Overview of HIV Data Sources

Wendy Garland, MPH
Chief Epidemiologist
Los Angeles County Department of Public Health
Division of HIV and STD Programs

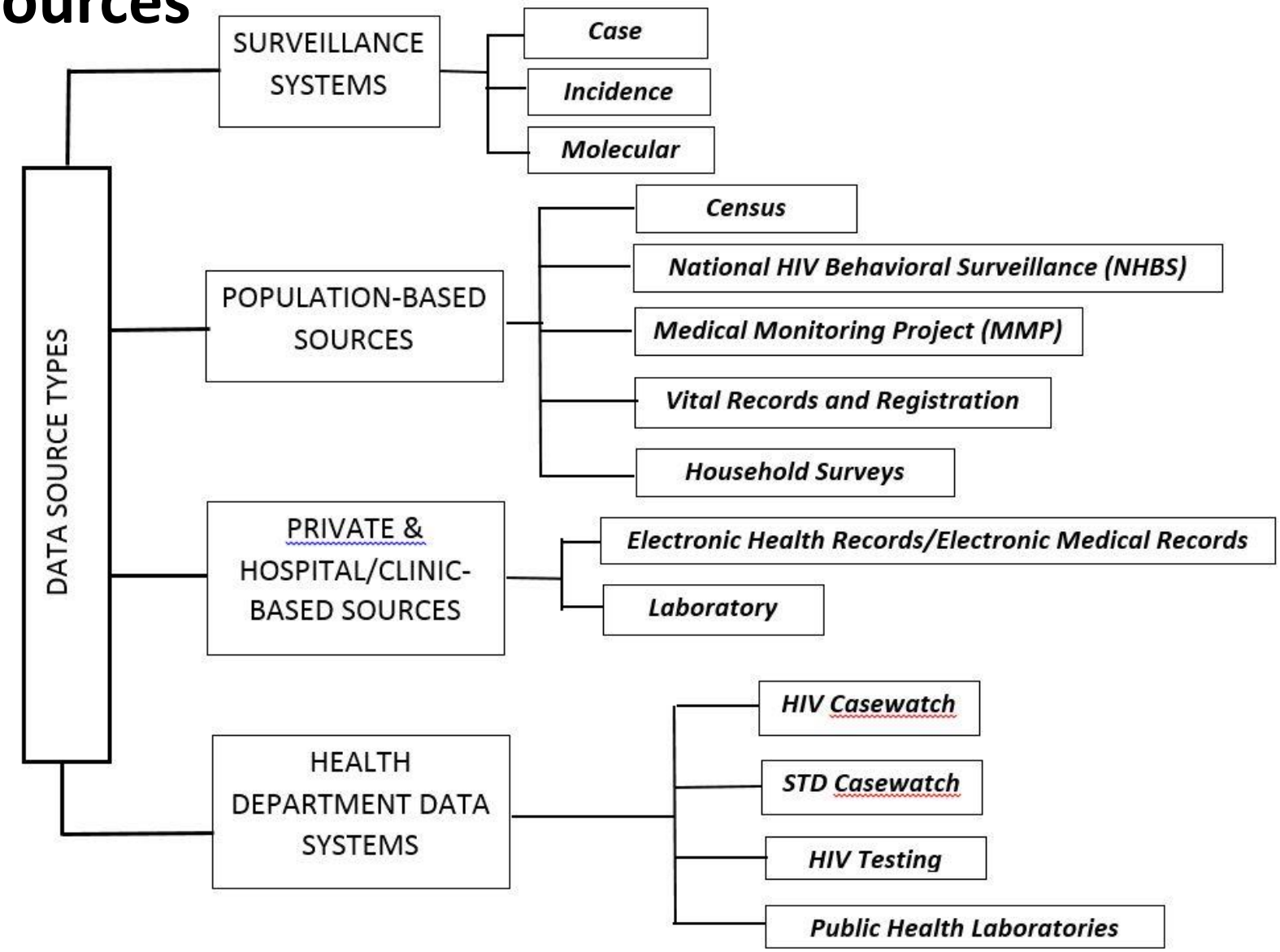
Commission on HIV Data Summit
July 23, 2019

Objectives

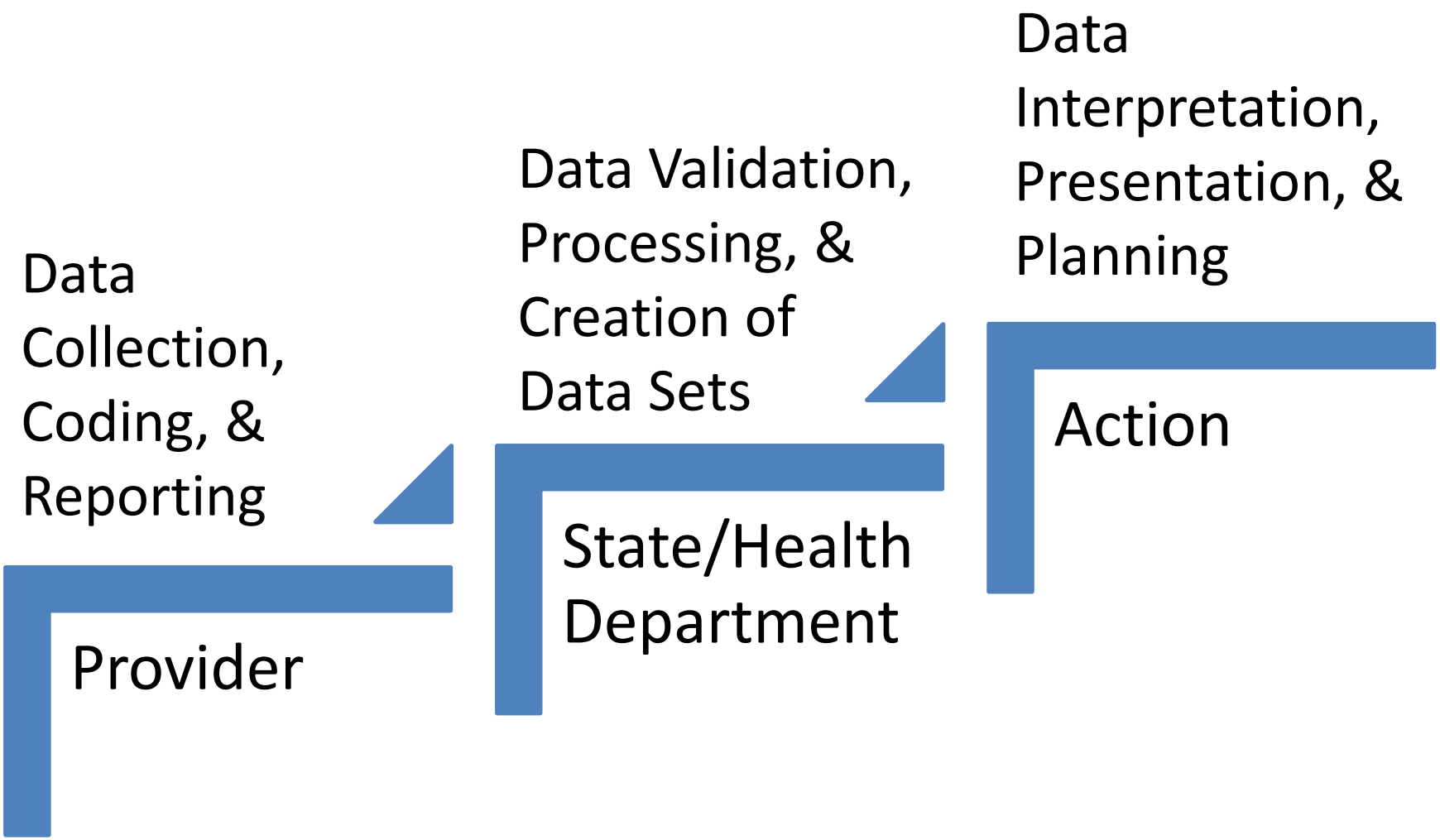
- Appreciate the varied sources and forms of information/data
- Understand the data flow process from data acquisition, validation, creation of data sets, interpretation, and presentation
- Describe the complexity, limitations, and uses of each data source
- Determine which data sources are most appropriate for a particular activity or question

Where do data come from?

Data Sources



From Provider to Action



10 Things you should know about each data source

1) Who is included in the data set?
2) How are the data collected and reported?
3) Where does the data go? Who receives the data?
4) How often are the data submitted?
5) What variables/measures are reported?
6) How are the data processed?
7) How are data validated?
8) How is the final data set created?
9) What question(s) can this data set answer?
10) What are the limitations of this data set?

National HIV Behavioral Surveillance (NHBS)

NHBS was developed to help state and local health departments establish and maintain supplemental surveillance data to monitor selected behaviors and prevention services among groups at highest risk for HIV infection.

Data Process/Step	Explanation or Example
1) Who is the sample population	• Men who have sex with men (MSM) • Persons who inject drugs • Heterosexuals at increased risk for HIV infection
2) Who collects and reports the data?	As of 2019, 22 high prevalence jurisdictions within the US are funded to conduct NHBS. DHSP is funded by CDC to conduct NHBS and collect the data in Los Angeles County.

NHBS(cont.)

Data Process/Step	Explanation or Example
3) Who receives the data?	Data collected by DHSP and sent to CDC.
4) How often are the data submitted?	• Data are collected from each risk group once every three years on a rotating basis. • NHBS data is submitted to CDC on a weekly basis during the data collection months.
5) What data variables or fields are reported?	• Behavioral risk factors for HIV acquisition (e.g. sexual behaviors, drug use) • HIV testing behaviors • Exposure to and use of HIV prevention strategies (e.g. condoms, PrEP). • HIV, HBV and HCV seroprevalence through testing activities

NHBS (cont.)

Data Process/Step	Explanation or Example
6) How are data processed?	• Survey data are collected on laptops in the field and then imported daily into a data warehouse on the secure DHSP network. • Data cleaning and analysis is performed using SAS (a special software program to run statistical analyses).
7) How are data validated?	Data are validated using visual reviews and SAS data verification programs.
8) How is the final dataset created? Do the data need to be matched with any other data sets?	The final dataset is packaged using CDC programs and includes behavioral data and testing data. All NHBS data is anonymous and thus cannot be matched with any other data sets.

NHBS (cont.)

Data Process/Step	Explanation or Example
9) What can we do with the data? What questions can this dataset answer?	NHBS is a population-based data that provides estimates of: 1) high-risk HIV-negative individuals, 2) HIV-positive persons unaware of their infection, and 3) HIV-positive persons aware of their HIV positive status who are in/out of care. Behavioral data in these populations is critical for tracking the epidemic, planning an effective response, and for monitoring and evaluating the local response.

NHBS (cont.)

Data Process/Step	Explanation or Example
10) What are some limitations of these data? What questions cannot be answered by this dataset?	• Self-reported data may be subject to social desirability and recall bias • Limited generalizability of results depending on sampling methods (i.e., venue-based sampling methods may only be generalizable to MSM who attend public venues and respondent-driven sampling methods may only be generalizable to socially networked persons who inject drugs and heterosexuals

Medical Monitoring Project

Data Process/Step	Explanation or Example
1) Who is the sample population?	Persons age 18 years and older who are diagnosed and aware of their HIV infection and has a LAC address in eHARS. In each cycle there are approximately 160 participants. *(Sample of 400 each cycle with about 40% response rate)*
2) Who collects and reports the data?	DHSP
3) Who receives the data?	DHSP sends the data to CDC
4) How often are the data submitted?	Monthly
5) What data variables or fields are reported?	Data come from survey interview and medical record abstraction: • Sociodemographic information • Clinical outcomes, care seeking and care utilization • Service needs and gaps • Substance use/mental health • Sexual behaviors

MMP (cont.)

Data Process/Step	Explanation or Example
6) How are data processed?	CDC processes all data (cleaning, creating calculated variables, applying formats and weighting the data)
7) How are data validated?	CDC has programmed validations in the data collection instruments. CDC also performs data consistency checks and DHSP reconciles these with CDC monthly and at the end of each cycle. DHSP also performs quality checks on 5% of interviews and abstractions.
8) How is the final dataset created? Do the data need to be matched with any other data sets?	Interview data and key variables from the abstraction data are matched and combined into an overlap file with population weights.

MMP (cont.)

