

# TX PHBPP: Finding More Cases with Fuzzy Matching

Virtual Perinatal Hepatitis B Prevention Summit, June 2020

Holly Hughes Garza, DVM, MPH

Immunization Unit, ACE Group, Epidemiology Team Lead

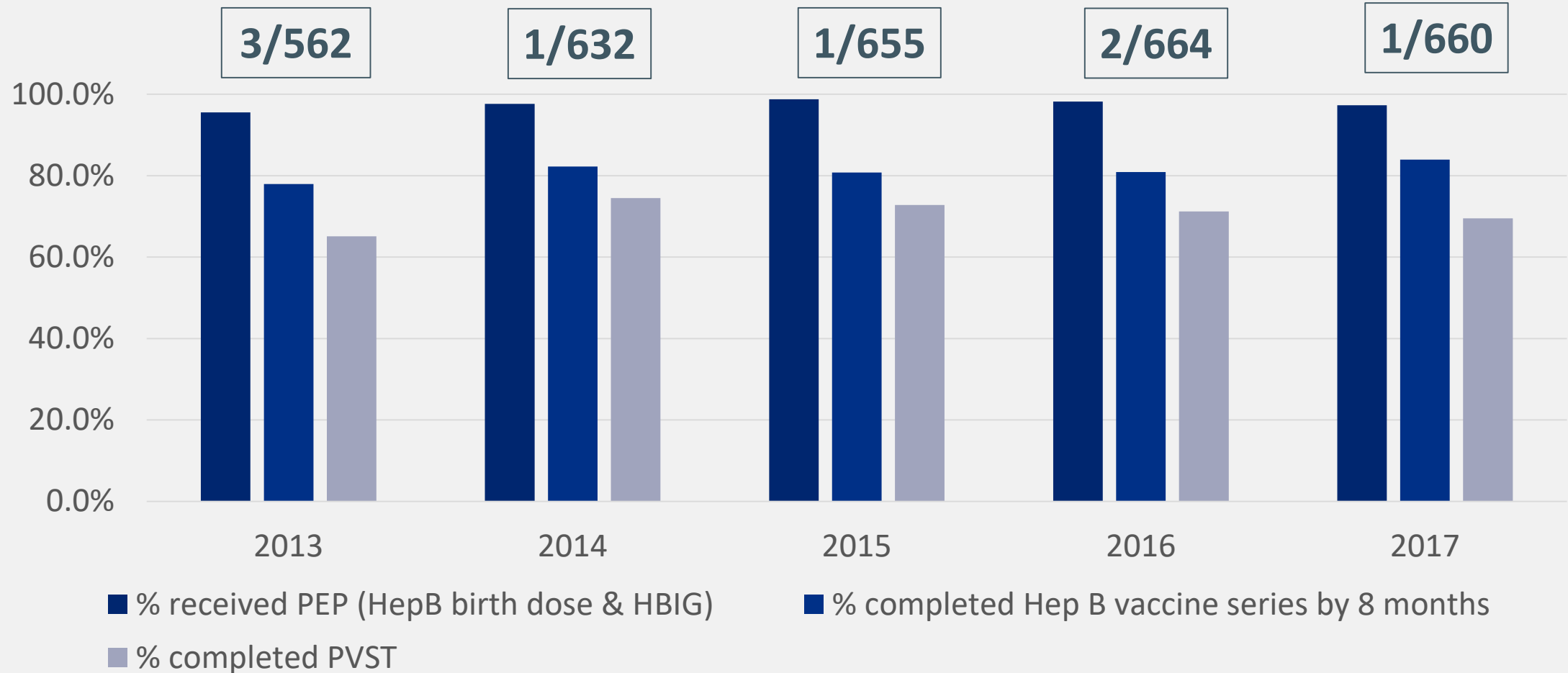


TEXAS  
Health and Human  
Services

Texas Department of State  
Health Services

# 2013-2017 Birth Cohorts

HBsAg+ infants/  
Exposed infants



# The Goal

- Enroll All Texas Hepatitis B Positive Pregnancies in PHBPP
  - CDC estimated at least 880 infants born to HBsAg+ women in Texas in 2017
  - Texas enrolled 660 cases (**75%**)
    - \*Not including City of Houston and San Antonio



# Matching Data from Multiple Sources

- Perinatal Hepatitis B Prevention Program (PHBPP) Database
- Vital Statistics Unit (VSU) Birth Records
- NEDSS – Lab results from expectant mothers



# How PHBPP Uses Birth Records

- Monthly file with data for all births
- Identify unreported moms with hepatitis B infection during pregnancy
  - Investigate!
- Match to mother-infant pairs already identified and fill in missing information
  - E.g. Infant's first name



# VSU Medical Data Worksheet

VS-109.2 (09/11)

### Medical Data Worksheet for Child's Birth Certificate

This form to be completed by hospital staff. This data will be used to populate the medical data portion of the birth certificate for the newborn. The medical data is required to be reported within five days of the birth. [HSC §192.003]

