National Center on Birth Defects and Developmental Disabilities

DD19-1906 Capacity Building for Sickle Cell Disease Surveillance

Session 10: The Evolution of Surveillance

May 7, 2020

SCDC

SICKLE CELL DATA COLLECTION

The Evolution of Surveillance

Angie Snyder, PhD

May 8, 2020

Georgia Health Policy Center

Georgia State University | Andrew Young School of Policy Studies

Surveillance Initiation

- Georgia began SCD surveillance with RuSH, started with 5 years of data: 2004-2008
  - Paper newborn screening records had to be reviewed for 50% of cases
  - Incidence could be reported once this data was available
- Identify additional data sources that could be used to find cases—
  - Older children and adults
- For states able to link Medicaid, clinical, and hospital discharge data
  - Who is still missing? Children & adults with private health insurance/no health insurance who didn't have 2-3 hospital-level encounters in 5 years

Georgia--5 year SCD Prevalence

Sickle cell cases in Georgia identified through RuSH, Cases 2004-2008		
Data Set	*Confirmed Cases*	*Probable Cases*
Newborn Screening	*730*	*98*
Augusta University (clinical)	*1,218*	*14*
Grady (clinical)	*1,661*	*2*
CHOA (clinical)	*1,908*	*242*
Medicaid/CHIP	*2,986*	*1,993*
State Health Benefit Plan	*209*	*215*
Hospital administrative data	*3,339*	*2,147*
De-duplicated Total	*4,288*	*3,011*
7,299 confirmed and probable cases in Georgia		

Georgia State University | Andrew Young School of Policy Studies

Chronic Disease Surveillance

- Considerations for SCD prevalence with additional years of data
  - Very few individuals will be cured, have SCD from birth
  - Should be able to link to death files
  - If you only count confirmed cases; likely missing 40% or more of the cases
  - Difficult to estimate out-migration and/or movement of individuals out of the state
  - Back of the envelope estimate of in-migration by sampling the birth cohort and seeing how many were linked to an in-state birth certificate or NBS record.

In-Migration of SCD patients

State	Confirmed Cases	Probable Cases
Georgia	8%	10%
North Carolina	14%	16%
California	10%	N/R

As reported May 2014, Children with SCD born between 2004-2008 without a matching in-state birth certificate or NBS record.

Prevalence Considerations

- In Georgia we observed approximately 155 babies born every year with SCD
- Estimates of death indicate approximately 10 patients per 1,000 die annually
- Is migration in and out of the state equal?
- Validation study of 2,000 children, 90% of confirmed cases had 6 or more encounters during the 5 year period. Using the same definition within one year of data we find similar sensitivity/specificity.

Snyder, A. B., Zhou, M., Theodore, R., Quarmyne, M.-O., Eckman, J., Peter, L. (2019). Improving an Administrative Case Definition for Longitudinal Surveillance of Sickle Cell Disease. *Public Health Reports*. 134(3): 274-281.
https://journals.sagepub.com/doi/pdf/10.1177/0033354919839072

5-year vs 14-year SCD Surveillance

Sickle Cell Disease Cases in Georgia				
	2004-2008 (RuSH)		2004 - 2017*	
Data Set	**Confirmed**	**Probable**	**Confirmed**	**Probable**
Newborn Screening	730	98	1,912	107
Augusta University (clinical)	1,218	14	1,793	23
Grady (clinical)	1,661	2	2,660	2
CHOA (clinical)	1,908	242	3,217	98
Savannah Memorial (clinical)	-	-	275	-
Medicaid/CHIP	2,986	1,993	5,407	3,441
State Health Benefit Plan**	209	215	246	205
Hospital Discharge	3,339	2,147	6,151	4,604
De-duplicated Total	4,288	3,011	7,457	5,875
Total Cases Identified		**7,299**		**13,332**

