

# Management of Hepatitis C cases in New Jersey Department of Health Surveillance System

June 8, 2022



# Presentation overview

- **Snapshot of New Jersey**
- **Burden of Hepatitis C in New Jersey**
- **New Jersey Communicable Disease Service surveillance system, CDRSS**
- **Hepatitis C and CDRSS**
- **Data Management: Hepatitis C Registry and PS2103**
- **Data Management in CDRSS**
- **Manual Deduplication using Link Plus**
- **Benefits and Limitations of Link Plus**
- **Manual Deduplication: Link Plus Process**
- **Computing Duplicates in Link Plus**
- **Link Plus: Matching Variables**
- **Alternative System for deduplication of cases**

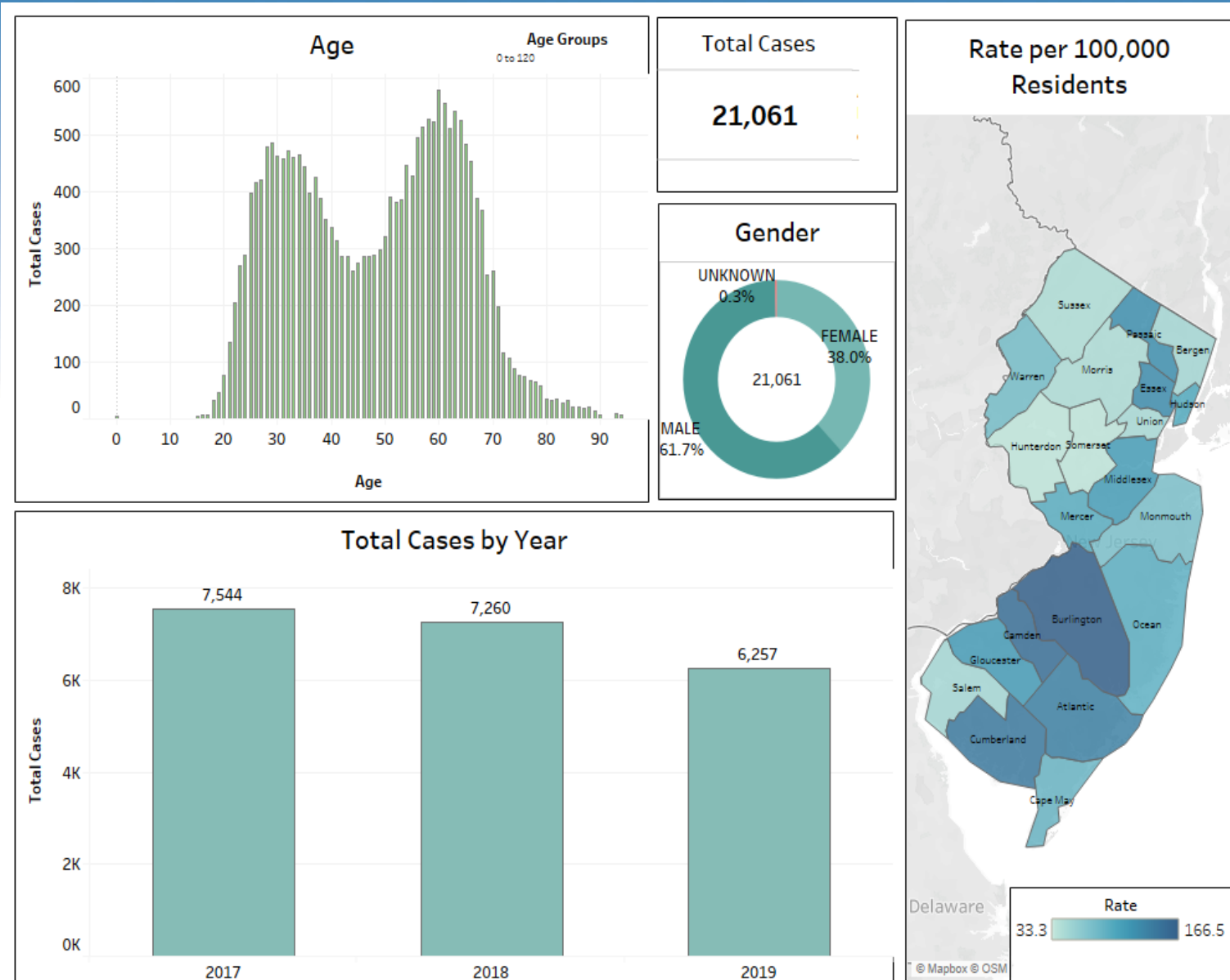
# Snapshot of New Jersey

- **New Jersey located in the Mid-Atlantic region and borders two metropolitan cities; New York City and Philadelphia**
  - Most densely populated state
  - Population ~9.0 million; 22% foreign -born
  - Jurisdictions of 21 counties and 565 municipalities
  - Trenton is the capital; located in Mercer county
  - Most people live close to the metropolitan cities and the Jersey shore



# Burden of Hepatitis C in New Jersey

- Hepatitis C Acute, Chronic and Perinatal are reportable