PATIENT REFERENCE:	
MOTHER MR# _____	NEWBORN MR# _____
MOTHER'S NAME _____	NEWBORN NAME _____
MEDICAID# _____	DOB _____
DELIVERING DR _____	DATE AOP SENT _____
MOTHER TRANSFERRED _____	SOURCE OF PAYMENT FOR DELIVERY _____
<input type="checkbox"/> Born at Facility <input type="checkbox"/> Born En Route <input type="checkbox"/> Foundling <input type="checkbox"/> Home Birth	

<b>Prenatal Care</b> <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown Date of First Visit: ____/____/____ Date of Last Visit: ____/____/____ Total Number of Prenatal Visits for this Pregnancy: _____ Date Last Normal Menses Began: ____/____/____	<b>Source of Prenatal Care</b> (check all that apply) <input type="checkbox"/> None <input type="checkbox"/> Midwife <input type="checkbox"/> Hospital Clinic <input type="checkbox"/> Other, Specify _____ <input type="checkbox"/> Public Health Clinic <input type="checkbox"/> Unknown <input type="checkbox"/> Private Physician
<b>Pregnancy History</b> <b>Live births now living</b> (Do not include this birth. For multiple deliveries, do not include the 1 <sup>st</sup> born in the set if completing this worksheet for that child. If none enter '0'): _____ <b>Live births now dead</b> (Do not include this birth. For multiple deliveries, do not include the 1 <sup>st</sup> born in the set if completing this worksheet for that child. If none enter '0'): _____ Date of last live birth: ____/____/____ MM YYYY <b>Number of other pregnancy outcomes</b> (include fetal losses of any gestational age. If this was a multiple delivery, include all fetal losses delivered before this infant in the pregnancy. If none enter '0'): _____ Date of last other pregnancy outcome: ____/____/____ MM YYYY	<b>Risk Factors in this Pregnancy</b> (check all that apply) <b>Diabetes</b> <input type="checkbox"/> Prepregnancy (diagnosis prior to this pregnancy) <input type="checkbox"/> Gestational (diagnosis in this pregnancy) <b>Hypertension</b> <input type="checkbox"/> Prepregnancy (chronic) <input type="checkbox"/> Gestational (PIH, preeclampsia) <input type="checkbox"/> Eclampsia <input type="checkbox"/> Previous preterm birth <input type="checkbox"/> Other previous poor pregnancy outcome (includes perinatal death, small-for-gestational age/intrauterine growth restricted birth) <input type="checkbox"/> Pregnancy resulted from infertility treatment <input type="checkbox"/> Fertility-enhancing drugs, artificial insemination or intrauterine insemination <input type="checkbox"/> Assisted reproductive technology <input type="checkbox"/> Mother had a previous cesarean delivery If yes, how many? _____ <input type="checkbox"/> Antiretrovirals administered during pregnancy or at delivery <input type="checkbox"/> None of the above

<b>Infections Present and/or Treated During Pregnancy</b> (check all that apply) <input type="checkbox"/> Gonorrhea <input type="checkbox"/> Hepatitis B <input type="checkbox"/> Syphilis <input type="checkbox"/> Hepatitis C <input type="checkbox"/> Chlamydia <input type="checkbox"/> None of the above	<b>Infections Present and/or Treated During Pregnancy</b> (check all that apply) <input type="checkbox"/> Gonorrhea <input checked="" type="checkbox"/> Hepatitis B <input type="checkbox"/> Syphilis <input type="checkbox"/> Hepatitis C <input type="checkbox"/> Chlamydia <input type="checkbox"/> None of the above
--	---

<b>HIV Test</b> HIV test done Prenatally <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown (check all that apply) <input type="checkbox"/> First Trimester <input type="checkbox"/> Third Trimester	1 of 2 HIV test done at delivery <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown Infant tested for HIV at birth <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
--	---

## Infections Present and/or Treated During Pregnancy (check all that apply)

- Gonorrhea
- Hepatitis B
- Syphilis
- Hepatitis C
- Chlamydia
- None of the above

