National Center on Birth Defects and Developmental Disabilities

DD19-1906 Capacity Building for Sickle Cell Disease Surveillance

Session 12: SCDC Data Nuts and Bolts

June 4, 2020

SCDC

SICKLE CELL DATA COLLECTION

Data Nuts & Bolts

Angie Snyder, PhD

June 6, 2020

Georgia Health Policy Center

Georgia State University | Andrew Young School of Policy Studies

Balancing Data Activities

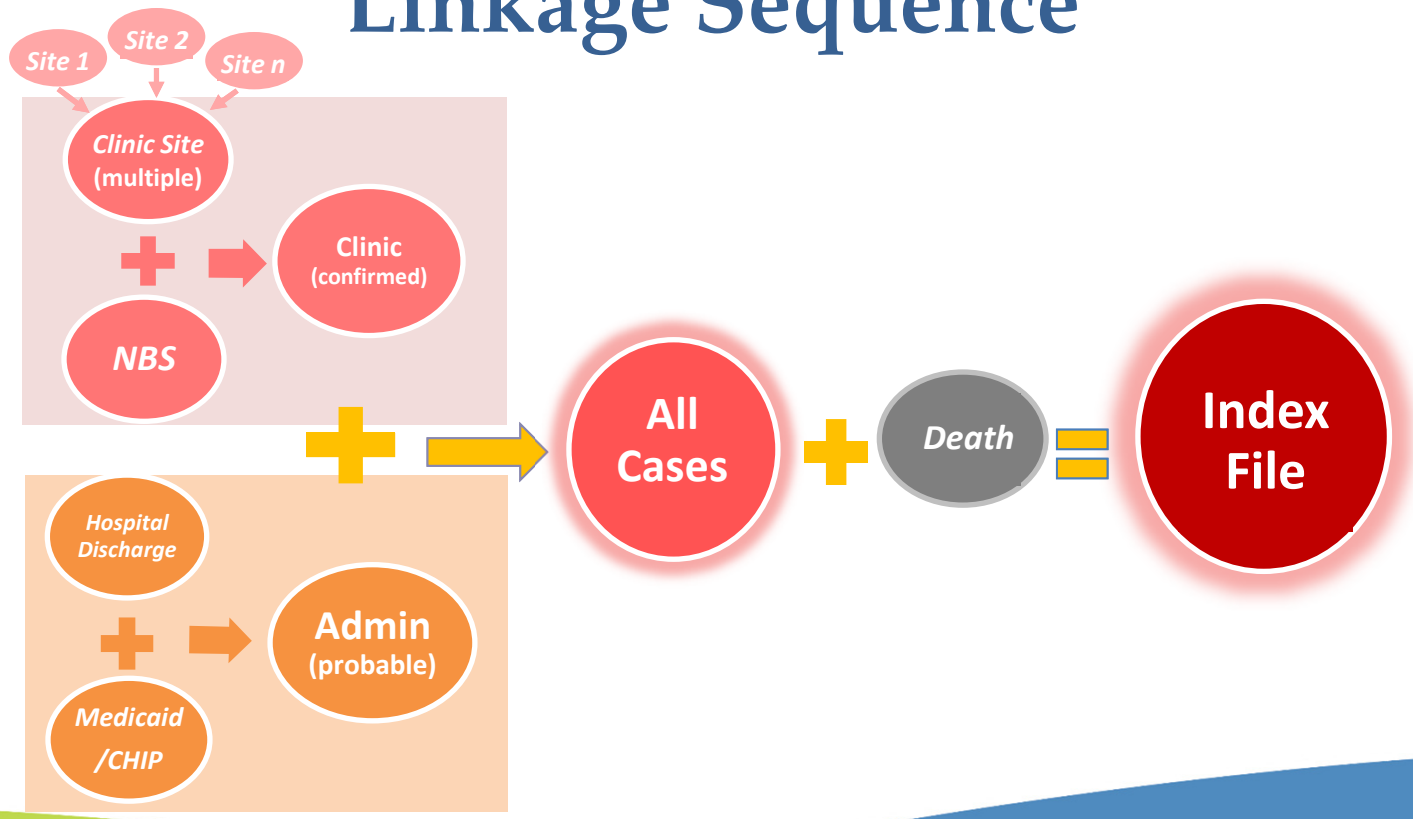
Building the Surveillance Dataset

- Random delivery of data (annual requests)
- Need to clean and deduplicate each new data delivery *by source* first
- More work for the longitudinal datasets (second round of finder files)
- Link across datasets (GA: wait until we have an entire new year of confirmed cases and/or probable cases before this step)
- Need for flexibility; process is never perfect or complete (iterative)

