National HIV Surveillance System (NHSS)

Attachment 4d.

Supplemental Surveillance Activity 2: Molecular HIV Surveillance (MHS) Technical Guidance

Technical Guidance for HIV Surveillance Programs

Molecular HIV Surveillance (MHS)

HIV Incidence and Case Surveillance Branch
Atlanta, Georgia

CONTENTS — MOLECULAR HIV SURVEILLANCE (MHS)

Introduction
- Objectives
- Disease surveillance activity

General Concepts
- Antiretroviral therapy (ART)
- Molecular epidemiology
- HIV drug resistance HIV genetic diversity
- HIV transmission patterns

Structural Requirements
- Policies and procedures
- Staffing needs

Process Standards
- Secure and confidential reporting of HIV surveillance data
- Reviewing and revising (if applicable) state or local HIV laws and regulations regarding the reporting of HIV nucleotide sequence data
- Collaborating with laboratories performing HIV genotypic drug resistance testing
- Obtaining HIV nucleotide sequence data from all genotyping tests performed for persons diagnosed
    - Electronic laboratory-based reporting (ELR)
        - Transmission with other HIV surveillance data
        - Transmission through the National Electronic Disease Surveillance System (NEDSS)
        - Transmission using Secure File Transfer Protocol (SFTP)
    - Non-ELR methods
    - Electronic file formats
- Validating HIV nucleotide sequence data
- Importing nucleotide sequence data into eHARS
- Collecting antiretroviral use history data for all newly diagnosed persons
- Transferring data monthly to CDC
- Ongoing monitoring of data quality and evaluation of data
- Analyzing data, including transmitted drug resistance data
- Developing and disseminating data reports and presentations

Outcome Standards

References

Suggested Readings

Appendices
- History and timeline of MHS
- Glossary of terms
- Acronyms

INTRODUCTION

The Centers for Disease Control and Prevention (CDC) maintains the National HIV Surveillance System (NHSS) and collects data on diagnoses of HIV infection for national, state, and local HIV program planning and evaluation. As an integral component of the NHSS, Molecular HIV Surveillance (MHS) —formerly referred to as Variant, Atypical, and Resistant HIV Surveillance (VARHS) —incorporates the collection of HIV nucleotide sequences into routine case reporting in the jurisdictions that participate in this activity. Together with case report data, HIV sequences can be used to achieve three primary goals: (1) to assess prevalence and trends in acquired and transmitted HIV drug resistance; (2) to evaluate HIV genetic diversity; and (3) to describe HIV transmission patterns for the purpose of evaluating the impact of HIV prevention strategies, guiding public health action, and enhancing the understanding of the burden of HIV in the United States. (Unless otherwise noted, all references to HIV in this document refer to HIV-1 infection.)

The objectives of MHS are as follows:

- Collect all HIV nucleotide sequence data from laboratories that perform HIV genotypic drug resistance testing;
- Use molecular epidemiologic techniques to assess HIV drug resistance, evaluate HIV genetic diversity, and describe HIV transmission patterns; and
- Disseminate results of molecular HIV data analyses to assist HIV treatment, prevention, and program planning and evaluation.

MHS offers a unique perspective about HIV and provides community-level information about the impact of prevention strategies to advance the understanding of the HIV burden in the surveillance areas that conduct MHS. MHS activities support the National HIV/AIDS Strategy goals of (1) reducing new HIV infections through the potential use of nucleotide sequence data to determine duration of infection for monitoring incidence; (2) increasing access to care and improving health outcomes by using nucleotide sequence data as a marker for linkage to and quality of care, and (3) reducing HIV-related disparities and health inequities by using nucleotide sequence data to reveal transmission patterns and provide insight into prevention.

Disease Surveillance Activity

In 2004, CDC determined that the collection of HIV nucleotide sequence data as a part of the National

HIV Surveillance System was a non-research disease surveillance activity and that a review by the Institutional Review Board, pursuant to Title 45 Code of Federal Regulations Section 46: Protection of Human Subjects, was not required.