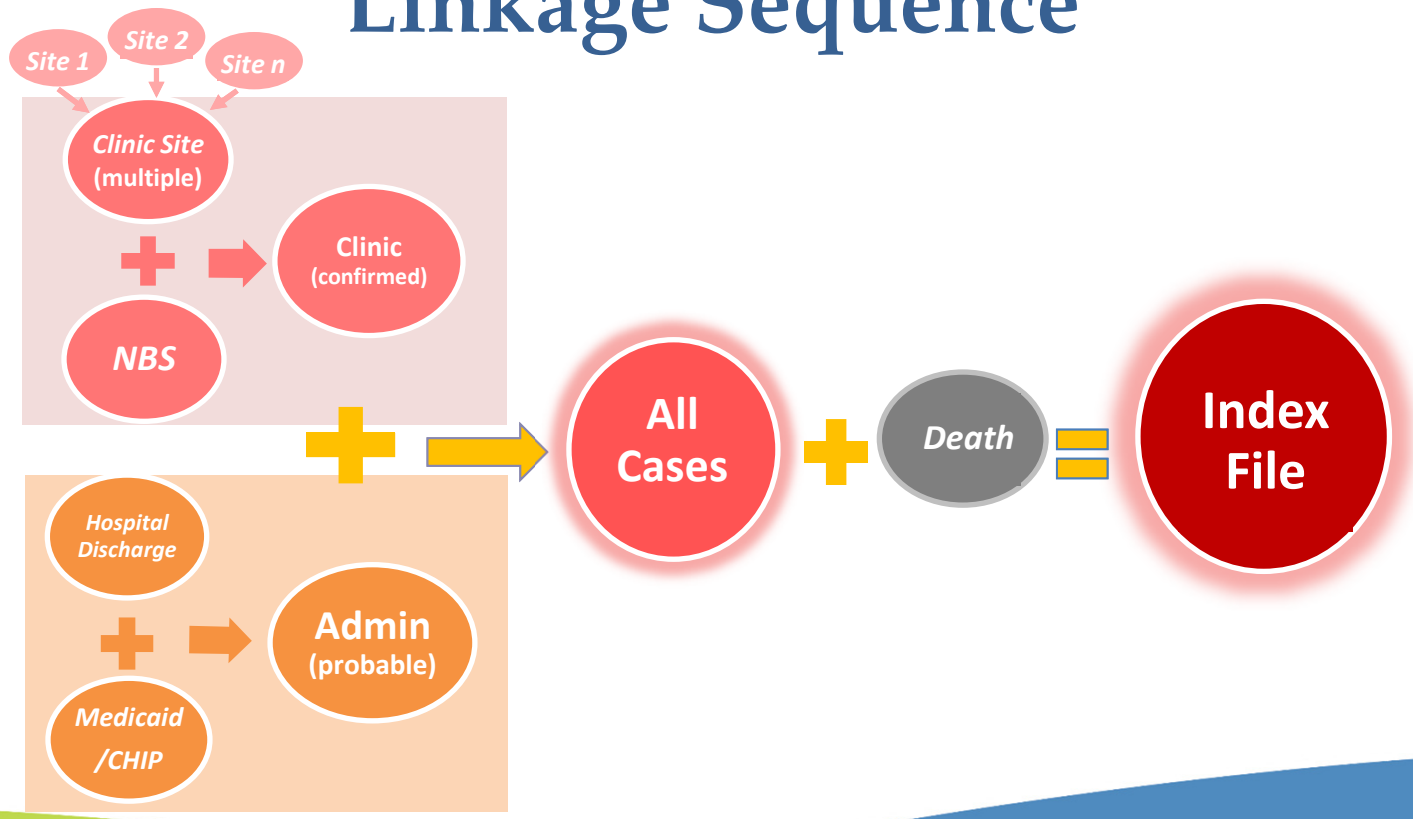
*The ending time varies from 2016– 2018 depending on the data source.

**State Health Benefit Plan data was collected for 2004-2008 time period only

Variables for Estimating Prevalence

- Has the individual had hospital/ER utilization in a given year using a GA address?
- Has the individual been enrolled in GA Medicaid?
- Is the individual actively being followed by a clinical treatment center (Need first and last date seen)
- See GA/CA annual reports on the CDC website for annual prevalence using utilization data
- Who are we still missing? Are we overestimating?

