med 2010;38:S512–521.
15. Claster S, Termuhlen A, Schrager SM, et al. Pitfalls of using administrative data sets to describe clinical outcomes in sickle cell disease. Pediatr Blood Cancer 2013;60:1936–1939.
16. Wolfson JA, Schrager SM, Coates TD, et al. Sickle-cell disease in California: A population-based description of emergency department utilization. Pediatr Blood Cancer 2011;56:413–419.
17. Grosse SD, Boulet SL, Amendah DD, et al. Administrative data sets and health services research on hemoglobinopathies: A review of the literature. Am J Prev Med 2010;38:S557–567.
18. Olney RS. Preventing morbidity and mortality from sickle cell disease. A public health perspective. Am J Prev Med 1999;16:116–121.
19. Yusuf HR, Lloyd-Puryear MA, Grant AM, et al. Sickle cell disease: The need for a public health agenda. Am J Prev Med 2011;41:S376–383.
20. Grosse SD, James AH, Lloyd-Puryear MA, et al. A public health framework for rare blood disorders. Am J Prev Med 2011;41:S319–323.
21. Dyson SM, Atkin K. Sickle cell and thalassaemia: Global public health issues come of age. Ethn Health 2011;16:299–311.
22. Hoots WK. The registry and surveillance in hemoglobinopathies: Improving the lives of individuals with hemoglobinopathies. Am J Prev Med 2010;38:S510–511.
23. National Heart, Lung, and Blood Institute. NHLBI, CDC Launch Surveillance and Research Program for Inherited Blood Diseases. 2010 [updated 2/18/2010; cited 1/17/2014]; Available from: http://www.nhlbi.nih.gov/news/press-releases/2010/nhlbi-cdc-launch-surveillance-and-research-program-for-inherited-blood-diseases.html.
24. Lee LM, Heilig CM, White A. Ethical justification for conducting public health surveillance without patient consent. Am J Public Health 2012;102:38–44.
25. Li X, Shen C. Linkage of patient records from disparate sources. Stat Methods Med Res 2013;22:31–38.
26. Clark DE, Hahn DR. Comparison of probabilistic and deterministic record linkage in the development of a statewide trauma registry. Proc Annu Symp Comput Appl Med Care 1995; 397–401.
27. Dean JM, Vernon DD, Cook L, et al. Probabilistic linkage of computerized ambulance and inpatient hospital discharge records: A potential tool for evaluation of emergency medical services. Ann Emerg Med 2001;37:616–626.
28. Zingmond DS, Ye Z, Ettner SL, et al. Linking hospital discharge and death records–accuracy and sources of bias. J Clin Epidemiol 2004;57:21–29.
29. Clark DE. Practical introduction to record linkage for injury research. Inj Prev 2004;10:186–191.
30. Wang Y, Kennedy J, Caggana M, et al. Sickle cell disease incidence among newborns in New York State by maternal race/ethnicity and nativity. Genet Med 2013;15:222–228.
31. Lanzkron S, Haywood C Jr, Segal JB, et al. Hospitalization rates and costs of care of patients with sickle-cell anemia in the state of Maryland in the era of hydroxyurea. Am J Hematol 2006;81:927–932.