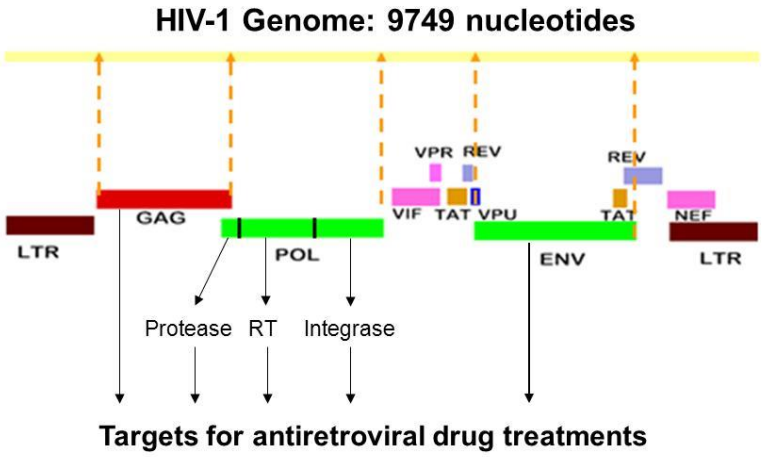
GENERAL CONCEPTS

Antiretroviral therapy (ART)

Current treatment of HIV disease has been designed and optimized for HIV-1 and includes five classes of antiretroviral drugs (ARVs). These ARVs have been approved by the Food and Drug Administration (FDA) for use in the treatment or prophylaxis of HIV infection and prevention of HIV replication at various stages of its life cycle (Figure 1). They include the following:

- Nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs), which prevent the HIV reverse transcriptase (RT) from transcribing HIV RNA to viral DNA;
- Protease inhibitors (PIs), which prevent the maturation of HIV proteins by the HIV protease;
- Integrase strand transfer inhibitors (INSTIs), which block the HIV integrase from integrating viral DNA into the genome of host cells with CD4 receptors; and
- Entry/fusion inhibitors (EIs), which prevent the HIV envelope (env) from binding to and allowing the HIV genome to enter the human cell.

Figure 1. Map of HIV-1 Genome and ARV Targets



Appropriately timed and consistent use of ARVs can suppress viral loads, which can lead to better health outcomes and a much lower chance of passing HIV on to partners (1). Lists of ARVs and recommended

HIV treatment regimens are available at:
http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf.

Molecular Epidemiology

Molecular epidemiology is the use of molecular biology techniques to study the factors associated with disease distribution. With regards to HIV, molecular techniques, such as viral sequencing and phylogenetic analyses, can be used to assess HIV drug resistance, evaluate HIV genetic diversity, and describe the dynamics of HIV infection, including the characterization of transmission patterns. Results of such analyses enhance the understanding of HIV infection and its spread through various populations and can substantially contribute to efforts to intervene to prevent HIV transmission.

HIV Drug Resistance

HIV has a rapid reproduction rate and a very high mutation rate. Because HIV lacks the mechanism to correct mistakes that occur as it reproduces, HIV in an infected person exists as a set of related, but non-identical viruses, some of which are intrinsically drug resistant even in the absence of ARV pressure (i.e., contain natural polymorphisms). Expanded and suboptimal uses of ART have also contributed to the emergence of drug-resistant strains of HIV, resulting in suboptimal virologic responses, treatment failures, acquired HIV drug resistance among infected persons, and transmission of drug-resistant strains of HIV to drug-naïve persons. CDC estimates that one in six individuals newly diagnosed with HIV infection and reported with nucleotide sequence data was infected with a strain that contained transmitted drug resistance mutations associated with at least one antiretroviral drug class (2).