Linkage Sequence



Linkage & Analysis

- **Clean, standardize, and de-duplicate each data source by itself**
  - NBS
  - Clinical site data (multiple sources)
  - Hospital discharge data (inpatient & ER)
  - Medicaid/CHIP data
  - Death data
- Link files
- Create analytic files

Planning a data linkage system

California Sickle Cell Data Collection Program

TRACKING CALIFORNIA

INFORMING ACTION FOR HEALTHIER COMMUNITIES

Things to consider when planning your linkage system

- Early on, start thinking about how new data will be incorporated into the linkage system.
- How often do you need (vs. want) to run the linkages.
- System should be flexible and able to adjust to changing data sources and/or data fields.

Always keep your objective/surveillance goals in mind.

Things to consider when planning your linkage system: New Data

- Additional year of NBS data
  - New to the system.
  - Easiest to incorporate into underlying data but it may be a while before it can be linked to any other data source
- Clinical updates: New case vs New information
  - Has this person been reported by another clinical site?
  - Is there new or better information that could improve existing or future linkages?
- Additional years of administrative data
  - Additional year to meet probable case definition
  - Additional year for a patient to show up in the data

Things to consider when planning your linkage system: Frequency

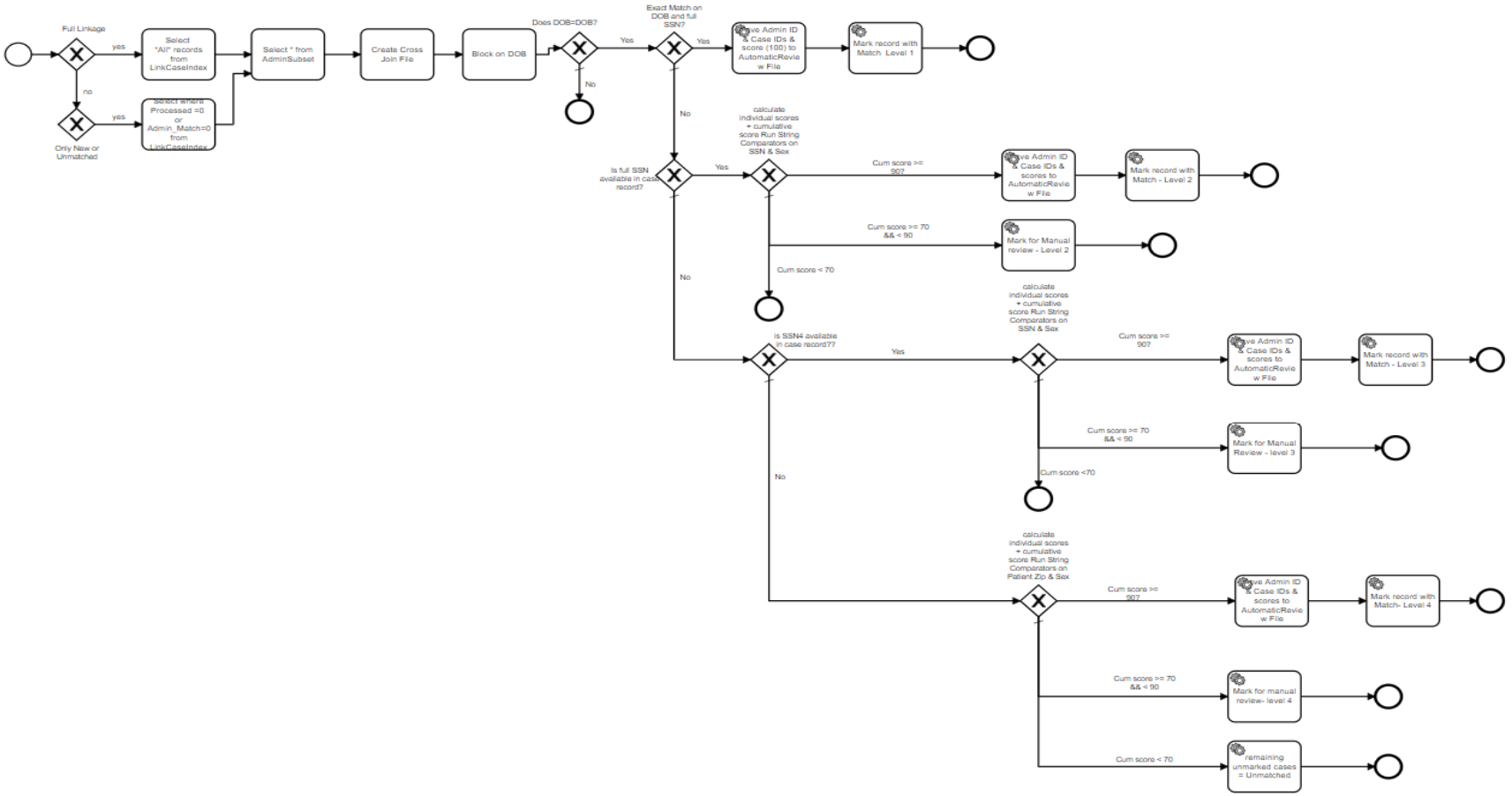
- How often do you need to run the linkages in order to meet your surveillance objectives?
  - Annually or some other set time point.
  - Once a “complete” set of data is in-house.
  - Both? Other?
- What does that linkage run like?
  - Rerun entire linkage process with all available data.
    - How will previous matches be dealt with?
    - What does the post- linkage data reconciliation process look like?
  - Run linkage on un-linked, and/or new, and/or updated case data only.
    - How are those categories defined and/or identified in the data?
  - Or something else..