# Fuzzy Matching?

“Sounds like” fun

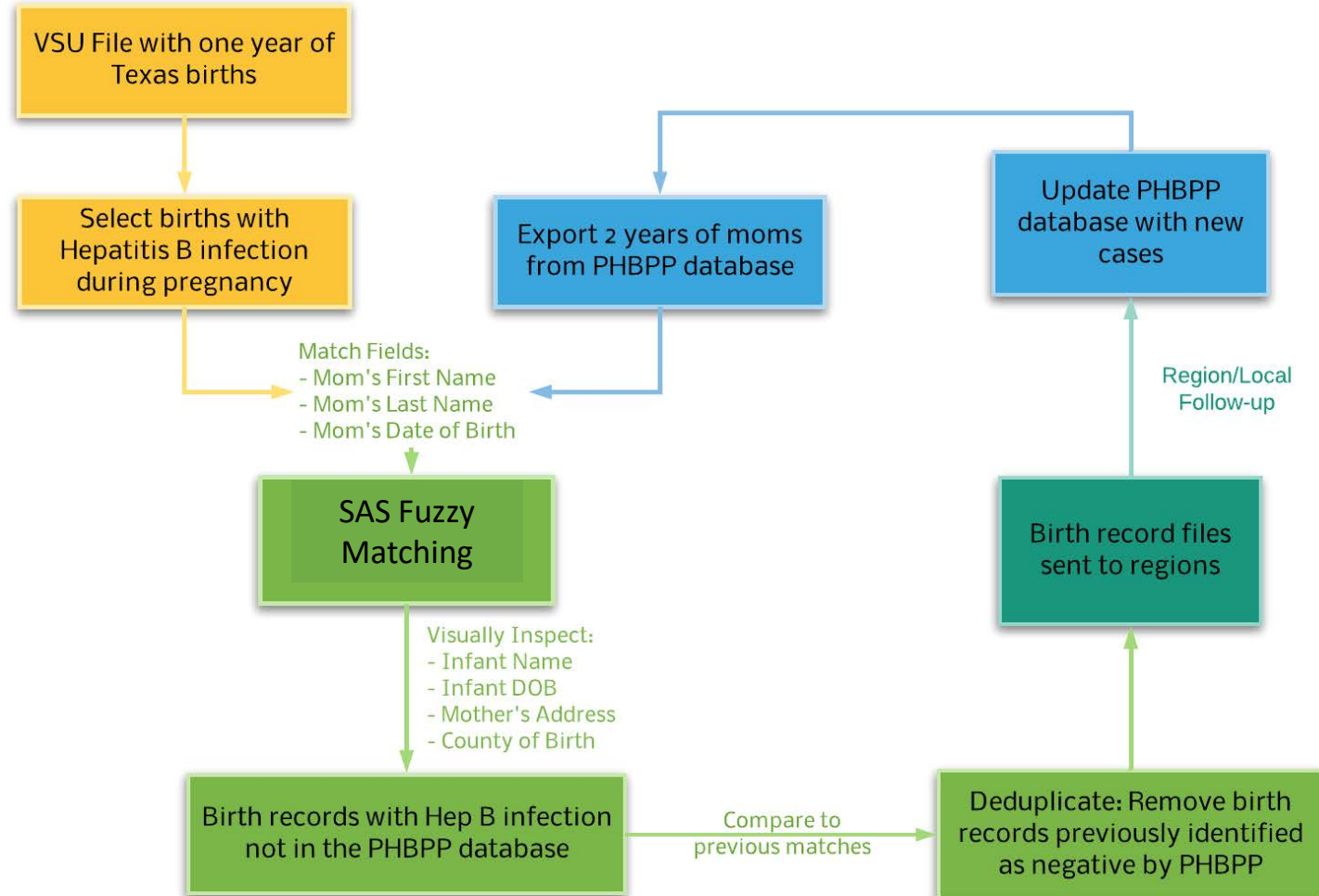
- SAS statistical analysis software
  - Data Cleaning
  - SOUNDEX
    - Invented and patented by Margaret K. Odell and Robert C. Russell in 1918 and 1922 to help match surnames that sound alike
    - Ignores case, spaces, punctuation
    - Best for English-sounding names
  - COMPGED
    - Creates a score of how similar two words are to one another

Paramont Pictures	3
Paramont Pictures	7
Universal Pictures	1
Universal Studios	2
Warner Bros	1
Warner Brothers	6

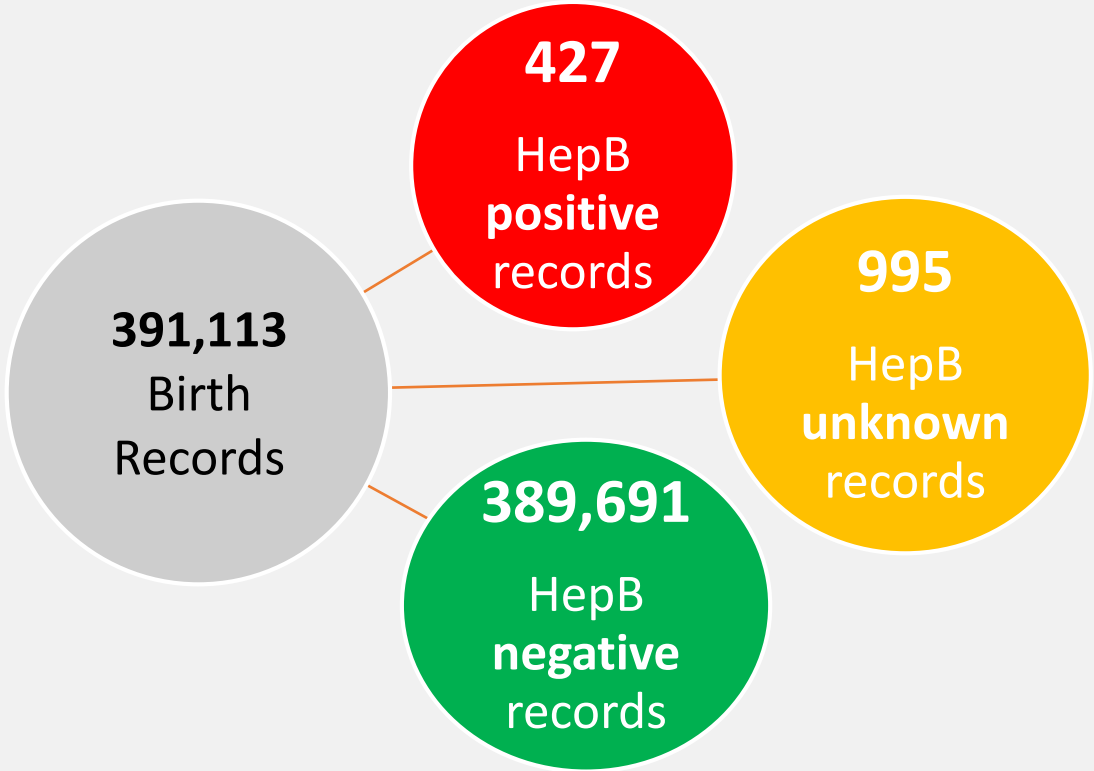
	Frequency
	2
	1
ation	1
ation	1
erica	1
Dracula	1
Dressed to Kill	1
Forrest Gump	2
Forrest Gumpp	1
Ghost	1