HIV genotypic assays evaluate HIV sequences to detect the presence of mutations associated with antiretroviral drug resistance. Standard HIV genotypic assays extract viral RNA from the blood plasma of an infected person and amplify regions of the HIV genome targeted by ARVs, mainly the PR and RT genes of the *pol* region (Figure 1). Other areas of the HIV genome, the integrase gene in the *pol* region and the envelope gene, can also be tested for mutations associated with antiretroviral drug resistance. A nucleotide sequence of the person's viral isolate is translated into corresponding amino acids and compared to a "wild-type" reference strain to identify mutations associated with drug resistance. A final report is generated that provides an interpretation of the level of resistance detected and can be used by health care providers in the clinical management of patients.

HIV genotypic testing is a part of the standard of care for persons infected with HIV. The U.S. Department of Health and Human Services and the International Antiviral Society-USA (IAS-USA)

recommend that HIV genotypic testing be performed at entry into care and again as needed to guide treatment (3,4).

HIV Genetic Diversity

There are two types of HIV: HIV-1 and HIV-2. Most HIV infections in the United States and around the world are caused by HIV-1. More than 90% of HIV-1 infections belong to Group M, of which at least 9 genetically distinct subtypes (or clades) are known: A, B, C, D, F, G, H, J and K. There are also "circulating recombinant forms" (CRFs) derived from the merging of different subtypes, "unique recombinant forms" (URFs), and numerous unknown variants.

Standard genotyping methods that identify mutations associated with antiretroviral drug resistance can be used to distinguish between subtype B and non-B variants of HIV. Phylogenetic analyses can also be applied to evaluate the genetic relationships between HIV strains and further classify potential non-B variants to the other subtypes, CRFs, and potential URFs.

HIV Transmission Patterns

Phylogenetic analysis and other analytical methods can be used to describe transmission patterns among persons who differ demographically, represent various risk groups, and reside in diverse geographical areas. Identification and characterization of potential transmission networks, consisting of persons infected with highly similar HIV sequences, can guide prevention efforts and optimize the allocation of resources.

STRUCTURAL REQUIREMENTS

Policies and Procedures

All persons diagnosed with HIV infection and all HIV-related laboratory test results (including, where applicable, results from HIV genotype tests) should be reported to HIV surveillance programs as required by local and state laws and regulations. Upon receipt, HIV surveillance staff should enter or import the information into eHARS in accordance with the *Technical Guidance for HIV Surveillance Programs, Vol. I: Policies and Procedures*. Because MHS is a component of HIV case surveillance, surveillance staff, in collaboration with CDC, should build upon the existing infrastructure to collect the information needed to achieve MHS objectives, including an implementation plan, a timeline for implementation, and the incorporation of MHS policies and procedures into local guidance documents.

MHS-specific policies and procedures should include information related to the following elements:

- Staff responsibilities and requirements
- Security and confidentiality
- Laws and regulations regarding the reporting of HIV nucleotide sequence data
- Collection of HIV nucleotide sequence data from laboratories
- Collection of ARV use history data
- Data management practices
- Data transmissions to CDC
- Analysis and dissemination
- Program evaluation

Staffing Needs

To implement MHS successfully, HIV surveillance staff should integrate MHS activities into the existing HIV surveillance system and infrastructure. The number of staff needed to conduct MHS activities at participating jurisdictions depends on the phase of program implementation, HIV morbidity, and resource availability. At a minimum, HIV surveillance staff should ensure that the following activities related to project coordination, surveillance and epidemiology, and data management are conducted:

Programmatic: Project Coordination

- Overall management of MHS to ensure integration of MHS activities into the existing HIV surveillance system
- Compliance with CDC and HIV surveillance security and confidentiality requirements
- Collaboration with appropriate staff to review and revise (if applicable) current laws and regulations to include HIV nucleotide sequences as test results indicative of HIV infection
- Development and implementation of MHS activities, including local MHS guidance, policies, and procedures
- Collaboration with laboratories to revise existing processes to include the electronic transfer of HIV nucleotide sequence data to the HIV surveillance program
- Collection of antiretroviral use history data for all newly diagnosed persons as part of HIV case reporting, including training of all data collectors and timely and accurate entry of these data on the Testing and Treatment History Document in eHARS

- Continued monitoring of MHS activities to ensure that process and outcome standards are achieved