Things to consider when planning your linkage system: Flexibility

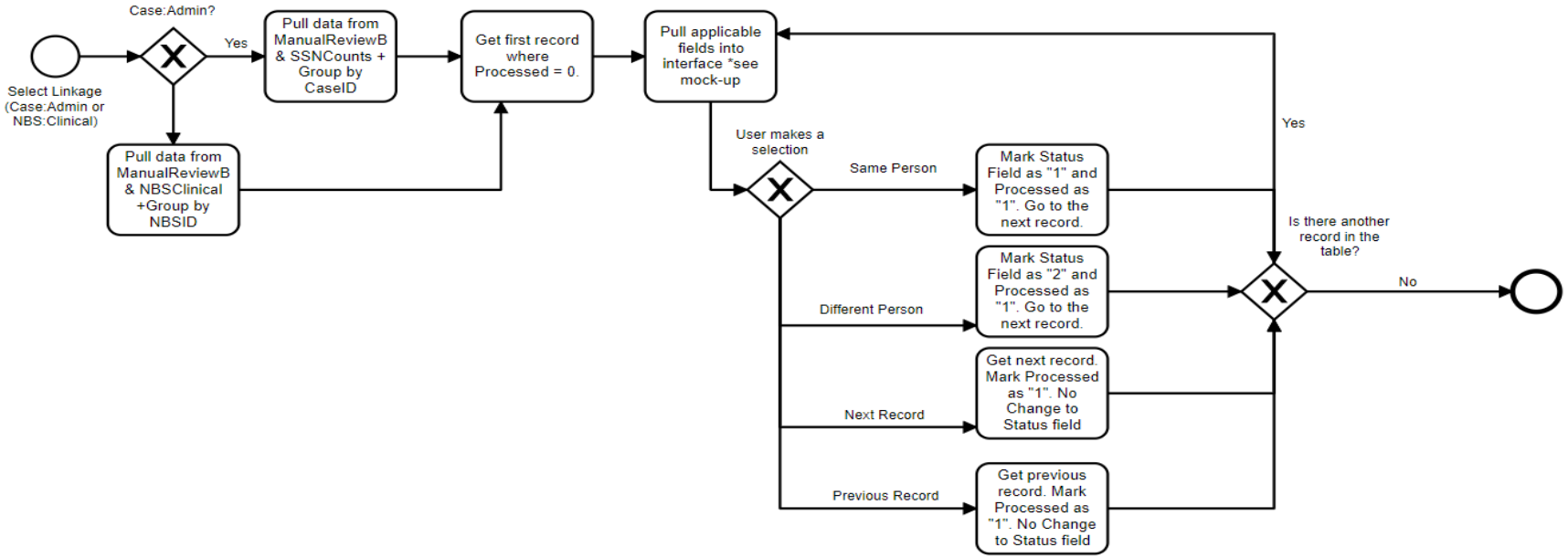
- Be flexible and expect changes
  - Changing data sources
    - Over time you may gain/lose access to data sources
  - Changing data fields
    - Increase/decrease in the identifying fields you get from data stewards can impact your linkage algorithm.
  - Evolving surveillance objectives/ tools/knowledge

- Whether your team is doing the linkage internally or another party is doing the linkage for you, it's important to keep this likelihood in mind. This may change how the linkage is coded/ programmed.