Data Analysis

- Up-to-date status of data by source is necessary to determine inclusion criteria
- Examples…
  - Incidence reports may only require access to the latest NBS data
  - Annual prevalence reports require all data sources to be up-to-date for the year in question
  - Some research questions may require access to certain years of Medicaid or Hospital Discharge Data

Linkage Sequence



Coordination of Data Activities

- The larger the data team & the more specialized their data activities the more systematic processes/protocols/documentation need to be to promote quality use of the data.
- Data analyses have dependencies on the availability of cleaned and linked data; building the surveillance dataset has dependencies related to the timing and quality of data deliveries.
- The work load is not consistent across time and some is outside of your control.

Key Considerations: Data Analyses

- Longitudinal data is powerful but includes temporal variables and follow up periods that must be defined for each research question.
- Key temporal variables: age, insurance status, address/zip/county/state of residence
- We know more about individuals with higher levels of utilization; from the hospital discharge data all ER/inpatient admissions and from Medicaid all outpatient visits and prescription fills as well; ex. to study prescription refills for HU must limit the analyses to Medicaid enrollees

Examples/Considerations

- Insurance status: Medicaid (need the eligibility file)
  - Include only individuals with 11/12 months of coverage during the year or 48/60 months of coverage for a 5 year follow-up period.
  - Some studies include everyone with any coverage during the follow-up period and report results in person/months.
- Age
  - Children reaching their 3rd birthday during a 5 year window to study receipt of TCD and/or PPV.
  - 5 year mortality for adults surviving to 40 years of age. Define age groups at start, middle, or end of follow-up period?

Examples/Considerations cont.

- Residence by County/zip code/state:
  - Different levels of accuracy from different data sources, some may contain PO Boxes or missing data
  - Most difficult for prevalence studies—Do you count confirmed cases that have no utilization in the past 5 years as still residing in the State?
  - Very interesting data to study allocation of resources geographically—can use it to study distance to a NBS follow-up Center, can use it to look at how far patients are travelling to the ER, etc.

Nuts and Bolts: Finder Files and Flow Charts

California Sickle Cell Data Collection Program

TRACKING CALIFORNIA

INFORMING ACTION FOR HEALTHIER COMMUNITIES

FINDER FILES FOR SURVEILLANCE

Finder Files in Disease Surveillance

- Not all claims and encounters for someone with SCD are about SCD
- No marker or variable that identifies as having SCD
- How do you find information on someone that isn't marked as SCD in claims or encounter data?

Finder Files for Disease Surveillance

- A finder file contains linkable identifiers of known or suspected cases of SCD
- Sent to data steward to allow them to search for these individuals in their data
- Sources that report all their data, no finder file needed
  - Ex. Full hospital discharge file, claims database with full access
- Sources that report only identified, confirmed cases of SCD, no finder file needed
  - Ex. Newborn screening, clinical case reports
- In California, we use a finder file for Medicaid claims and will use one for Medicare claims