- About **9000** reports for Hepatitis C received annually
- > **6000** newly reported Acute and Chronic cases yearly; a slight decrease is observed
- New reports in multiple generations
- Highest rate of Hepatitis C (HCV) mainly in the most populated areas



# Communicable Disease Reporting and Surveillance in New Jersey

- CDRSS, NJDOH official disease surveillance system with over 100 reportable and non-reportable diseases and work-related conditions
- Integrated electronic lab reporting (ELR) system that is patient-centric and web-based
- Has advanced tools for data analysis, visualization, mapping and exportable reports
- Has the capacity to manage outbreak investigations, schools and other facility surveillance
- Structured to accept positive and non-positive reports according to case definitions
- Industry-standard security features

# Hepatitis C and CDRSS

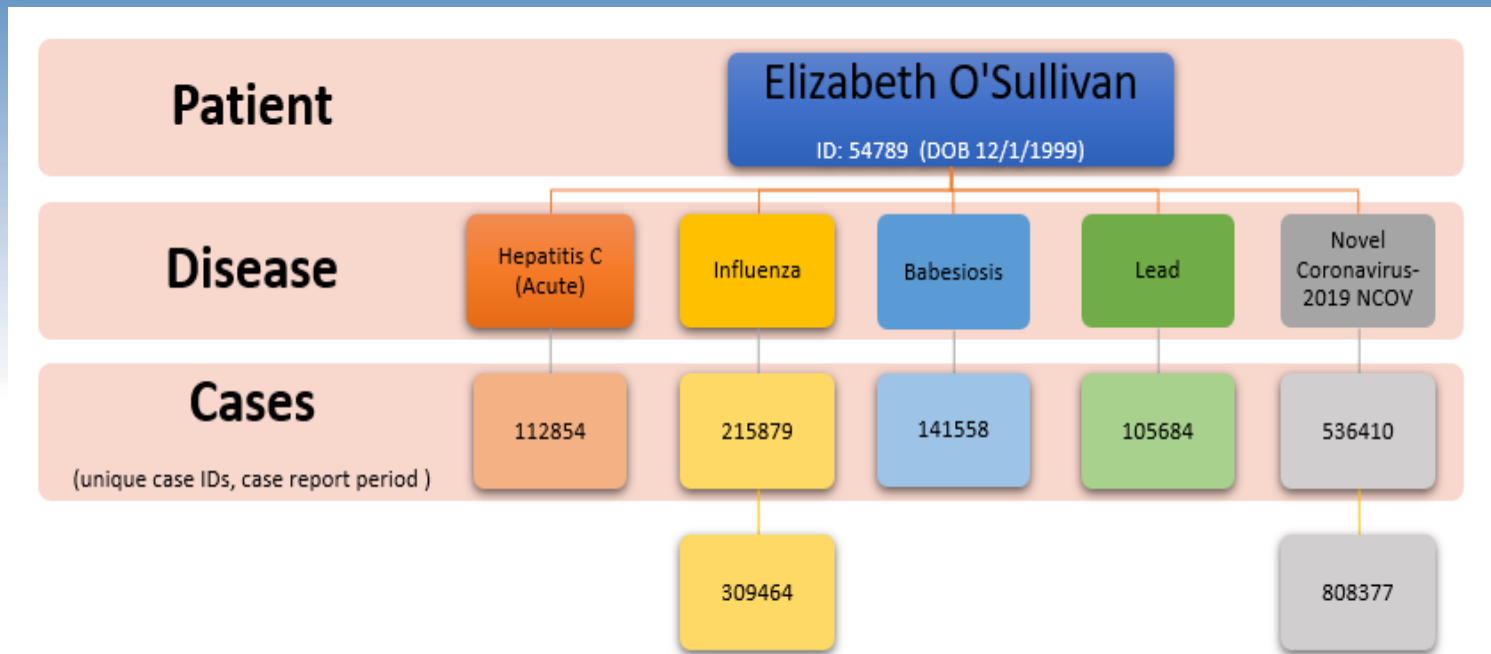
- At least **90% of HCV labs are reported electronically** in CDRSS
- Some manual entry from paper laboratory reports
- HCV Investigators provide updates on HCV reports for residents in their jurisdiction in CDRSS
- Multiple person and case merges per year not automatically identified
- **2779** cases manually merged in CDRSS, in 2021