Technical: Data Management

- Securely store and manage and timely transfer MHS data according to CDC requirements
- Development of processes to ensure that MHS meets or exceeds data quality outcome standards
- Ongoing proficiency in Base SAS software and other locally-required data management/analysis applications

Scientific: Surveillance and Epidemiology

- Analysis of MHS data
- Development and dissemination of MHS data reports and publications

PROCESS STANDARDS

MHS activities involve the following general processes:

- Secure and confidential reporting of HIV surveillance data
- Reviewing and revising (if applicable) state or local laws and regulations regarding the reporting of HIV nucleotide sequence data
- Collaborating with laboratories performing HIV genotypic drug resistance testing
- Obtaining HIV nucleotide sequence data from all genotyping tests performed for persons diagnosed
- Validating HIV nucleotide sequence data received
- Importing nucleotide sequence data into eHARS
- Collecting antiretroviral use history data for all persons newly diagnosed with HIV infection
- Transferring data monthly to CDC
- Ongoing monitoring of data quality and evaluation of local activities
- Analyzing data, including transmitted drug resistance
- Developing and disseminating data reports and presentations

Secure and confidential reporting of HIV surveillance data

The secure and confidential transfer of any HIV-related data, including HIV nucleotide sequence data, is of utmost importance. HIV surveillance programs should review existing state and local policies on maintaining patient confidentiality and develop specific steps that will be taken to secure the data. These

standard operating procedures should be reviewed and approved by the overall responsible party (ORP) and should be consistent with the guidelines and standards specified in *Data Security and Confidentiality Guidelines for HIV, Viral Hepatitis, Sexually Transmitted Disease, and Tuberculosis Programs: Standards to Facilitate Sharing and Use of Surveillance Data for Public Health Action*.

All staff responsible for the transmission and receipt of HIV surveillance data, including HIV nucleotide sequence data obtained through MHS, must be trained in the security and confidentiality procedures for HIV surveillance.

Reviewing and revising (if applicable) state or local HIV laws and regulations regarding the reporting of HIV nucleotide sequence data

HIV surveillance programs conducting MHS should review existing state laws and regulations regarding the reporting of HIV test results to the health department. If HIV nucleotide sequence data are not specifically referenced on the list of reportable HIV laboratory test results, staff should consult the state office of legal counsel to determine whether current laws and regulations can be interpreted to include the reporting of HIV nucleotide sequence data (e.g., "all tests indicative of HIV infection and care"). Otherwise, consider modifying existing laws and regulations to ensure the reporting of HIV nucleotide sequence data as another HIV diagnostic test needed for HIV surveillance.

Additional information about reviewing reporting laws is available in the chapter on *Reporting* in the *Technical Guidance for HIV Surveillance Programs, Volume I: Policies and Procedures*.

Collaborating with laboratories performing HIV genotypic drug resistance testing

MHS relies on collaboration with laboratories performing HIV genotypic testing. HIV surveillance programs conducting MHS should identify public, commercial, and private laboratories that perform HIV genotypic testing for area residents. Sources of this information include Clinical Laboratory Improvement Amendments (CLIA) state and regional offices and state and local licensing boards for clinical laboratories.

HIV surveillance programs should consider developing a survey to collect information that will enhance the collaboration between the laboratories and the HIV surveillance program. Examples of topics to be addressed include the following:

- Contact information for key laboratory personnel, both administrative and technical
- Whether HIV genotypic testing occurs on-site or at a reference laboratory, and document how

their HIV genotypic testing system can generate and export nucleotide sequence data in standard text-based file formats

- Level of information technology (IT) support available

Laboratory staff should be informed about the purpose of MHS and made aware of any laws or regulations that require the reporting of HIV nucleotide sequence data to the HIV surveillance program. Collaboration between the laboratory staff and the HIV surveillance staff to fulfill MHS activities should be encouraged.

Additional information about laboratories and reporting of data is available in the chapter on *Reporting* in the *Technical Guidance for HIV Surveillance Programs, Volume I: Policies and Procedures*.