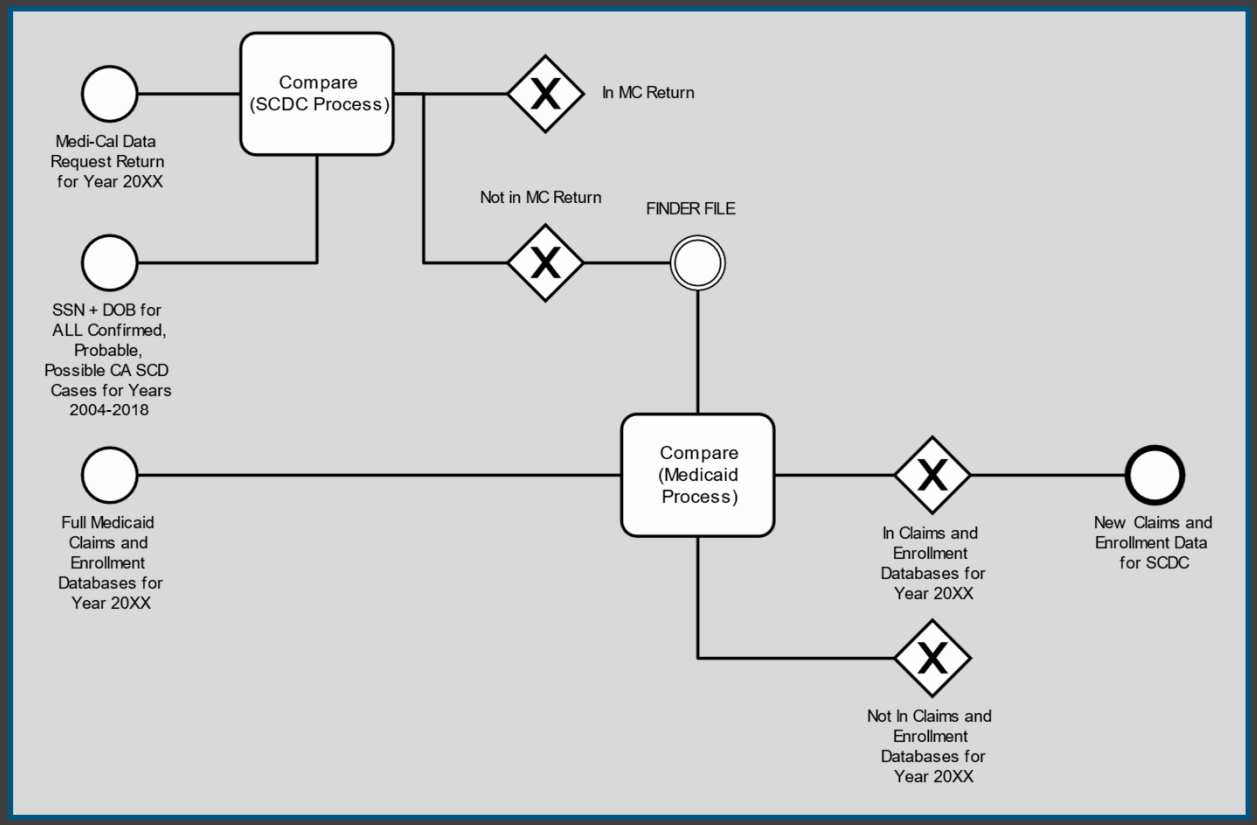
Medicaid Finder File – Missing Information About Known Persons

- Initial data request:
  - Per calendar year
  - If person enrolled at any time in Medi-Cal AND
  - Had one or more SCD or SCT ICD codes (any order) in that year THEN
  - Send all claims for that year

- Follow up finder file:
  - SSN + DOB OF
  - Persons with one or more SCD or SCT codes in ANY data source in that calendar year or any prior calendar year (2004-2018) UNLESS
  - That SSN + DOB is included in the data request returned file

- Finder file linked back to all Medicaid claims and enrollment returns:
  - Claims and enrollment information for anyone in finder file that links