# VSU-PHBPP Database Match





# 2017 Birth Records



# Data Cleaning (Names)

## SAS Functions: *STRIP*, *COMPBL*, *SCAN*, *COUNT*

TABLE: Work.Test				VIEWTABLE: Work.Test2			
	infirst	inmid	inlast		infirst	inmid	inlast
1	Jessica M		Cheng	1	Jessica	M	Cheng
2	Jessica	M	Cheng	2	Jessica	M	Cheng
3	Jessica Marie		Cheng	3	Jessica	Marie	Cheng
4	Jessica	Marie	Cheng	4	Jessica	Marie	Cheng
5	Jessica M		Cheng	5	Jessica	M	Cheng
6	Jessica Marie	Marie	Cheng	6	Jessica	Marie	Cheng
7	Jessica M		Cheng	7	Jessica	M	Cheng
8	Jessica	Marie Marie	Cheng	8	Jessica	Marie	Cheng
9	Jessica Marie	Marie	Cheng	9	Jessica	Marie	Cheng
10	Jessica Marie	Marie	Cheng	10	Jessica	Marie	Cheng
11	Jessica Marie	Marie	Cheng	11	Jessica	Marie	Cheng
12	Jessica		Marie Cheng	12	Jessica	Marie	Cheng
13	Jessica	Marie	Marie Cheng	13	Jessica	Marie	Cheng
14	Jessica	Marie	Cheng Young	14	Jessica	Marie	Cheng Young

# Data Cleaning (Address)

Concatenates the new value 'ZTE' with a blank before and after.

```
&VAR2 = LEFT(PRXCHANGE('s/ #|STE | SUITE | BUILDING | BLDG / ZTE /',-1,CAT(' ',&VAR2,' ')));
```

Removes leading blanks

Needed for substitutions

Tells SAS to look to the end for matches

- Using PERL expressions
- Can add other common strings (e.g. apt, apt., apt #, unit)

VIEWTABLE: Work.Test		
	inadd	STREET1
1	100 PACIFIC AVE APT 133	100 PACIFIC ZTE 133
2	100 PACIFIC AVE APT. 133	100 PACIFIC ZTE 133
3	100 PACIFIC AVE # 133	100 PACIFIC ZTE 133
4	100 PACIFIC AVE STE 133	100 PACIFIC ZTE 133
5	100 PACIFIC AVE UNIT 133	100 PACIFIC ZTE 133
6	100 PACIFIC ST UNIT 133	100 PACIFIC ZTE 133
7	100 PACIFIC STREET UNIT 133	100 PACIFIC ZTE 133

# Matching Process

## **SOUNDEX** function

- 1) Retain the 1<sup>st</sup> letter
- 2) Discard A E H I O U W Y
- 3) Recode
  - 1: B F P V
  - 2: C G J K Q S X Z
  - 3: D T
  - 4: L
  - 5: M N
  - 6: R
- 4) If two or more adjacent lets have the same recode value, then discard all but the first

\*Better for English names

VIEWTABLE: Work.Name_test		
	Name	sname
1	Jessica	J22
2	Jessika	J22
3	Jesica	J22
4	Jess	J2
5	Jessie	J2
6	Hessica	H22
7	Sessica	S22
8	Anne	A5
9	Ann	A5
10	Annie	A5
11	Anna	A5
12	Annabelle	A514
13	Annabell	A514
14	Anny	A5

# Matching Process

## **COMPGED** function

- Lower values indicate better matches
- Similar SAS functions: **COMPLEV** and **SPEDIS**