Obtaining HIV nucleotide sequence data from all genotyping tests performed for persons diagnosed

HIV surveillance programs conducting MHS should work with laboratories to obtain HIV nucleotide sequence data from all genotyping tests performed for persons diagnosed in the jurisdiction. Existing processes can be modified to incorporate HIV nucleotide sequence data into the routine transfer of other HIV laboratory results.

Laboratories that perform HIV genotypic testing use HIV genotyping systems to analyze HIV nucleotide sequence data. The primary output of most of these systems is a genotype report that provides a drug resistance interpretation for clinical use. The HIV nucleotide sequence, which is the basis of MHS activities, is an intermediate product of these HIV genotypic testing systems. HIV surveillance programs should work with laboratory staff to develop procedures for extracting HIV nucleotide sequence data from the HIV genotypic testing system. Sequence data are often not stored with other laboratory test results, and the HIV genotypic testing system that holds the sequence data may not be connected to the laboratory's main computer network. Procedures should be developed that link the sequence data extracted from the genotypic system to standard patient demographic information and other data submitted to the laboratory (e.g., collection date, ordering physician, and facility). Additional linking procedures may be necessary if a reference laboratory is used to conduct the HIV genotypic testing, as the reference laboratory may not have access to the standard patient and provider data. Close collaboration with the laboratory administration and technical staff, including IT staff, is needed to extract the HIV nucleotide sequence data from HIV genotypic testing systems. If the laboratory staff are not familiar with the data export capabilities of the genotyping system, the manufacturer of the HIV genotypic testing system should be consulted.

Depending on the surveillance program's infrastructure, technical capabilities, and state reporting laws and regulations, surveillance staff might rely on a combination of the data transfer methods described below. Ideally, HIV nucleotide sequence data should be transferred to surveillance programs in electronic formats, which facilitate efficient entry into eHARS.

Electronic Laboratory-based Reporting (ELR)

Electronic Laboratory-based Reporting (ELR) is the electronic transfer of public health data from clinical laboratories to public health agencies in pre-established formats (e.g., HL7 messaging, ASCII, spreadsheet) that do not require extensive manual data entry, cutting and pasting, or translation to add to a database. Ideally, data transmitted by ELR should be automated and should use standardized codes for tests (e.g., Logical Observation Identifiers Names and Codes or LOINC) and results, allowing for timely, complete, and accurate reporting. Additional information about ELR is available in the chapter on *Reporting* of the *Technical Guidance for HIV Surveillance Programs, Volume I: Policies and Procedures*. Possible methods of receiving HIV nucleotide sequence data from laboratories include coordinated transmission with other HIV surveillance data (e.g., CD4 and viral load results), coordinated transmission with data from existing state systems, and transmission of sequence data from the laboratory directly using the Secure File Transfer Protocol (SFTP).

Transmission with other HIV surveillance data

HIV surveillance programs that use ELR to receive HIV serology data, CD4 data, and viral load data, may receive HIV nucleotide sequence data through this same mechanism. If the HIV surveillance program does not use ELR, other surveillance programs within the health department (e.g., general infectious diseases, TB, STDs) may use ELR and can include HIV nucleotide sequence data. Surveillance staff should consult with staff within these programs and with IT staff at the health department about the possibility of modifying existing ELR processes that ensure the secure and confidential receipt of HIV nucleotide sequence data.

Transmission through existing state systems

The CDC Public Health Information Network (PHIN) is a national initiative to improve the capacity of public health to use and exchange information electronically by promoting the use of standards and defining functional and technical requirements. As a component of PHIN, CDC has developed the National Electronic Disease Surveillance System (NEDSS), an Internet-based infrastructure for public health surveillance data exchange. The NEDSS Messaging solution (NMS) supports electronic messaging

between public health partners (e.g., commercial entities and local, state, and federal agencies) and relies on industry standards (i.e., LOINC, SNOMED, and HL7), policy-level agreements on data access, and the protection of confidentiality. Additional information on NMS is available at: http://www.cdc.gov/phin/library/phin_fact_sheets/111759_NMS_NEDSS.pdf. Additional information on NEDSS and HIE are available in the chapter on *Reporting of the Technical Guidance for HIV Surveillance Programs, Volume I: Policies and Procedures*.