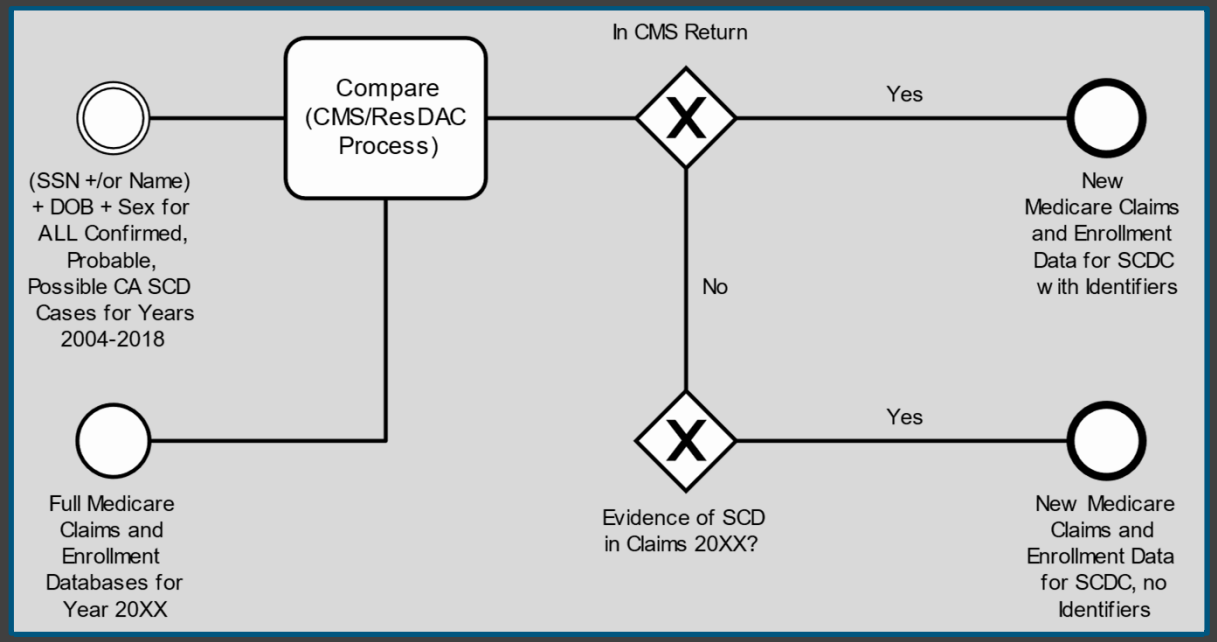
Medicaid Finder File



Medicare Finder File – New Case Identification and All Claims

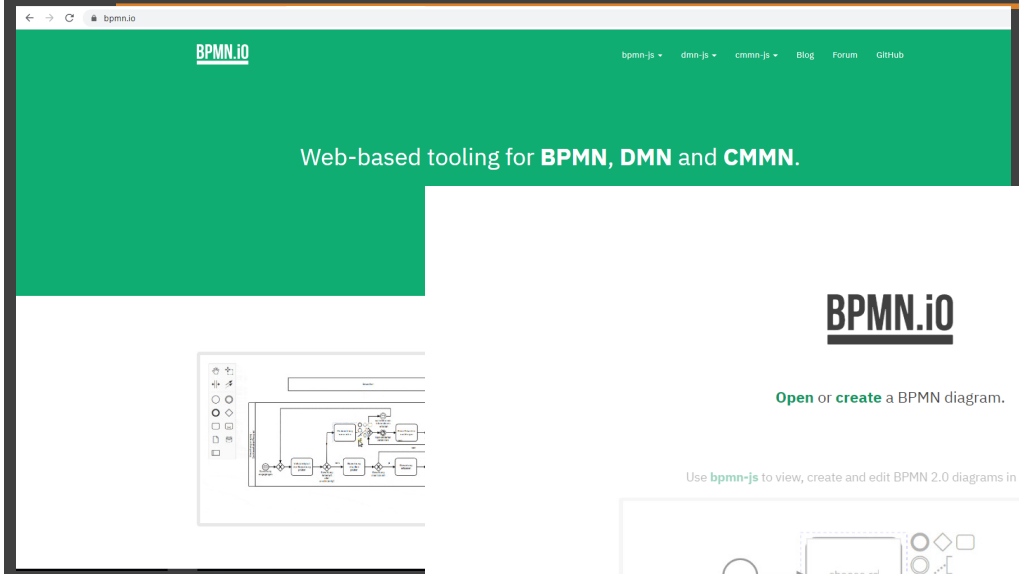
- Initial data request:
  - Per calendar year IF
  - Linked to finder file and residing in CA OR
  - Not linked to finder file but reside in CA and have >0 SCD ICD codes
- Initial finder file:
  - (SSN and/or Name) AND Sex AND Date of birth OF
  - Persons with one or more SCD or SCT codes in ANY data source in that calendar year or any prior calendar year (2004-2018)
- Returns:
  - Claims and enrollment info for anyone in finder file that links (have identifiers)
  - Claims and enrollment info for anyone with SCD ICD codes in that year (no identifiers)