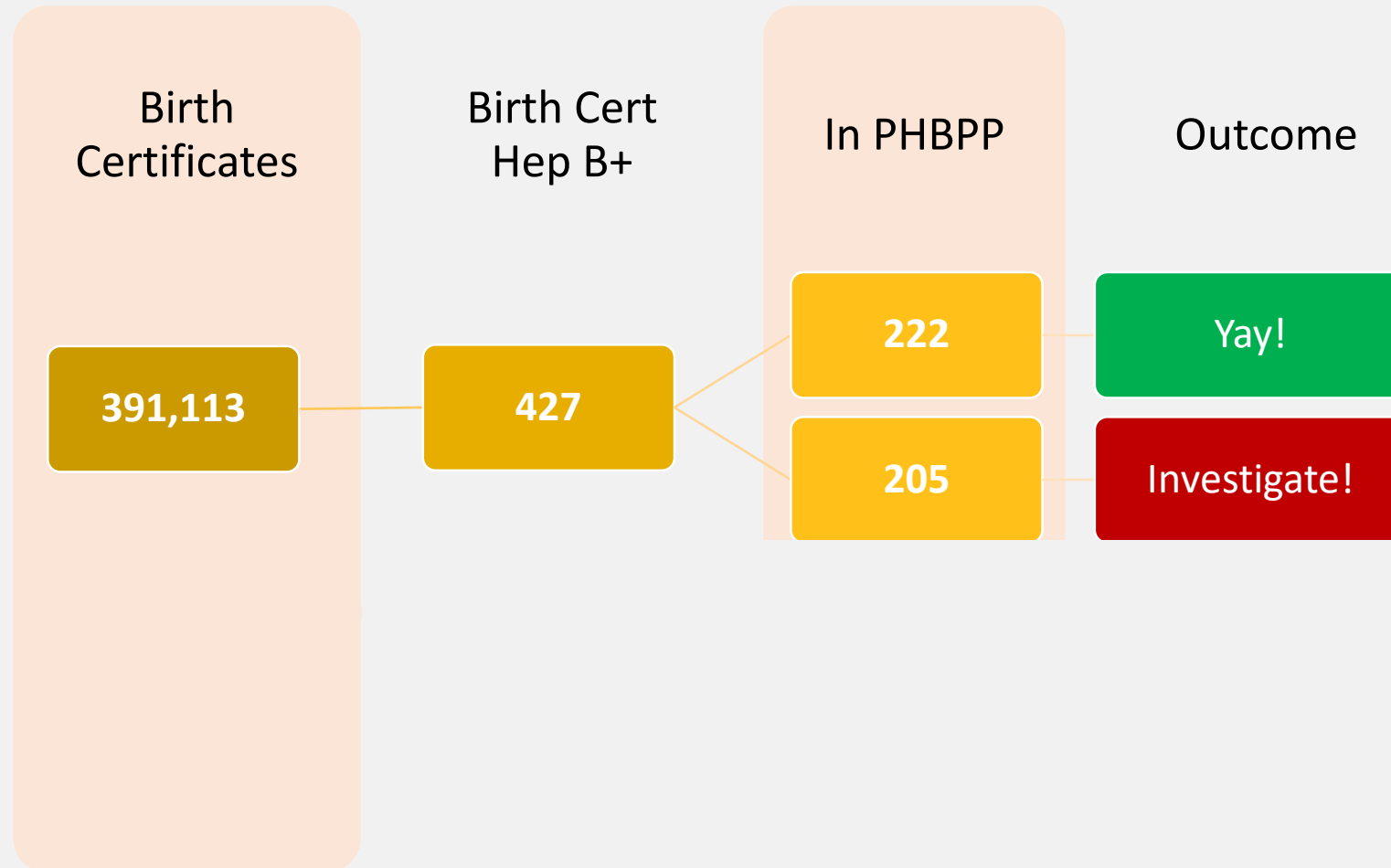
VIEWTABLE: Work.Compged_test			
	Name1	Name2	score
1	Jessica	Jessica	0
2	Jessica	Jessicaa	10
3	Jessica	Jess ica	10
4	Jessica	Jesica	20
5	Jessica	Jesscia	20
6	Jessica	Jessika	100
7	Jessica	Jess	150
8	Jessica	Jessie	150
9	Jessica	Jennifer	410
10	Jessica	Marie	650
11	Jessica	Linda	700