# Data Management: Hepatitis C Registry (PS21-2103)

- As part of the grant requirement, overall data quality is essential to maintain NJDOH HCV registry and avoid duplication of cases
- Management of jurisdictional laboratory-based case cascade for hepatitis to monitor testing to cure efforts
- New Jersey CDRSS has the capacity to capture negative lab reports – thus increase the need for deduplication, as more laboratories increase HCV RNA reflex testing
- NJDOH runs the deduplication program quarterly to reduce incidence of duplicates
  - Manual process when demographics are not matched one to one
  - Takes about 2.5 hours a week

# Data Management in CDRSS

- Managing a patient-centric system
  - ELR rules for incoming reports
  - Disease/Condition algorithms: Confirmed, Probable, Not a Case
  - New cases or reinfections?
- Duplicate Persons and Cases
  - ELR deduplication
  - Deduplication by system merge: One to one demographics
  - Manual deduplication: Link Plus Software**



How do duplicates get into the system?

## Data Entry Errors (99% of errors)

- Demographics Discrepancies:  
Name: Elizabeth/Liz/Elisabeth, O'Sullivan/Osullivan  
DOB: 12/1/1999 – 12/11/1999



# Manual Deduplication: Using Link Plus Software

- Link Plus is a CDC software originally designed as cancer record matching and deduplication tool for cancer registries
- Computes probabilistic record linkage (matching) and has the capacity to detect duplicates using a choice of two phonetic coding systems (Soundex and NYSIIS)
- The program works for any number of records with a limit of 4.5 to 4.8 million records
- Provides Best Match options of all potential data matches or duplicates based on user defined threshold (cut off value)
- Link Plus is FREE: \$0.00



# Benefits and Limitations of Link Plus

## Benefits

- Designed especially for cancer registry work but can be used with ANY DATA
- Probabilistic math and computation largely hidden from user
- System default values provided deduplication
- Software includes robust “Help Menu” and test examples
- Output results are clear and accessible to new or advanced users
- 🎵 **FREE FREE FREE FREE...FREE** 🎵

## Limitations

- Creating external data files to be imported
- Default values may need to be changed multiple times
- Duplicates identified will require manual review and merging