Medicare Finder File

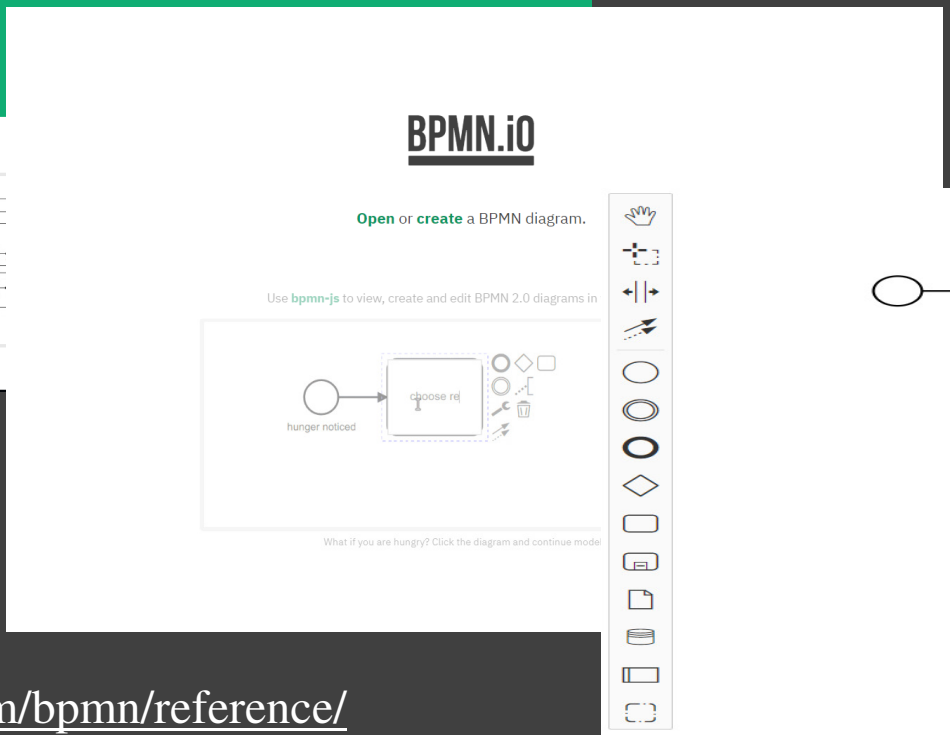


FLOW CHARTS FOR DATA LINKAGE

Business Process Modeler- BPM

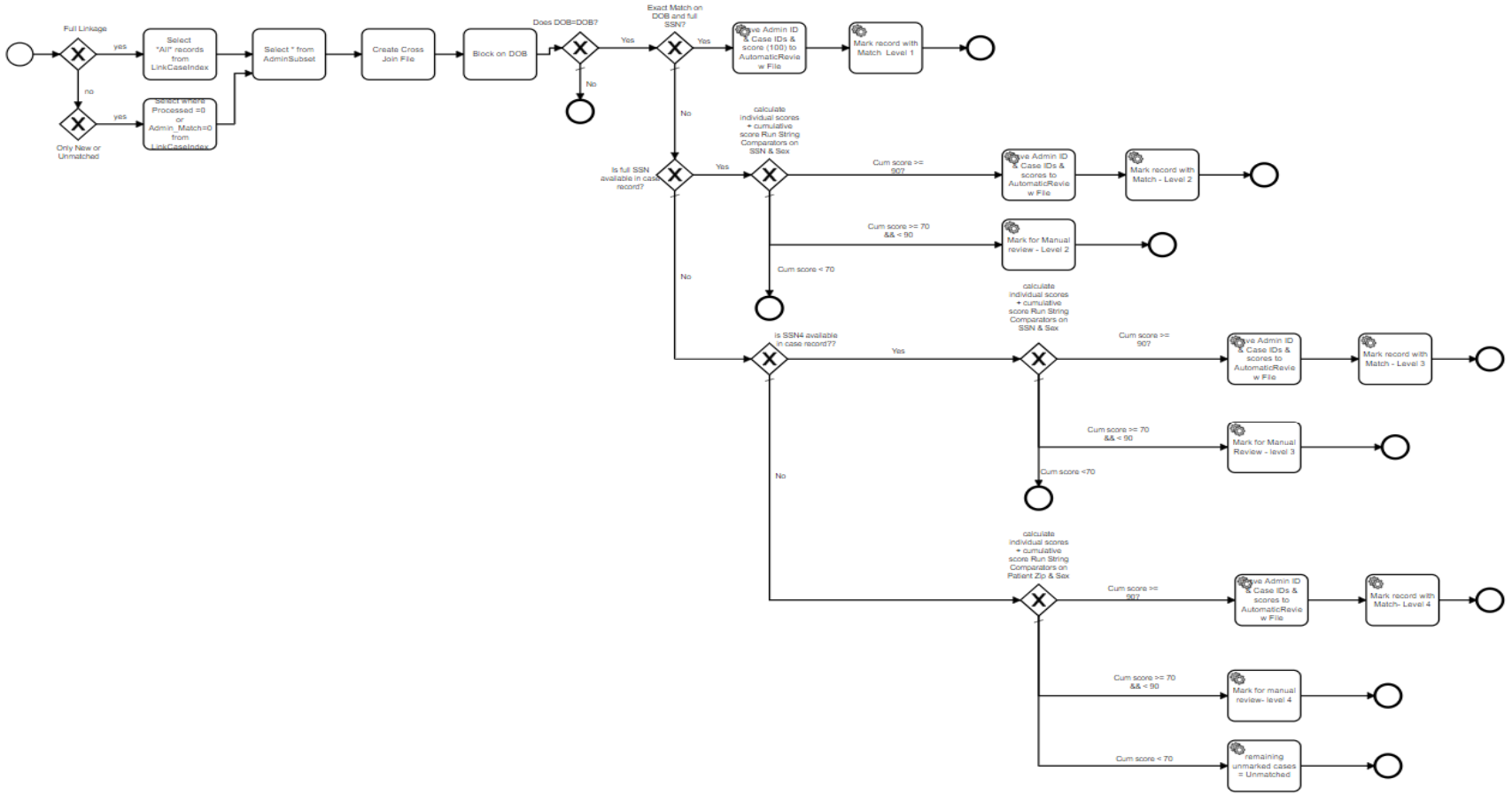