# Match On (iterative process)

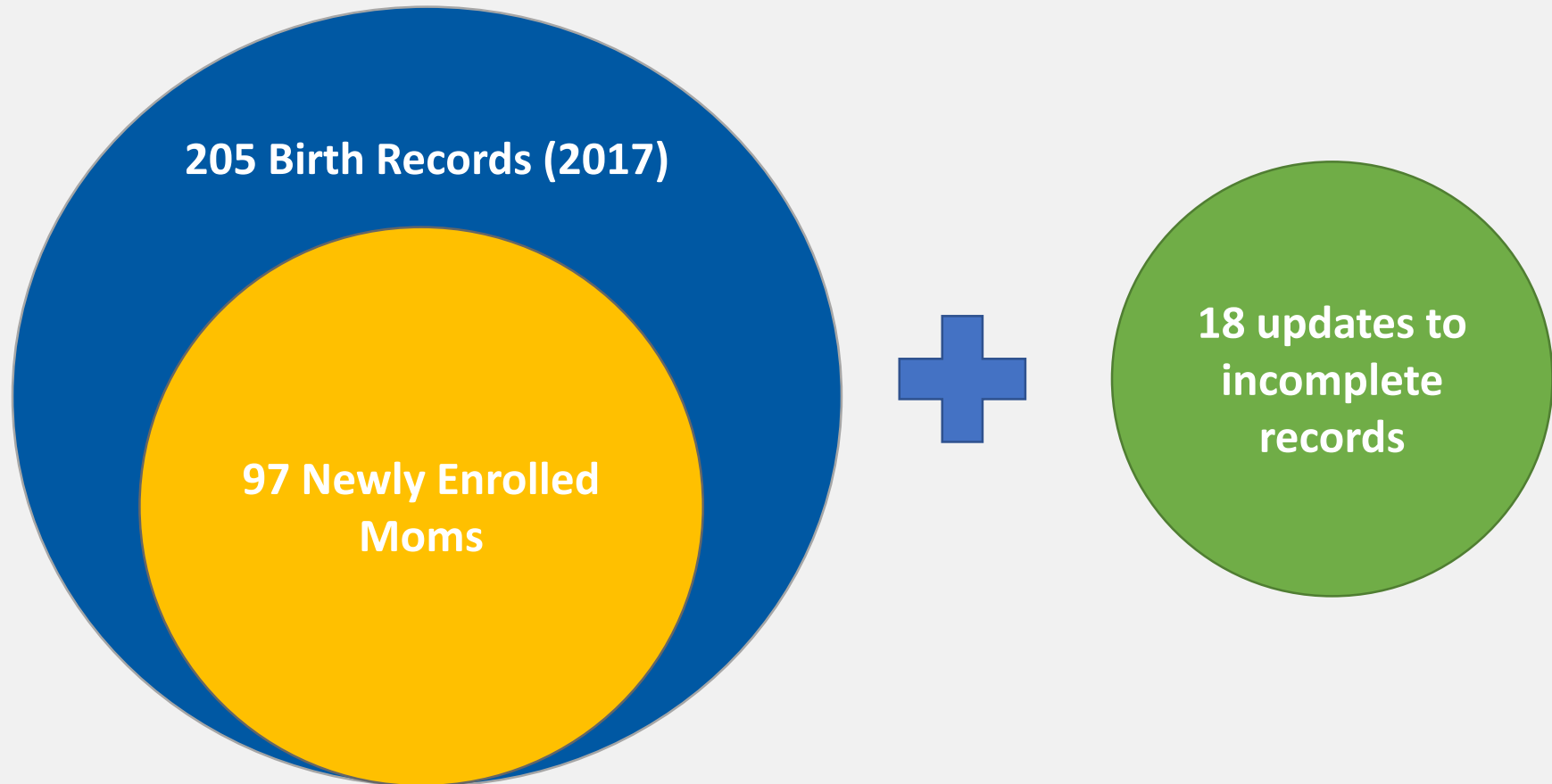
- Mom's DOB
- Mom's First to First Name (Soundex)
- Mom's Last to Last Name (Soundex)
- Mom's First to Last Name (Soundex)
- Mom's Last to First Name (Soundex)
- Etc with COMPGED Scores for each
- De-duplication



# Match Yield—2017 Cohort



# Infants Identified from Matching





# Monthly Report: Vital Statistics Unit Report (VSU)

- **Overview**

- The VSU report is a list of women who indicated that Hepatitis B was a problem during pregnancy on the birth certificate medical worksheet.

- **Purpose**

- This is another tool CO has to help identify possible HBsAg positive women. Case managers should contact the delivery hospital to verify Hepatitis B status and enroll eligible women in PHBPP.



# Monthly Report: Vital Statistics Unit Report (VSU)

In 2019

(not by birth cohort)

- Sent **256** birth records to PHBPP for follow-up
- Updated missing information on **87** infants



Texas Department of State  
Health Services

**TxEVER** Texas Electronic  
Vital Events  
Registrar

# VSU Medical Data Worksheet

VS-109.2 (09/11)

### Medical Data Worksheet for Child's Birth Certificate

This form to be completed by hospital staff. This data will be used to populate the medical data portion of the birth certificate for the newborn. The medical data is required to be reported within five days of the birth. [HSC §192.003]

PATIENT REFERENCE:	
MOTHER MR# _____	NEWBORN MR# _____
MOTHER'S NAME _____	NEWBORN NAME _____
MEDICAID# _____	DOB _____
DELIVERING DR _____	DATE AOP SENT _____
MOTHER TRANSFERRED _____	SOURCE OF PAYMENT FOR DELIVERY _____
<input type="checkbox"/> Born at Facility <input type="checkbox"/> Born En Route <input type="checkbox"/> Foundling <input type="checkbox"/> Home Birth	

<b>Prenatal Care</b> <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown Date of First Visit: ____/____/____ Date of Last Visit: ____/____/____ Total Number of Prenatal Visits for this Pregnancy: _____ Date Last Normal Menses Began: ____/____/____	<b>Source of Prenatal Care</b> (check all that apply) <input type="checkbox"/> None <input type="checkbox"/> Midwife <input type="checkbox"/> Hospital Clinic <input type="checkbox"/> Other, Specify _____ <input type="checkbox"/> Public Health Clinic <input type="checkbox"/> Unknown <input type="checkbox"/> Private Physician
<b>Pregnancy History</b> <b>Live births now living</b> (Do not include this birth. For multiple deliveries, do not include the 1 <sup>st</sup> born in the set if completing this worksheet for that child. If none enter '0'): _____ <b>Live births now dead</b> (Do not include this birth. For multiple deliveries, do not include the 1 <sup>st</sup> born in the set if completing this worksheet for that child. If none enter '0'): _____ Date of last live birth: ____/____/____ MM YYYY <b>Number of other pregnancy outcomes</b> (include fetal losses of any gestational age. If this was a multiple delivery, include all fetal losses delivered before this infant in the pregnancy. If none enter '0'): _____ Date of last other pregnancy outcome: ____/____/____ MM YYYY	<b>Risk Factors in this Pregnancy</b> (check all that apply) <b>Diabetes</b> <input type="checkbox"/> Prepregnancy (diagnosis prior to this pregnancy) <input type="checkbox"/> Gestational (diagnosis in this pregnancy) <b>Hypertension</b> <input type="checkbox"/> Prepregnancy (chronic) <input type="checkbox"/> Gestational (PIH, preeclampsia) <input type="checkbox"/> Eclampsia <input type="checkbox"/> Previous preterm birth <input type="checkbox"/> Other previous poor pregnancy outcome (includes perinatal death, small-for-gestational age/intrauterine growth restricted birth) <input type="checkbox"/> Pregnancy resulted from infertility treatment <input type="checkbox"/> Fertility-enhancing drugs, artificial insemination or intrauterine insemination <input type="checkbox"/> Assisted reproductive technology <input type="checkbox"/> Mother had a previous cesarean delivery If yes, how many? _____ <input type="checkbox"/> Antiretrovirals administered during pregnancy or at delivery <input type="checkbox"/> None of the above

<b>Infections Present and/or Treated During Pregnancy</b> (check all that apply) <input type="checkbox"/> Gonorrhea <input type="checkbox"/> Hepatitis B <input type="checkbox"/> Syphilis <input type="checkbox"/> Hepatitis C <input type="checkbox"/> Chlamydia <input type="checkbox"/> None of the above	<b>Infections Present and/or Treated During Pregnancy</b> (check all that apply) <input type="checkbox"/> Gonorrhea <input checked="" type="checkbox"/> Hepatitis B <input type="checkbox"/> Syphilis <input type="checkbox"/> Hepatitis C <input type="checkbox"/> Chlamydia <input type="checkbox"/> None of the above
--	---

<b>HIV Test</b> HIV test done Prenatally <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown (check all that apply) <input type="checkbox"/> First Trimester <input type="checkbox"/> Third Trimester	1 of 2 HIV test done at delivery <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown Infant tested for HIV at birth <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Unknown
--	---

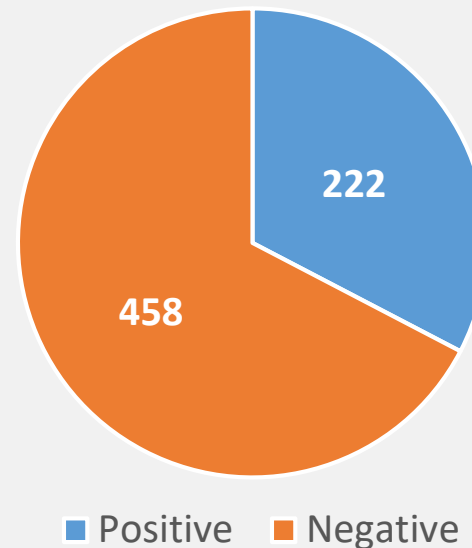
## Infections Present and/or Treated During Pregnancy (check all that apply)