Data Process/Step	Explanation or Example
9) What we can do with the data? What questions can this dataset answer?	Provide key estimates for HIV-infected persons diagnosed and aware of their infection in LAC that are comparable to national estimates produced by CDC. Questions: • What are patients' risk behaviors and clinical outcomes? • What are patients' health-related behaviors? • How are patients accessing and using prevention, care, and support services? • What care and treatment are being provided? • Are there variations of factors by respondent characteristics?
10) What are some limitations of these data? What questions cannot be answered by this dataset?	Data are representative of PLWH in LAC but due to sampling strategy and small numbers, estimates by demographics as well as out-of-care persons are small may be unstable and unreliable

HIV Case Surveillance Systems

Data Process/Step	Explanation or Example
1) Which individuals are included in the data?	• Persons newly diagnosed with HIV in LAC and previous diagnoses from LAC or other areas.
2) Who collects and reports the data?	• Medical providers and • Laboratory that process HIV tests and tests ordered by providers for viral load, CD4 and genotype • HIV Surveillance Unit staff in DHSP collects the data.
3) Who receives the data?	• Data first received by DHSP HIV Surveillance Unit staff. • Data also transferred to California State office of AIDS and CDC (with personal identifiers removed).
4) How often are the data submitted?	• Ongoing. • CA HIV reporting regulation requires a case report or a test result of HIV be submitted within 7 days. • Aggregated HIV surveillance summary is available on an annual basis or upon request.

HIV Case Surveillance Systems (cont.)

Data Process/Step	Explanation or Example
5) What data variables or fields are reported?	Data reported through case report and laboratory data • Demographic, HIV transmission category, and residential information • Laboratory test results, physician's diagnoses, treatments, and other clinical information • Childbirth (for women) • HIV care provider • Insurance type and partner services (often incomplete)
6) How are data processed?	• Paper reports entered into Enhanced HIV/AIDS Reporting System (eHARS) • Electronic reports and lab results processed, cleaned, validated, matched, and imported into eHARS

HIV Case Surveillance Systems (cont.)

Data Process/Step	Explanation or Example
7) How are data validated?	• Routine clerical reviews and de-duplication • Daily executed quality check procedures • Monthly random sample check of 5-10% of the cases • Locally developed and CDC-mandated quality check procedures for accuracy, completeness, and timeliness. • Feedbacks of summarized tables to reporting laboratories and providers.
8) How is the final dataset created? Do the data need to be matched with any other data sets?	• Final dataset created with CDC programs at the end of June • Matched with State and CDC HIV registry for de-duplication purposes • Matched with local, state, and national vital registries to identify deceased persons • Matched with STD and TB registries to include unreported cases • Matched with Ryan White program data to estimate HIV outcomes

HIV Case Surveillance Systems (cont.)

Data Process/Step	Explanation or Example
9) What can we do with the data? What questions can this dataset answer?	• Estimate overall HIV disease burden in LAC, in selected geographic areas, or by specific subgroups. • Estimate HIV incidence, prevalence, and related deaths. • Trend of HIV • Estimate HIV care indicators, such as linkage to care, engagement and retention in care, and HIV viral suppression • Data-to-Care activities (locating PLWH not in HIV care)
10) What are some limitations of these data? What questions cannot be answered by this dataset?	• Undiagnosed HIV • Delayed or missed HIV reporting • Deaths among PLWH moved out of county • PLWH move out of LAC

HIV Casewatch

Data Process/Step	Explanation or Example
1) Which individuals are included in the data?	PLWH receiving Ryan White Program services in LAC
2) Who collects and reports the data?	DHSP-supported HIV care services providers
3) Who receives the data?	DHSP
4) How often are the data submitted?	As required by contract
5) What data variables or fields are reported?	Demographic, services received, assessment data (some service categories)
6) How are data processed?	Vendor managed
7) How are data validated?	Vendor managed

HIV Casewatch (cont.)

Data Process/Step	Explanation or Example
8) How is the final dataset created? Do the data need to be matched with any other data sets?	SAS created final dataset. HIV Casewatch data are matched with HIV surveillance data monthly.
9) What can we do with the data? What questions can this dataset answer?	• HIV care outcomes among PLWH in RWP • Evaluation of RWP care system and service categories • Service utilization
10) What are some limitations of these data? What questions cannot be answered by this dataset?	• Not population based • Limited to Ryan White program clients

Questions??