HIV surveillance programs that are considering receiving HIV nucleotide sequence data through their state NEDSS program should coordinate with NEDSS program staff to make relevant modifications in NEDSS. Upon meeting security and confidentiality guidelines, HIV nucleotide sequence data can be routed through NEDSS to the HIV surveillance program. Programs that have implemented a system that uses standards similar to NEDSS or receive healthcare information electronically from different organizations through health information exchanges (HIE) can also consider the possibility of receiving HIV surveillance data via these mechanisms.

Additional information about NEDSS and HIE are available in the chapter on *Reporting* of the *Technical Guidance for HIV Surveillance Programs, Volume I: Policies and Procedures*.

Transmission using Secure File Transfer Protocol (SFTP)

A common method used by health departments to receive HIV surveillance data is the Secure File Transfer Protocol (SFTP). SFTP is a type of protocol that provides a set of rules that govern the syntax, semantics, and synchronization of communication across computer networks. SFTP encrypts both commands and data, requires certification on at least the sending or receiving end, and allows the secure transmission of passwords and sensitive information.

To establish SFTP transmission, both the host laboratory and the HIV surveillance program should maintain a SFTP server, preferably located within a secure area, behind a firewall, and with other physical security measures to prevent access by non-authorized staff. Authorized staff within the HIV surveillance program may be granted permission to log into a laboratory's host network, access the host server using a password, retrieve the sequence data files (e.g., HL7 file or FASTA file) and initiate the transfer of the files from the laboratory to a SFTP server located at the health department. Alternatively, authorized laboratory staff may be granted permission to access and transfer the data files to the secure server at the health department. Regardless of the process chosen, close collaboration between the laboratory staff, HIV surveillance staff, and staff from both IT departments is essential.

Ideally, electronic data transfers should be conducted over a secure data network (SDN), virtual private network (VPN) connection with certificates on both the sending and receiving ends, or a similar secure connection. At a minimum, when transferring data electronically, surveillance programs should encrypt data using a secure application such as the SFTP described above. If either the sender or the recipient of the data is not part of a defined security zone appropriate for sensitive data, the data should be encrypted by a method that meets the Federal Information Processing Standard for Advanced Encryption Standard, or FIPS 140-2, before transmission (5).

HIV surveillance programs that have not yet implemented ELR are strongly encouraged to do so. Programs should contact their assigned CDC HIV Incidence and Case Surveillance Branch epidemiologist and request assistance in establishing ELR for their HIV surveillance program. Refer to the ELR section in the chapter on *Reporting* of the *Technical Guidance for HIV Surveillance Programs, Volume I: Policies and Procedures*.

Non-ELR methods

If ELR is not a feasible option, HIV surveillance programs can develop other methods to obtain HIV nucleotide sequence data from the laboratory to the health department. Acceptable methods of data storage and transport include CDs, DVDs, flash drives, or external hard drives, provided that the data have been encrypted to meet FIPS 140-2 requirements. The selected procedures should also comply with security and confidentiality requirements at the state and CDC.

All removable or external storage devices containing identifiable public health data must include only the minimum amount of information necessary to accomplish assigned tasks as determined by the designated official or ORP; must be encrypted and stored under lock and key when not in use; and must be sanitized immediately following a given task (except for those used as back-ups). Methods used to sanitize a storage device must ensure that any data on the device cannot be retrieved by using "undelete" or data retrieval software. Such policies and procedures should be consistent with the guidelines and standards specified in *Data Security and Confidentiality Guidelines for HIV, Viral Hepatitis, Sexually Transmitted Disease, and Tuberculosis Programs: Standards to Facilitate Sharing and Use of Surveillance Data for Public Health Action*.