https://demo.bpmn.io/

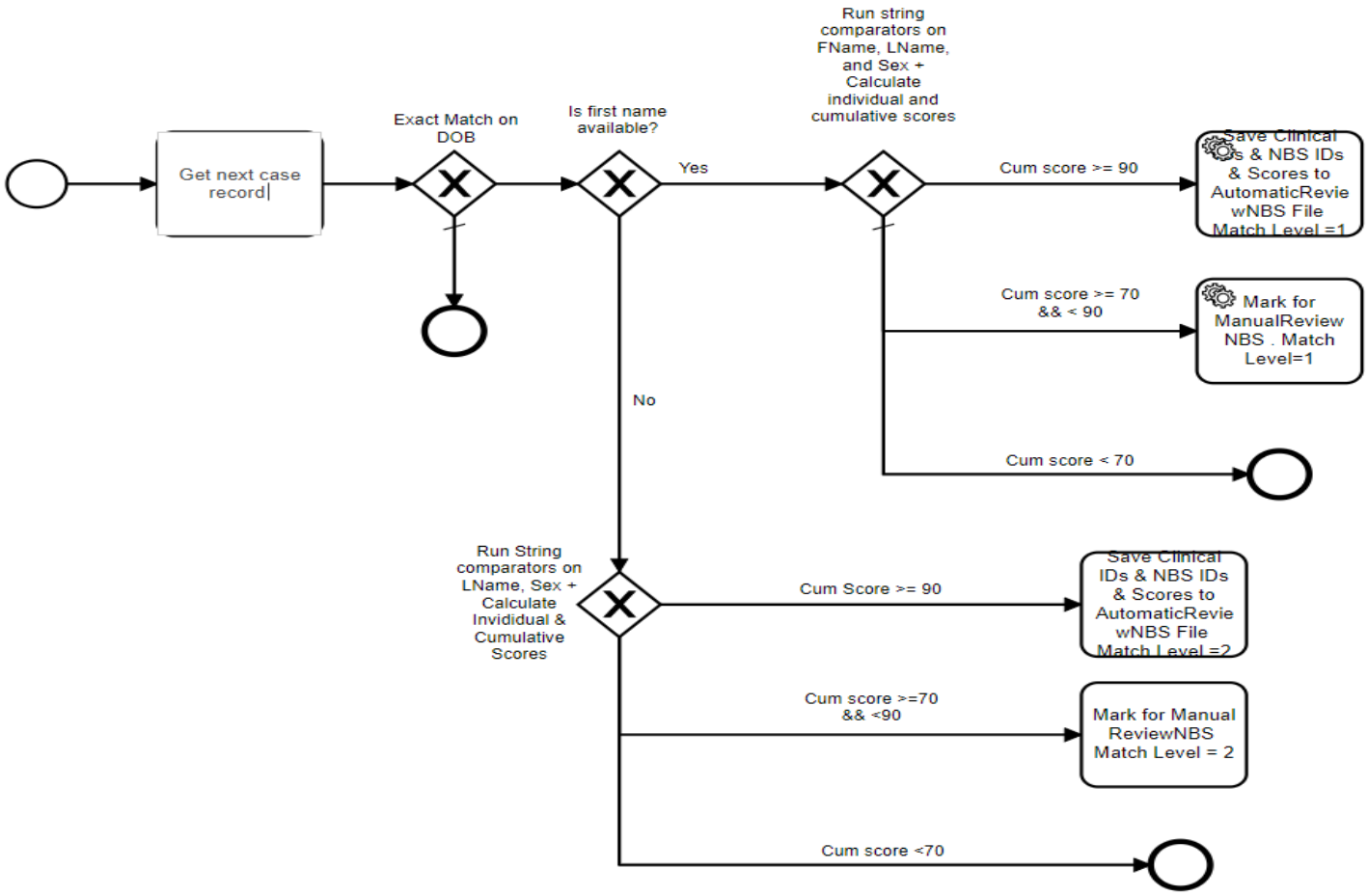


https://camunda.com/bpmn/reference/

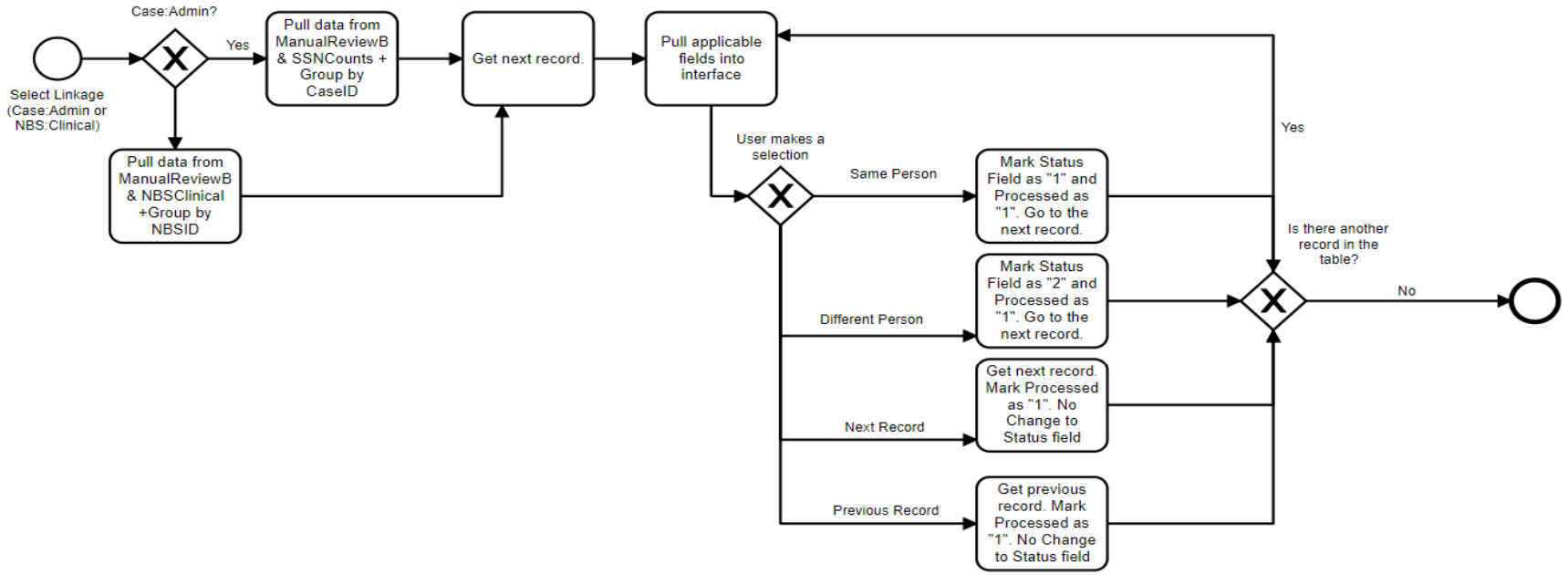
Case to Administrative Data Linkage



NBS:Clinical Linkage



Manual Review



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