Case to Administrative Data Linkage



Manual Review



An Example: CA RuSH vs. SCDC Data

California Sickle Cell Data Collection Program

TRACKING CALIFORNIA

INFORMING ACTION FOR HEALTHIER COMMUNITIES

RuSH Era Data (2005)

- Five years of data (2004-2008) gathered and linked at once
- Early case definitions
- Counting everyone we saw, regardless of follow up time

LEVEL 1: HIGH

- Results of a state newborn screening program with confirmatory testing, or
- Clinical diagnosis by a physician with documented confirmatory laboratory testing after the newborn period.

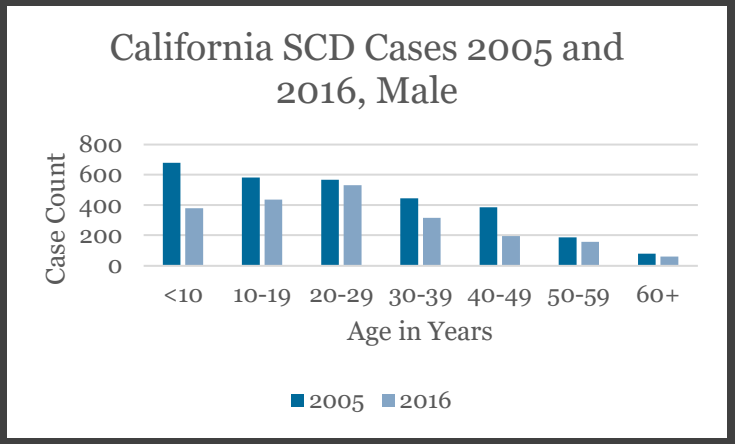
LEVEL 2: PROBABLE

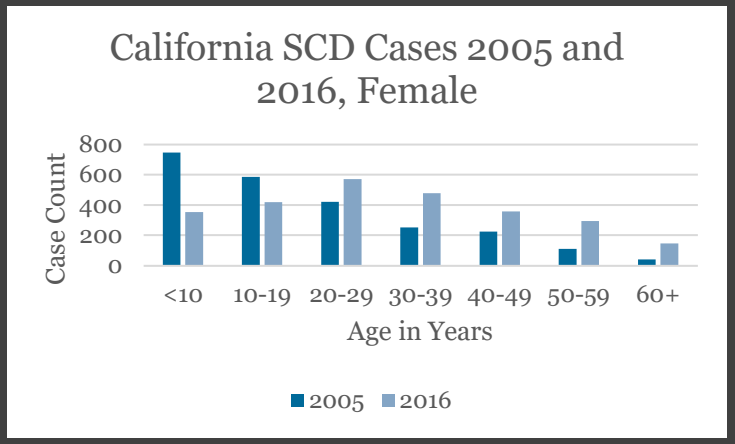
- Results of a state newborn screening program without confirmatory testing, or
- Hemoglobinopathy-related International Classification of Disease (ICD) -9 or -10 code (see following codes; excluding sickle cell trait) used during two or more separate health care encounters, plus one or more hemoglobinopathy-associated complication, treatment, or procedure (see following lists)

SCDC Era Data (2016)

- Thirteen years of data
  - More possibility of accurate linking
  - More possibility of case finding
- Simplified case definition (validated)
- Counting only those we could “see” in data in 2016

Case Counts 2005 and 2016, California





Takeaways

- The “data” (and conclusions) change with your methods and with the information you have available
- If system is constantly in development, data and conclusions will change
  - Next California annual reports have a significant increase in cases due to new data
- This is confusing for stakeholders
  - Were the previous data wrong? How do we know these data are right?
  - Where did people come from?
  - Over-explaining methodology is not helpful

Mary Hulihan (CDC): ibx5@cdc.gov
Susan Paulukonis (CA): Susan.Paulukonis@cdph.ca.gov
Angie Snyder (GA): angiesnyder@gsu.edu

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov