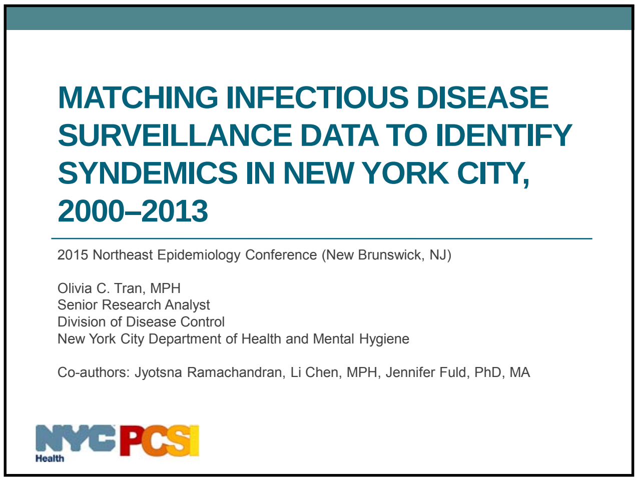
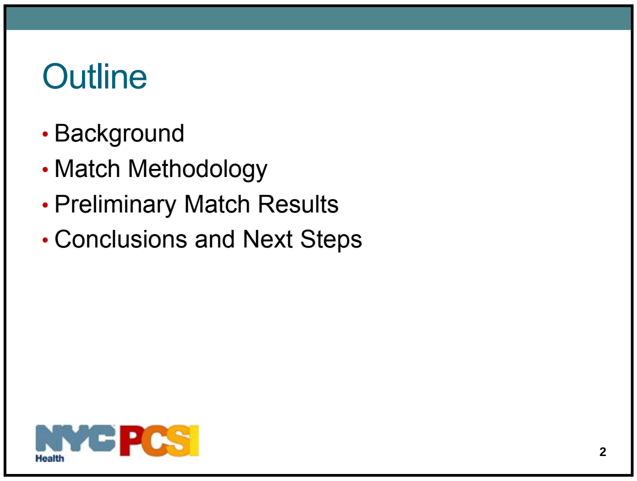


MATCHING INFECTIOUS DISEASE SURVEILLANCE DATA TO IDENTIFY SYNDEMICS IN NEW YORK CITY, 2000–2013

2015 Northeast Epidemiology Conference (New Brunswick, NJ)

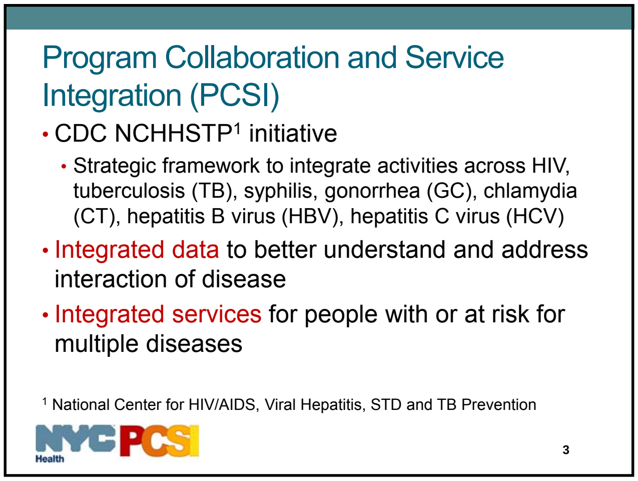
Olivia C. Tran, MPH
Senior Research Analyst
Division of Disease Control
New York City Department of Health and Mental Hygiene

Co-authors: Jyotsna Ramachandran, Li Chen, MPH, Jennifer Fuld, PhD, MA



Outline

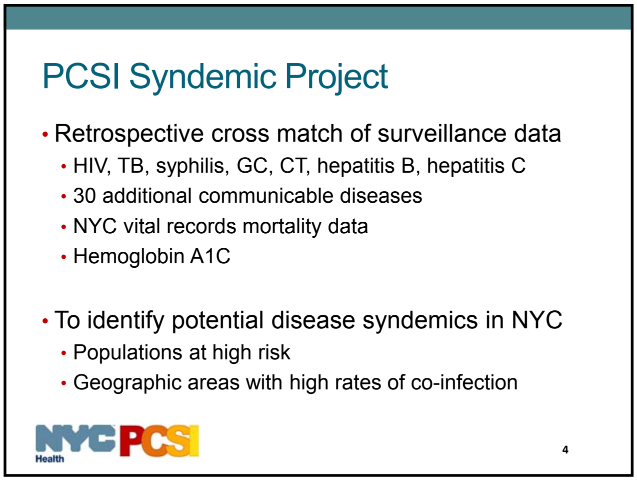
- Background
- Match Methodology
- Preliminary Match Results
- Conclusions and Next Steps



Program Collaboration and Service Integration (PCSI)

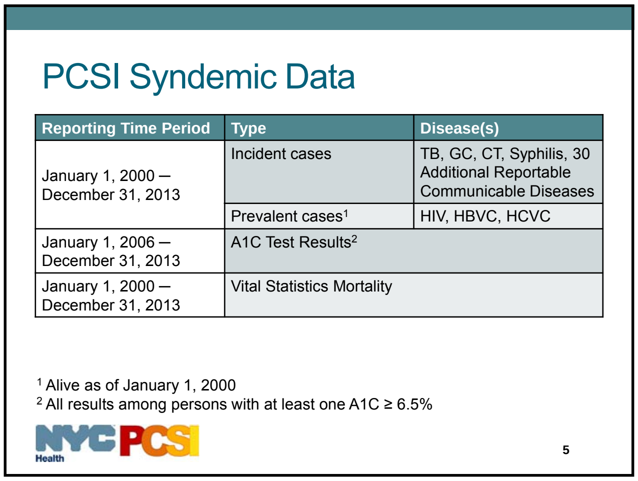
- CDC NCHHSTP[1] initiative
  - Strategic framework to integrate activities across HIV, tuberculosis (TB), syphilis, gonorrhea (GC), chlamydia (CT), hepatitis B virus (HBV), hepatitis C virus (HCV)
- Integrated data to better understand and address interaction of disease
- Integrated services for people with or at risk for multiple diseases

[1] National Center for HIV/AIDS, Viral Hepatitis, STD and TB Prevention



PCSI Syndemic Project

- Retrospective cross match of surveillance data
  - HIV, TB, syphilis, GC, CT, hepatitis B, hepatitis C
  - 30 additional communicable diseases
  - NYC vital records mortality data
  - Hemoglobin A1C
- To identify potential disease syndemics in NYC
  - Populations at high risk
  - Geographic areas with high rates of co-infection

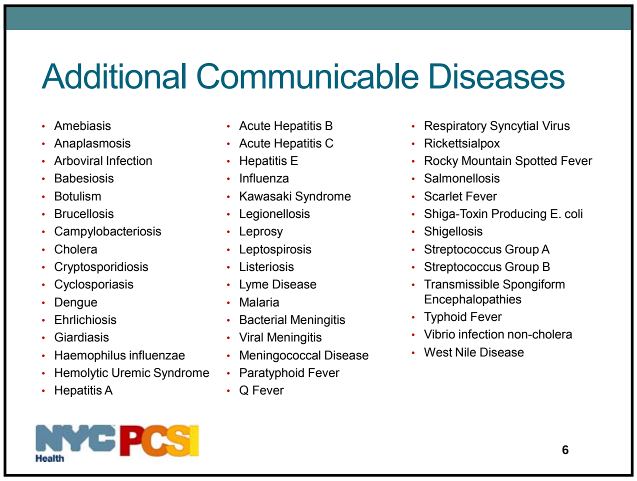


PCSI Syndemic Data

Reporting Time Period	Type	Disease(s)
January 1, 2000 – December 31, 2013	Incident cases	TB, GC, CT, Syphilis, 30 Additional Reportable Communicable Diseases
	Prevalent cases[1]	HIV, HBVC, HCVC
January 1, 2006 – December 31, 2013	A1C Test Results[2]	
January 1, 2000 – December 31, 2013	Vital Statistics Mortality	

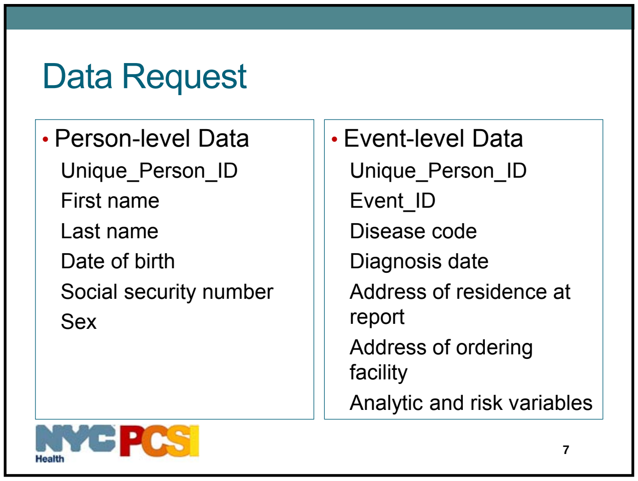
[1] Alive as of January 1, 2000

[2] All results among persons with at least one A1C ≥ 6.5%



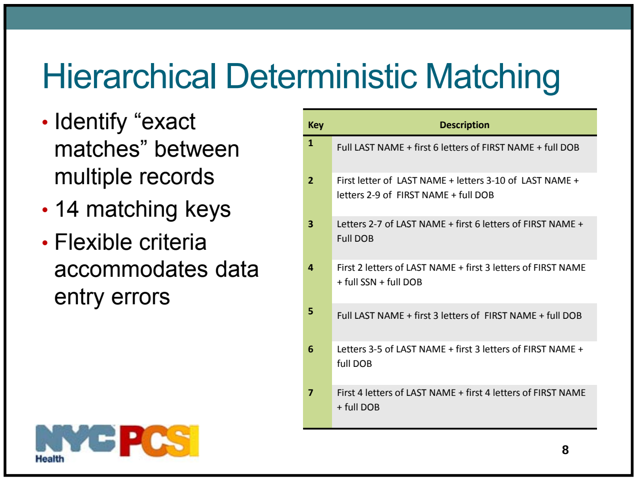
Additional Communicable Diseases

- Amebiasis
- Anaplasmosis
- Arboviral Infection
- Babesiosis
- Botulism
- Brucellosis
- Campylobacteriosis
- Cholera
- Cryptosporidiosis
- Cyclosporiasis
- Dengue
- Ehrlichiosis
- Giardiasis
- Haemophilus influenzae
- Hemolytic Uremic Syndrome
- Hepatitis A
- Acute Hepatitis B
- Acute Hepatitis C
- Hepatitis E
- Influenza
- Kawasaki Syndrome
- Legionellosis
- Leprosy
- Leptospirosis
- Listeriosis
- Lyme Disease
- Malaria
- Bacterial Meningitis
- Viral Meningitis
- Meningococcal Disease
- Paratyphoid Fever
- Q Fever
- Respiratory Syncytial Virus
- Rickettsialpox
- Rocky Mountain Spotted Fever
- Salmonellosis
- Scarlet Fever
- Shiga-Toxin Producing E. coli
- Shigellosis
- Streptococcus Group A
- Streptococcus Group B
- Transmissible Spongiform Encephalopathies
- Typhoid Fever
- Vibrio infection non-cholera
- West Nile Disease



Data Request

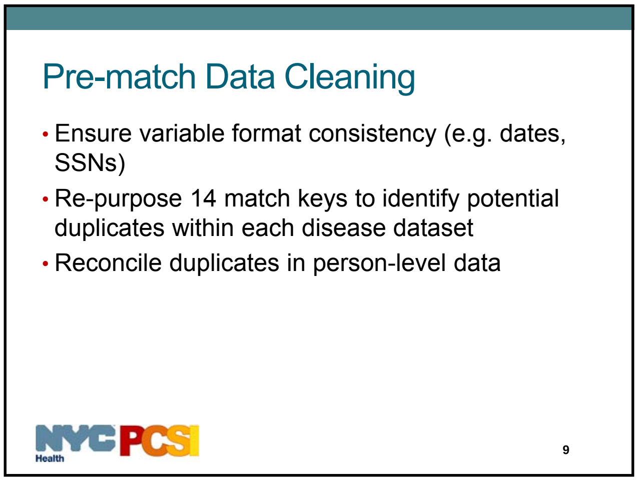
- Person-level Data
  - Unique_Person_ID
  - First name
  - Last name
  - Date of birth
  - Social security number
  - Sex
- Event-level Data
  - Unique_Person_ID
  - Event_ID
  - Disease code
  - Diagnosis date
  - Address of residence at report
  - Address of ordering facility
  - Analytic and risk variables



Hierarchical Deterministic Matching

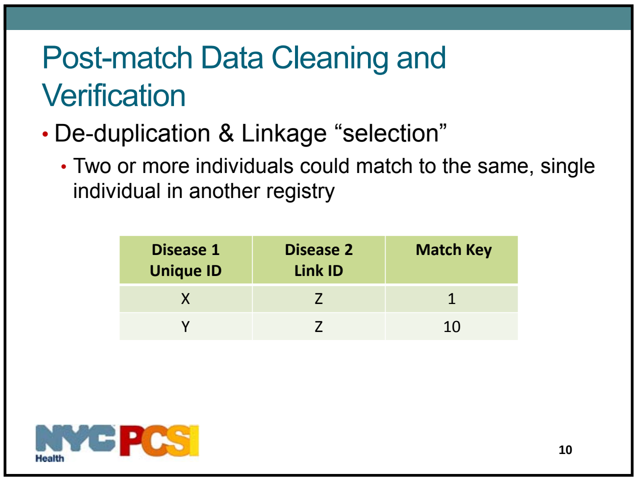
- Identify "exact matches" between multiple records
- 14 matching keys
- Flexible criteria accommodates data entry errors

Key	Description
1	Full LAST NAME + first 6 letters of FIRST NAME + full DOB
2	First letter of LAST NAME + letters 3-10 of LAST NAME + letters 2-9 of FIRST NAME + full DOB
3	Letters 2-7 of LAST NAME + first 6 letters of FIRST NAME + Full DOB
4	First 2 letters of LAST NAME + first 3 letters of FIRST NAME + full SSN + full DOB
5	Full LAST NAME + first 3 letters of FIRST NAME + full DOB
6	Letters 3-5 of LAST NAME + first 3 letters of FIRST NAME + full DOB
7	First 4 letters of LAST NAME + first 4 letters of FIRST NAME + full DOB



Pre-match Data Cleaning

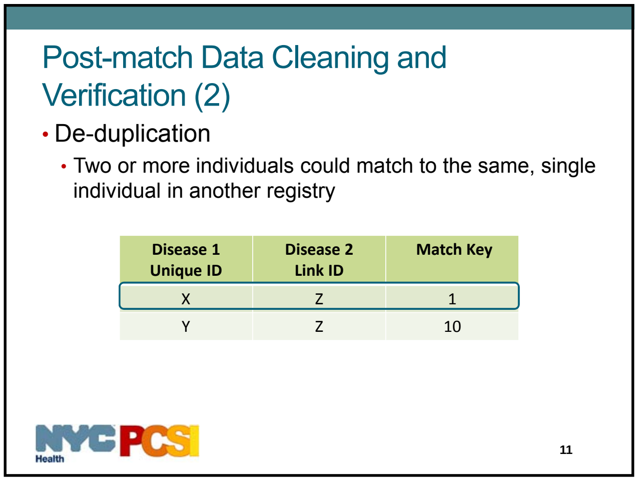
- Ensure variable format consistency (e.g. dates, SSNs)
- Re-purpose 14 match keys to identify potential duplicates within each disease dataset
- Reconcile duplicates in person-level data



Post-match Data Cleaning and Verification

- De-duplication & Linkage “selection”
  - Two or more individuals could match to the same, single individual in another registry

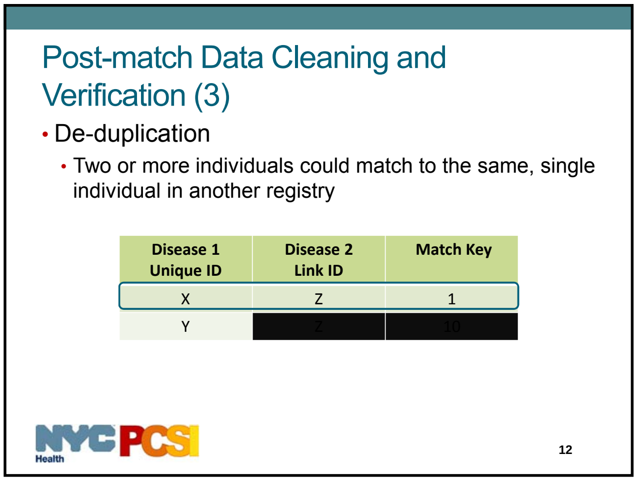
Disease 1 Unique ID	Disease 2 Link ID	Match Key
X	Z	1
Y	Z	10



Post-match Data Cleaning and Verification (2)

- De-duplication
  - Two or more individuals could match to the same, single individual in another registry

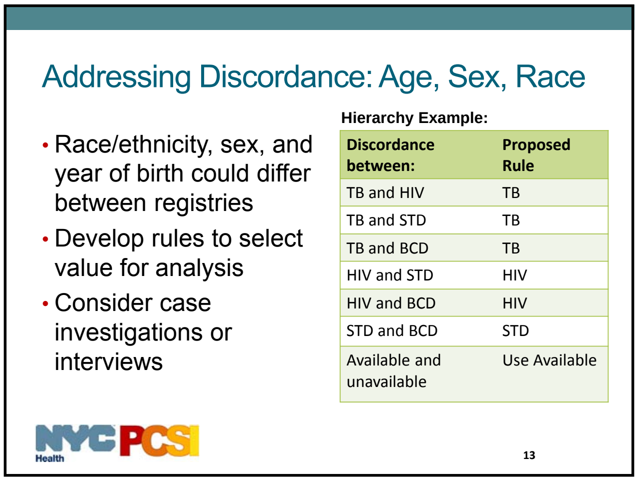
Disease 1 Unique ID	Disease 2 Link ID	Match Key
X	Z	1
Y	Z	10



Post-match Data Cleaning and Verification (3)

- De-duplication
  - Two or more individuals could match to the same, single individual in another registry

Disease 1 Unique ID	Disease 2 Link ID	Match Key
X	Z	1
Y	[redacted]	[redacted]

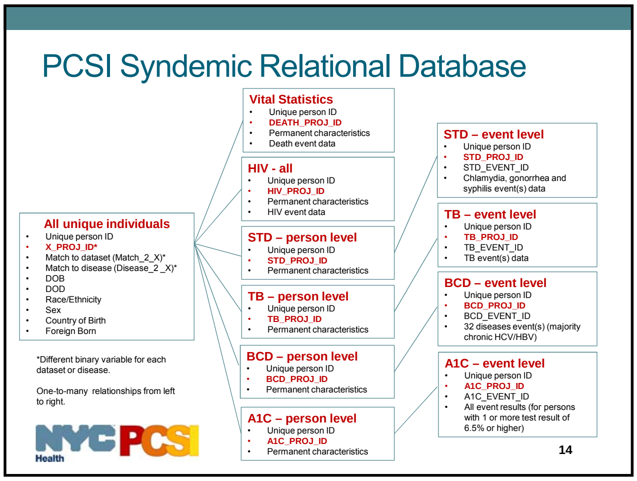


Addressing Discordance: Age, Sex, Race

- Race/ethnicity, sex, and year of birth could differ between registries
- Develop rules to select value for analysis
- Consider case investigations or interviews

Hierarchy Example:

Discordance between:	Proposed Rule
TB and HIV	TB
TB and STD	TB
TB and BCD	TB
HIV and STD	HIV
HIV and BCD	HIV
STD and BCD	STD
Available and unavailable	Use Available



PCSI Syndemic Relational Database

All unique individuals
- Unique person ID
- X_PROJ_ID*
- Match to dataset (Match_2_X)*
- Match to disease (Disease_2 _X)*
- DOB
- DOD
- Race/Ethnicity
- Sex
- Country of Birth
- Foreign Born

*Different binary variable for each dataset or disease.

One-to-many relationships from left to right.

Vital Statistics
- Unique person ID
- DEATH_PROJ_ID
- Permanent characteristics
- Death event data

HIV - all
- Unique person ID
- HIV_PROJ_ID
- Permanent characteristics
- HIV event data

STD – person level
- Unique person ID
- STD_PROJ_ID
- Permanent characteristics

TB – person level
- Unique person ID
- TB_PROJ_ID
- Permanent characteristics

BCD – person level
- Unique person ID
- BCD_PROJ_ID
- Permanent characteristics

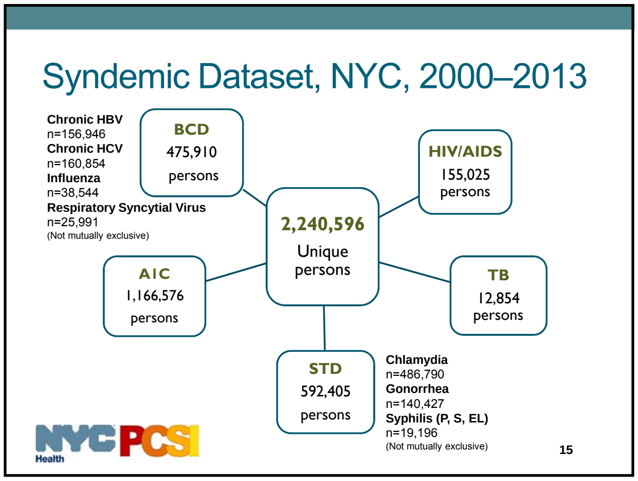
A1C – person level
- Unique person ID
- A1C_PROJ_ID
- Permanent characteristics

STD – event level
- Unique person ID
- STD_PROJ_ID
- STD_EVENT_ID
- Chlamydia, gonorrhea and syphilis event(s) data

TB – event level
- Unique person ID
- TB_PROJ_ID
- TB_EVENT_ID
- TB event(s) data

BCD – event level
- Unique person ID
- BCD_PROJ_ID
- BCD_EVENT_ID
- 32 diseases event(s) (majority chronic HCV/HBV)

A1C – event level
- Unique person ID
- A1C_PROJ_ID
- A1C_EVENT_ID
- All event results (for persons with 1 or more test result of 6.5% or higher)



Syndemic Dataset, NYC, 2000–2013

Chronic HBV
n=156,946
Chronic HCV
n=160,854
Influenza
n=38,544
Respiratory Syncytial Virus
n=25,991
(Not mutually exclusive)

BCD
475,910
persons

HIV/AIDS
155,025
persons

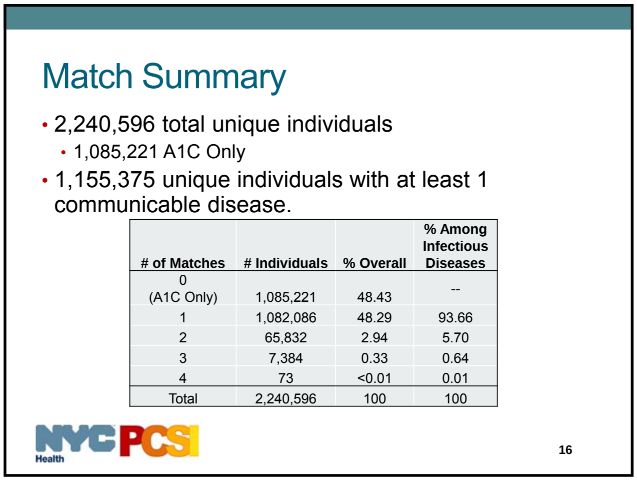
2,240,596
Unique
persons

AIC
1,166,576
persons

TB
12,854
persons

STD
592,405
persons

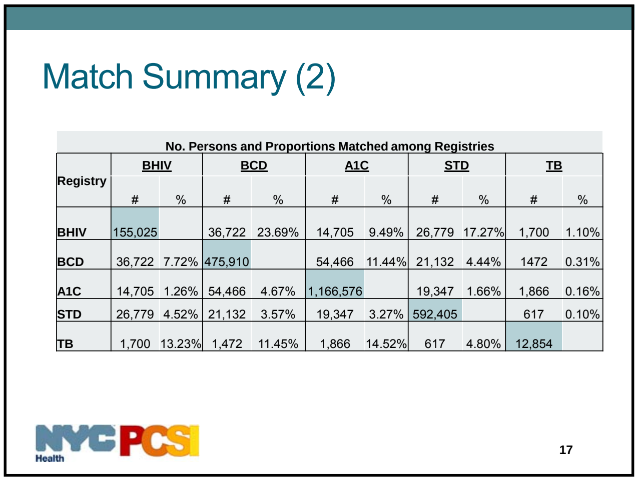
Chlamydia
n=486,790
Gonorrhea
n=140,427
Syphilis (P, S, EL)
n=19,196
(Not mutually exclusive)



Match Summary

- 2,240,596 total unique individuals
  - 1,085,221 A1C Only
- 1,155,375 unique individuals with at least 1 communicable disease.

# of Matches	# Individuals	% Overall	% Among Infectious Diseases
0 (A1C Only)	1,085,221	48.43	--
1	1,082,086	48.29	93.66
2	65,832	2.94	5.70
3	7,384	0.33	0.64
4	73	<0.01	0.01
Total	2,240,596	100	100



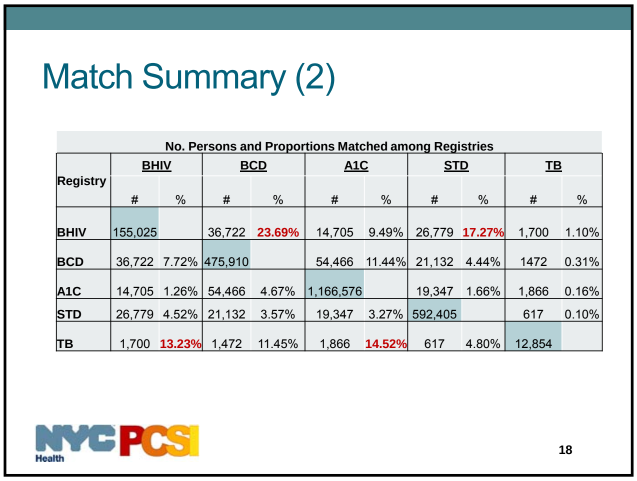
Match Summary (2)

No. Persons and Proportions Matched among Registries

Registry	BHIV		BCD		A1C		STD		TB	
	#	%	#	%	#	%	#	%	#	%
BHIV	155,025		36,722	23.69%	14,705	9.49%	26,779	17.27%	1,700	1.10%
BCD	36,722	7.72%	475,910		54,466	11.44%	21,132	4.44%	1472	0.31%
A1C	14,705	1.26%	54,466	4.67%	1,166,576		19,347	1.66%	1,866	0.16%
STD	26,779	4.52%	21,132	3.57%	19,347	3.27%	592,405		617	0.10%
TB	1,700	13.23%	1,472	11.45%	1,866	14.52%	617	4.80%	12,854	

NYC Health PCSI

17



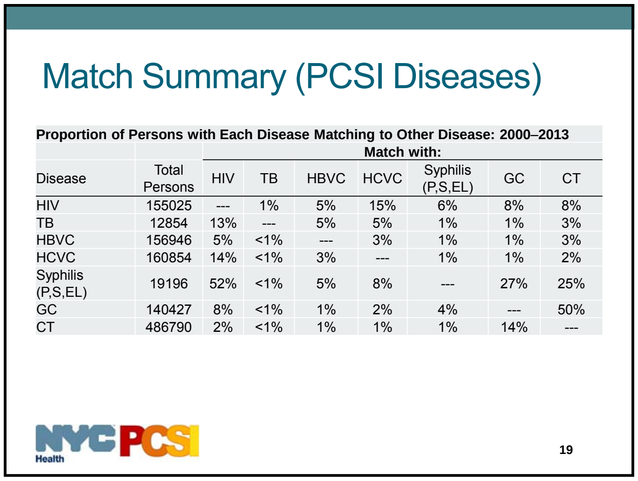
Match Summary (2)

No. Persons and Proportions Matched among Registries

Registry	BHIV		BCD		A1C		STD		TB	
	#	%	#	%	#	%	#	%	#	%
BHIV	155,025		36,722	**23.69%**	14,705	9.49%	26,779	**17.27%**	1,700	1.10%
BCD	36,722	7.72%	475,910		54,466	11.44%	21,132	4.44%	1472	0.31%
A1C	14,705	1.26%	54,466	4.67%	1,166,576		19,347	1.66%	1,866	0.16%
STD	26,779	4.52%	21,132	3.57%	19,347	3.27%	592,405		617	0.10%
TB	1,700	**13.23%**	1,472	11.45%	1,866	**14.52%**	617	4.80%	12,854	

NYC Health PCSI

18



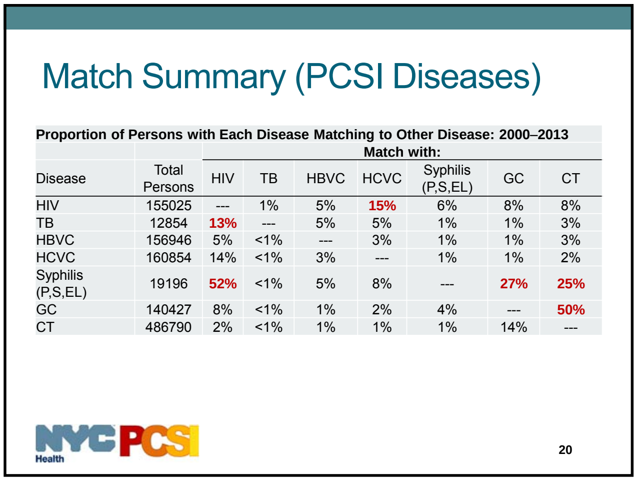
Match Summary (PCSI Diseases)

Proportion of Persons with Each Disease Matching to Other Disease: 2000–2013

		Match with:						
Disease	Total Persons	HIV	TB	HBVC	HCVC	Syphilis (P,S,EL)	GC	CT
HIV	155025	---	1%	5%	15%	6%	8%	8%
TB	12854	13%	---	5%	5%	1%	1%	3%
HBVC	156946	5%	<1%	---	3%	1%	1%	3%
HCVC	160854	14%	<1%	3%	---	1%	1%	2%
Syphilis (P,S,EL)	19196	52%	<1%	5%	8%	---	27%	25%
GC	140427	8%	<1%	1%	2%	4%	---	50%
CT	486790	2%	<1%	1%	1%	1%	14%	---

NYC Health PCSI

19



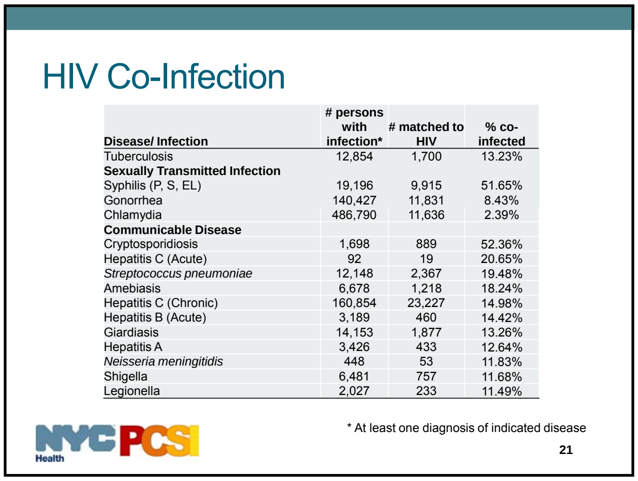
Match Summary (PCSI Diseases)

Proportion of Persons with Each Disease Matching to Other Disease: 2000–2013

		Match with:						
Disease	Total Persons	HIV	TB	HBVC	HCVC	Syphilis (P,S,EL)	GC	CT
HIV	155025	---	1%	5%	**15%**	6%	8%	8%
TB	12854	**13%**	---	5%	5%	1%	1%	3%
HBVC	156946	5%	<1%	---	3%	1%	1%	3%
HCVC	160854	14%	<1%	3%	---	1%	1%	2%
Syphilis (P,S,EL)	19196	**52%**	<1%	5%	8%	---	**27%**	**25%**
GC	140427	8%	<1%	1%	2%	4%	---	**50%**
CT	486790	2%	<1%	1%	1%	1%	14%	---

NYC Health PCSI

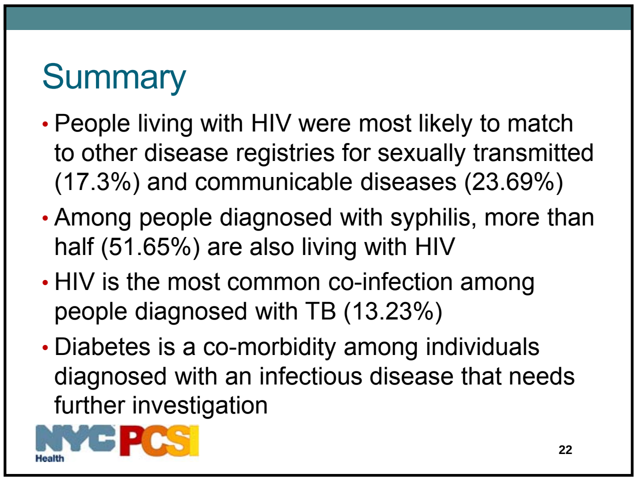
20



HIV Co-Infection

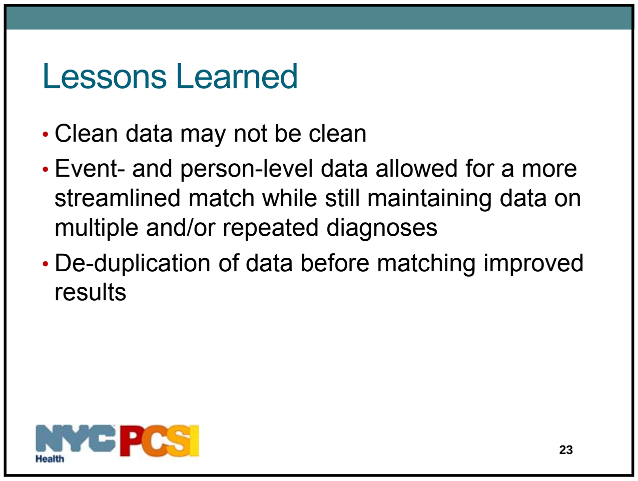
Disease/ Infection	# persons with infection*	# matched to HIV	% co-infected
Tuberculosis	12,854	1,700	13.23%
Sexually Transmitted Infection			
Syphilis (P, S, EL)	19,196	9,915	51.65%
Gonorrhea	140,427	11,831	8.43%
Chlamydia	486,790	11,636	2.39%
Communicable Disease			
Cryptosporidiosis	1,698	889	52.36%
Hepatitis C (Acute)	92	19	20.65%
Streptococcus pneumoniae	12,148	2,367	19.48%
Amebiasis	6,678	1,218	18.24%
Hepatitis C (Chronic)	160,854	23,227	14.98%
Hepatitis B (Acute)	3,189	460	14.42%
Giardiasis	14,153	1,877	13.26%
Hepatitis A	3,426	433	12.64%
Neisseria meningitidis	448	53	11.83%
Shigella	6,481	757	11.68%
Legionella	2,027	233	11.49%

* At least one diagnosis of indicated disease



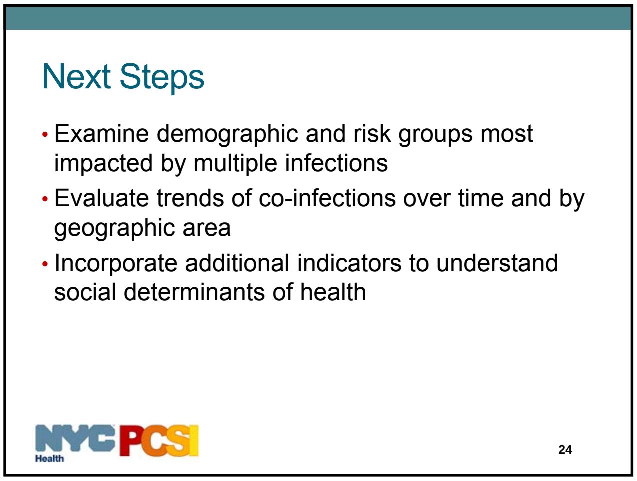
Summary

- People living with HIV were most likely to match to other disease registries for sexually transmitted (17.3%) and communicable diseases (23.69%)
- Among people diagnosed with syphilis, more than half (51.65%) are also living with HIV
- HIV is the most common co-infection among people diagnosed with TB (13.23%)
- Diabetes is a co-morbidity among individuals diagnosed with an infectious disease that needs further investigation



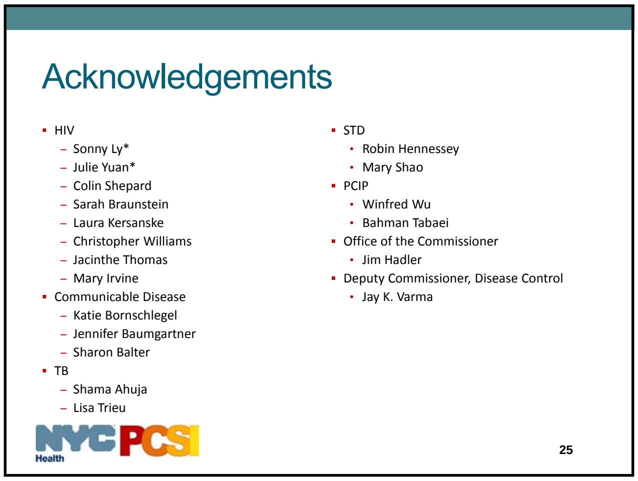
Lessons Learned

- Clean data may not be clean
- Event- and person-level data allowed for a more streamlined match while still maintaining data on multiple and/or repeated diagnoses
- De-duplication of data before matching improved results