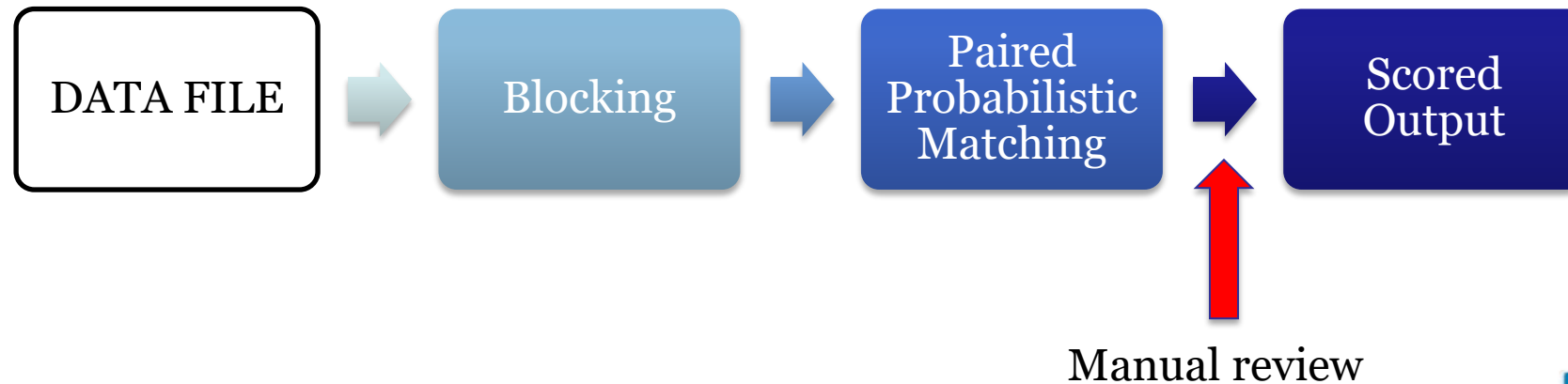
# Manual Deduplication: Link Plus Process

**Link Plus is REALLY EASY to use!**

Step 1: Select data file

Step 2: Select blocking variables and phonetic system

Step 3: Click run...and the software will do the rest!



# Computing Duplicates in Link Plus

Data for Deduplication



Blocking



Matching and Scoring



**High Score**



**Gray Area**



**Low Score**

# Link Plus: Matching Variables

## Demographic data

- First name
- Last name
- Middle name
- Date of birth
- Social Security Number (*not collected in CDRSS*)



## Other matching variables

- Zip code
- Generic ID
- Exact string data
- Value specific data



### Select ID Variables

	DISEASE
	PATIENT_ID
*	CASE_ID
	GENDER
	COUNTY_1

## Other variables in data file

	Data Item	Matching Method
	LAST_NAME	Last Name
	FIRST_NAME	First Name
	DOB	Date
*		

Blocking and matching variables

Cancel

☒ Direct Method

20

C:\RegPlus\LinkPlus\report\LinkageReport.txt

Remove

Linkage Process is Done

Start time=6/6/2022 12:00:20 PM      End time=6/6/2022 12:05:56 PM      Total time=0:05:36

Link Plus is preparing the linkage configuration, and will begin reading data files shortly

Reading data

Computing scores for pairs of records in each block.

The result has been saved to: 'W:\USD\PDISEASE INFORMATION\NOVEL CORONAVIRUS\INVESTIGATION\PUICONFIRMED CASES\Case Summary Reports\Death and LTC Line Lists Report\EDRS\Duplicate deaths\dupicates.txt'

Computing score for each block...      Block      33773

Conduct Manual Review Now      Close

Run



C:\RegPlus\LinkPlus\Configuration\HepC Dedup.cfg

Data = X:\DISEASE INFORMATION\HEPATITIS C\DATA &amp; STATS (BY YEAR)\Dedup\2021 to 2017 with DOB.csv