Validating HIV nucleotide sequence data

Data validation should be implemented and incorporated into existing data processes to ensure that the HIV nucleotide sequence data received are of good quality. Reported nucleotide sequences vary in length, and HIV surveillance programs should be able to identify potential problems. (Refer to the table in the next section for maximum expected lengths by genotype test type.)

HIV surveillance staff should develop internal processes for validating the sequences, including:

- Using LOINC to compare the length of sequences received to expected lengths. Although HIV genotyping systems generate sequences that can sometimes vary in length, sequences outside of the system's expected length or outside of a laboratory's usual range of variation may indicate a problem with transmission.
- Assessing sequences for embedded, non-sequence characters. Some characters may be legitimate, while others are not and may interfere with proper interpretation of the sequence. Sequences that contain numbers, letters other than ATUCGNRWYMKSHBVDX, and/or punctuation other than ~!#*-.()' might indicate a problem.

Importing nucleotide sequence data into eHARS

In January 2013, CDC released eHARS 4.0, which stores and manages nucleotide sequence data. eHARS 4.0 and later allows for the import of sequences up to 10,000 characters for five genotype test types:

Genotype test type in eHARS 4.0 and later	Maximum length for test type
HIV-1 Genotype (PR Nucleotide Sequence)	297
HIV-1 Genotype (RT Nucleotide Sequence)	1,680
HIV-1 Genotype (PR/RT Nucleotide Sequence)	864
HIV-1 Genotype (IN Nucleotide Sequence)	1,977
HIV-1 Genotype (PR/RT/IN Nucleotide Sequence)	2,841

Through the NEDSS program, CDC has provided most NEDSS-funded jurisdictions with the Rhapsody Integration Engine (Orion Health, CA), a software tool that can be used to read laboratory data in various formats. HIV surveillance programs using ELR methods can use this software, or an alternative software, to parse laboratory data reported through ELR and translate the laboratory results into a format that can be imported into eHARS. Unmatched sequence data may indicate a reporting delay in eHARS or may indicate a new case that has not been reported to the HIV surveillance system; additional follow-up or field investigation of the case may be warranted.

Sequences should be imported into eHARS in their original format. Therefore, nucleotide sequences that are received as separate records should be stored as separate laboratory results for the same individual. Inclusion of LOINC is important to distinguish the type of HIV nucleotide sequence generated by the HIV genotype test. The data may exist as two separate nucleotide sequences from the PR and/or RT genes of the *pol* region, and the data may also exist as a single, combined PR and RT sequence. The paired vs. single-record format of the result depends on the genotyping system used by the laboratory performing the test. Though rare, the data may also represent the integrase gene of the *pol* region, and sequences may exist as separate or combined PR, RT, and integrase sequences.

Values entered in the genotype sequence field are compressed during importing, which eliminates new line feeds, carriage control characters, literal spaces, and line breaks. Several validation checks and error warnings have also been incorporated into eHARS to help ensure good quality of MHS sequences:

Validation /Edit Check Rule	Error Message	Comment
Up to 10,000 characters may be entered per nucleotide sequence	"You entered a genotype test with more than 10,000 characters."	eHARS rejects the record
These characters are allowed in the genotype sequence field: AaTtUuCcGgNnRrWwYyMmKkSsHhBbVvDdXx~!#*-.()'	"You entered a nucleotide sequence that contained non-standard characters."	eHARS accepts the record
Genotype test name is present, but the corresponding collection date and genotype sequence fields do not have values	"The lab test being added is missing a collection date and a genotype sequence. Lab tests missing collection dates and genotype sequences are excluded from the processes that generate calculated variables."	A record is not created on the LAB_GENOTYPE table for the imported record.
Collection date is populated for a genotype test, but a genotype sequence is not	"The lab genotype test being added is missing a genotype sequence. Lab genotype tests missing genotype sequences are excluded from the processes that generate calculated variables."	A record is not created on the LAB_GENOTYPE table for the imported record.
Length of the entered string exceeds values established for certain tests	"You attempted to enter an <test type> with more than <max length of test> characters."	eHARS accepts the record

Surveillance staff can contact the Division of HIV/AIDS Prevention Helpdesk at dhapsupport@cdc.gov for assistance with the importing of nucleotide sequence data into eHARS. Staff can also refer to the *eHARS 4.0 (or later) Technical Reference Guide* for additional information about genotype sequence data in eHARS.