Next Steps

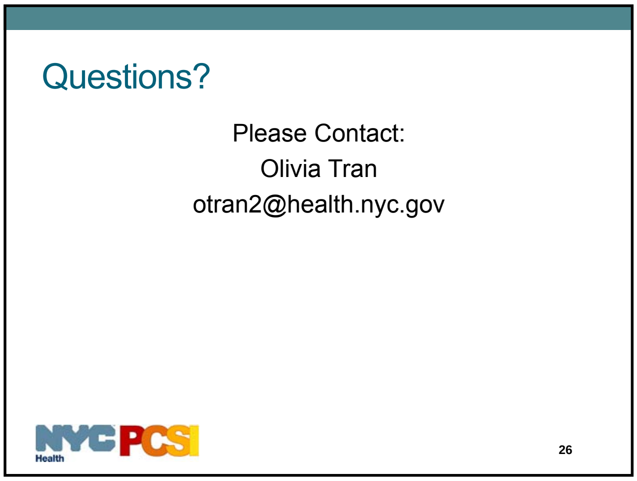
- Examine demographic and risk groups most impacted by multiple infections
- Evaluate trends of co-infections over time and by geographic area
- Incorporate additional indicators to understand social determinants of health



Acknowledgements

- HIV
  - Sonny Ly*
  - Julie Yuan*
  - Colin Shepard
  - Sarah Braunstein
  - Laura Kersanske
  - Christopher Williams
  - Jacinthe Thomas
  - Mary Irvine
- Communicable Disease
  - Katie Bornschlegel
  - Jennifer Baumgartner
  - Sharon Balter
- TB
  - Shama Ahuja
  - Lisa Trieu
- STD
  - Robin Hennessey
  - Mary Shao
- PCIP
  - Winfred Wu
  - Bahman Tabaei
- Office of the Commissioner
  - Jim Hadler
- Deputy Commissioner, Disease Control
  - Jay K. Varma

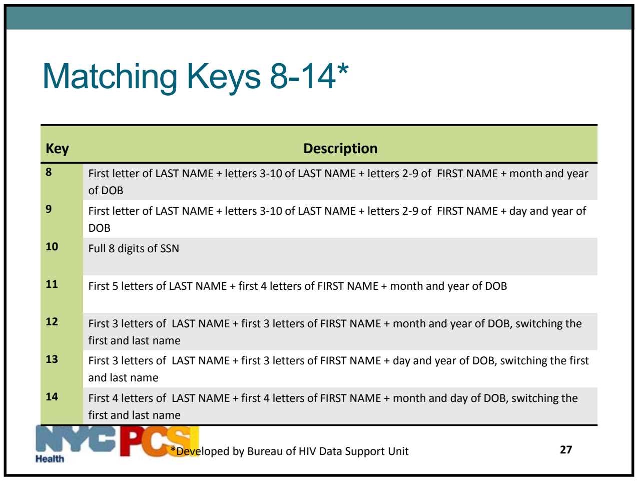
25



Questions?

Please Contact:
Olivia Tran
otran2@health.nyc.gov

26

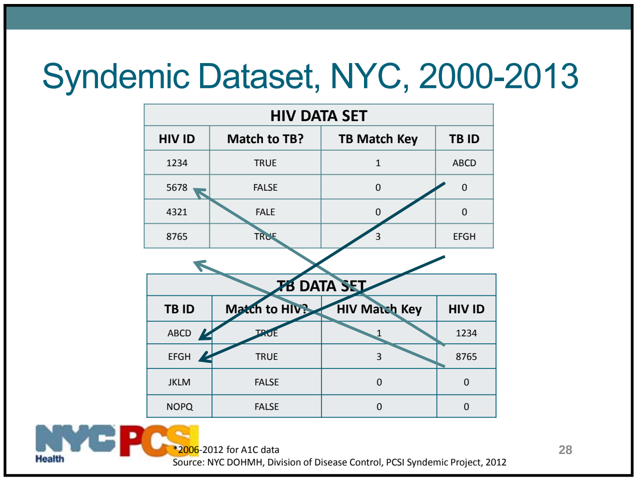


Matching Keys 8-14*

Key	Description
8	First letter of LAST NAME + letters 3-10 of LAST NAME + letters 2-9 of FIRST NAME + month and year of DOB
9	First letter of LAST NAME + letters 3-10 of LAST NAME + letters 2-9 of FIRST NAME + day and year of DOB
10	Full 8 digits of SSN
11	First 5 letters of LAST NAME + first 4 letters of FIRST NAME + month and year of DOB
12	First 3 letters of LAST NAME + first 3 letters of FIRST NAME + month and year of DOB, switching the first and last name
13	First 3 letters of LAST NAME + first 3 letters of FIRST NAME + day and year of DOB, switching the first and last name
14	First 4 letters of LAST NAME + first 4 letters of FIRST NAME + month and day of DOB, switching the first and last name

*Developed by Bureau of HIV Data Support Unit

27



Syndemic Dataset, NYC, 2000-2013

HIV DATA SET			
HIV ID	Match to TB?	TB Match Key	TB ID
1234	TRUE	1	ABCD
5678	FALSE	0	0
4321	FALE	0	0
8765	TRUE	3	EFGH

TB DATA SET			
TB ID	Match to HIV?	HIV Match Key	HIV ID
ABCD	TRUE	1	1234
EFGH	TRUE	3	8765
JKLM	FALSE	0	0
NOPQ	FALSE	0	0

*2006-2012 for A1C data
Source: NYC DOHMH, Division of Disease Control, PCSI Syndemic Project, 2012

28