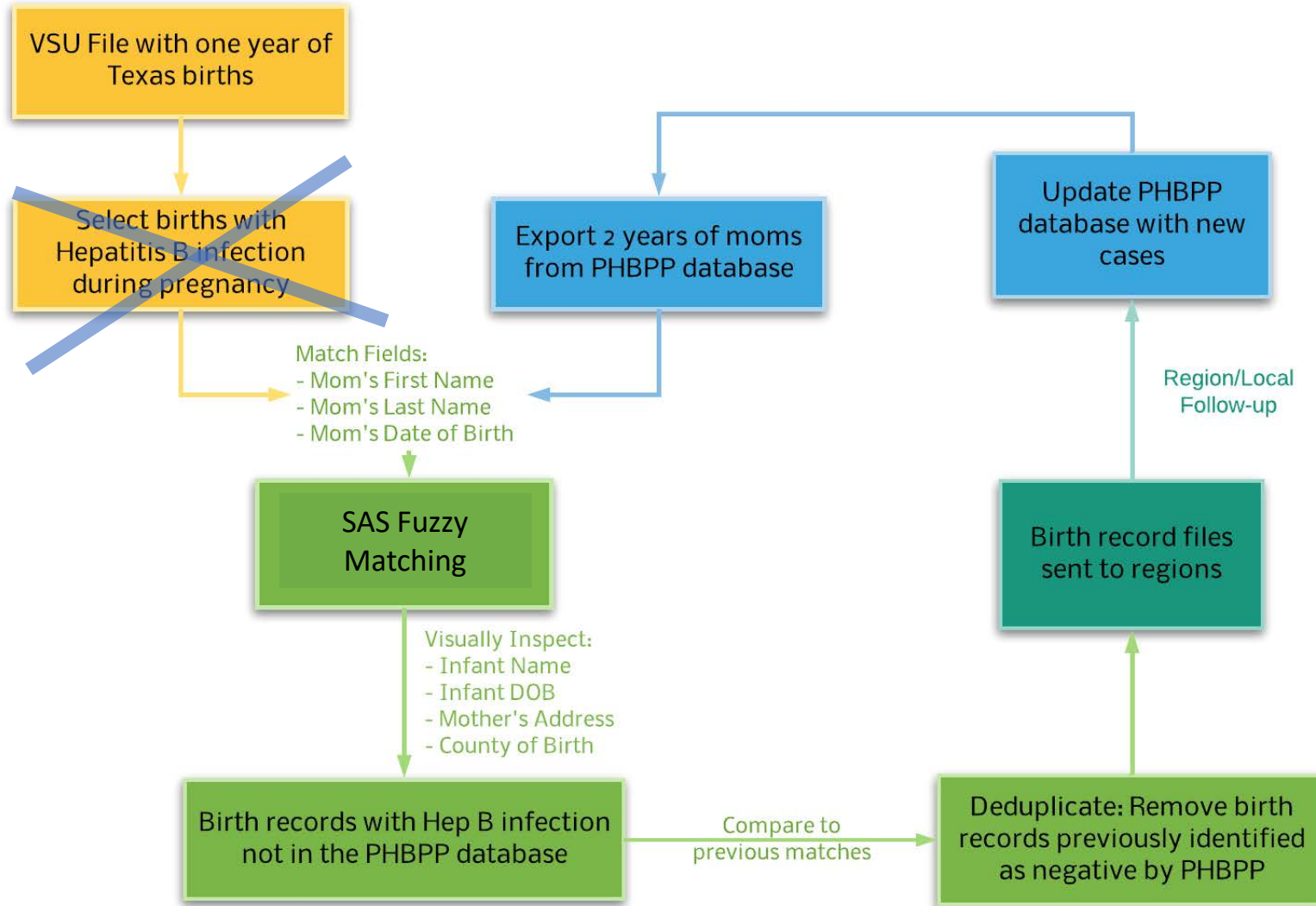
- Gonorrhea
- Hepatitis B
- Syphilis
- Hepatitis C
- Chlamydia
- None of the above

# Reverse Match to All Birth Records

Of all the infants born in 2017 who enrolled in PHBPP, only about **1 in 3** infants had indicated on their birth certificate that mom was HepB+

HepB Indication on the Birth Record for Infants Enrolled in the PHBPP





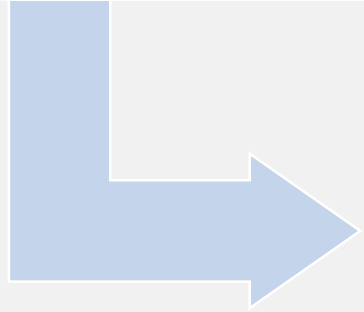
# NEDSS+VSU+PHBPP Database

- Lab Results
  - Hep B surface antigen test
  - In a specific date range (weekly for the week before)
  - Not male (female or unknown, very small unknown number)
  - Of childbearing age
- Match against all VSU Birth Records
- Match against the PHBPP Database
  - if mom not there, route for follow up



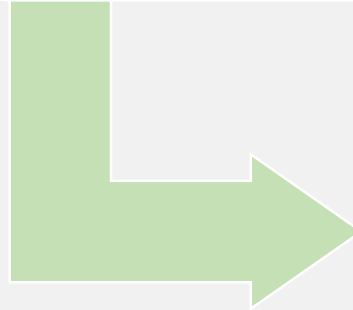
NEDSS

- Lab results



VSU

- Birth Records (400K)



PHBPP

- Database



# Fall 2019 NEDSS-VSU-PHBPP Match Results

2018 and partial 2019 births

---





# Fall 2019 NEDSS-VSU-PHBPP Match Results

2018 and partial 2019 births

- Eligible Women Found on prior Monthly Report
  - VSU report n= 5
  - Lab Report n=4
- **31% could have been enrolled earlier** using monthly report
- None found on both lab & VSU report
- **69% of eligible cases were not found using standard VSU & NEDSS reports**



# ImmTrac and HBIG

ImmTrac Hbig Records  
N=120  
(Infants born in 2017 or 2018 and who received  
Hbig within 7 days of birth)

In PHBPP  
N=48

Not in PHBPP  
N=72

Cases  
N=47

Not Cases (ineligible)  
N=1

Entered In Peri First  
N=1

Entered in ImmTrac First  
N=9

Entered In Peri First  
N=38

As of 1/18/19


# The Future



## New PHBPP Database

- Import data from NEDSS
  - Include more lab results

## Matching

- New DUA with Vital Stats
  - Streamline SAS code using PROC SQL and compare results to existing match process
  - Explore other options such as CDC Link Plus
  - Continue to track progress
- 

# Thank you!

This presentation made possible by:

Rachel Wiseman

Eva Hansson

Khalilah Loggins

Jessica Jupe

Srishty Arora

Lakshmi Sahini

Jess Cheng

Lauren Hoffman-Arriaga

Holly Hughes Garza – Epidemiology Team Lead

[holly.garza@dshs.texas.gov](mailto:holly.garza@dshs.texas.gov)