Number of Records = 38365

## Linking Process

Direct method is employed

Cutoff Value=20.0

Field for Blocking

soundex on LAST\_NAME

soundex on FIRST\_NAME

soundex on DOB

## Matching Parameters

Matching Field	m-prob	u-prob	agree	disagree	matching method
DOB	0.96000	0.00008	8.54606	-2.92987	date
LAST_NAME	0.97000	0.00055	6.80272	-3.19130	last name
FIRST_NAME	0.96000	0.00354	5.10070	-2.92672	first name

Cutoff Value=20.0

m-prob: The probability that a matching variable agrees given that the comparison pair being examined is a match

u-prob: The probability that a matching variable agrees given that comparison pair being examined as a non-match

agree: The agreement weight assigned for an agreement on a given matching variable

disagree: The disagreement weight assigned for a disagreement on a given matching variable

matching method: The method used for computing the weight on a given matching variable.

## Match Report

"Sorted by score descently (higher score, higher likelihood of being a matched pair)"

Number of Matched Pairs = 1014

Linkage ID	Set ID	Record #	DISEASE	PATIENT_ID	CASE_ID	DOB	LAST_NAME	FIRST_NAME	COUNTY_1	MUNICIPALITY_1	Class	Score
1	1013	12519	HEPATITIS C	2838663	1220987	12/1/1999	OSULLIVAN	ELIZABETH	PASSAIC	PASSAIC CITY		
1	1013	17802	HEPATITIS C - CHRONIC	3143925	1373103	12/11/1999	O' SULLIVAN	ELISABETH	PASSAIC	PASSAIC CITY	7	27.1
2	1012	13423	HEPATITIS C - CHRONIC	3034218	1311037	7/4/2002	PATEL	SITUKKA A	HUDSON	SECAUCUS TOWN		
2	1012	38293	HEPATITIS C - ACUTE	3639071	1700103	7/4/2002	PATEL	SITTUKA	ESSEX	NEWARK CITY	8	26.8
3	1011	17582	HEPATITIS C - CHRONIC	735475	1410118	1/1/1961	SMITH	JOHN	HUDSON	JERSEY CITY		
3	1011	38174	HEPATITIS C - CHRONIC	9900128	5917396	1/1/1961	SMITH	JOHN	HUDSON	JERSEY CITY	10	26.8



# Alternative system – Match\*Pro

- Match\*Pro is a Java-based application developed by Information Management Services, Inc (IMS) for probabilistic record linkages .
- Also based on the Fellegi and Sunter model
- It has a user-friendly tabbed interface for the configuration of linkages and the manual review of uncertain matches
- Match\*Pro can run on any fixed-width or delimited data file as Link Plus but can also support XML files
- Match\*Pro is flexible and allows users to filter incoming data, specify blocking and matching methods, define unknown values, adjust weights, identify cutoff points, and more during the deduplication process
- **Match\*Pro has 9 distinct blocking strategies**



# Match\*Pro vs Link Plus?

- Match\*Pro is more advanced than Link Plus and can handle more parameters for a complicated probabilistic record linkage or deduplication
- Users have access to **more tools** during the manual review process including a Query Editor
- Probabilistic math and computation visible to user during deduplication process
- Match\*Pro has the option to filter input file before deduplication process
- Also FREE but requires registration before download
- Overall, Link Plus is easier to use for everyday work

# How to get Link Plus

**Download from CDC National Program of Cancer Registries (NPCR)**

**<https://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm#installation>**

- Follow installation directions

**Questions ?**

**THANK YOU**

# Contact Information

*Bernice Carr, M.P.H., M.S.*

Epidemiologist- Hepatitis C  
Communicable Disease Service  
New Jersey Department of Health  
(609) 826.5964

[Bernice.Carr@doh.nj.gov](mailto:Bernice.Carr@doh.nj.gov)

*Mojisola Ojo, M.P.H*

Epidemiologist- Data Coordinator  
Communicable Disease Service  
New Jersey Department of Health  
(609) 826.5964

[Mojisola.Ojo@doh.nj.gov](mailto:Mojisola.Ojo@doh.nj.gov)