Collecting antiretroviral use history data for all newly diagnosed persons

Information on the prior use of ARVs for all newly diagnosed persons is needed to assess the prevalence of acquired and transmitted HIV drug resistance. HIV surveillance programs can incorporate the collection of ARV use history data into routine HIV case reporting using the CDC Adult HIV Confidential Case Report Form or other forms (e.g., state-based case report forms or National HIV Monitoring and Evaluation Program [NHM&E] form). Sources of the information may be patient self-reports; medical charts that contain physician's notes, laboratory reports, and pharmacy records; and the AIDS Drug Assistance Program (ADAP) records. Surveillance staff reporting ARV use data should comply with reporting procedures that meet the routine security and confidentiality guidelines for HIV surveillance.

Information on ARV use should be entered into eHARS on the Testing and Treatment History (TTH) Document. To be consistent with the principles of document-based data entry, new eHARS documents should be used to enter data from multiple forms, from multiple sources, or to update previously entered information.

The TTH variables related to ARV use in eHARS are listed below.

- **Ever Taken Any Antiretroviral Medications —** this variable is used to determine whether the patient took any ARVs at any time. The information collected should indicate whether antiretroviral drugs have ever been used (i.e., no time limit should be placed on the history of antiretroviral drug use) and the date on which that information was obtained.
- **Name(s) of ARV Medication Taken —** this variable lists at least one of the ARVs that the patient has taken, but may not include all medications used, and verifies that at least one medication taken was an ARV. A list of current medications used to treat HIV is available at http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf.
- **Date ARVs First Began —** this date represents the earliest date of any ARV use and determines whether the patient took any ARVs before the collection date of the specimen from which the nucleotide sequence was obtained.

- **Date of Last ARV Use —** this variable represents the date when ARVs were last taken by the patient.

Additional information about the collection of ARV use history data and entry of these data into eHARS can be found in the *Guidance for Collection and Data Entry of HIV Incidence Surveillance Information*.

Transferring data monthly to CDC

HIV surveillance programs are expected to transmit HIV surveillance data, including MHS data, to CDC each month via standard CDC data transfer processes. Data transmitted to CDC must be encrypted and password protected and must not include personal identifiers as specified in *Data Security and Confidentiality Guidelines for HIV, Viral Hepatitis, Sexually Transmitted Disease, and Tuberculosis Programs: Standards to Facilitate Sharing and Use of Surveillance Data for Public Health Action*.

Ongoing monitoring of data quality and evaluation of data

HIV surveillance programs should review and determine the completeness and quality of the MHS data received, including, but not limited to, demographic, HIV transmission risk, clinical, and laboratory data elements.

Analyzing data, including transmitted drug resistance

CDC uses a regularly updated program to analyze the HIV nucleotide sequence data submitted by HIV surveillance programs. HIV-1 *pol* sequences are screened for subtype B (cut-off 90%) and potential non-B variants using Sierra—The Stanford HIV Web Service, Version 1.0 (http://hivdb.stanford.edu/pages/webservices/). This program translates the nucleotide sequence data and incorporates information on individual mutations of interest, the level of resistance to each antiretroviral drug in common use, and HIV-1 subtype. Sequences classified to potential non-B variants are further analyzed and assigned to non-B subtypes, CRFs, and URFs. Records that meet CDC data quality standards are aggregated for national analyses, including analyses on transmitted drug resistance-associated mutations (TDRMs).

Persons are classified as having sequences that contain transmitted drug resistance-associated mutations (TDRMs) based on the CDC HIV-1 surveillance mutation list if the following criteria are met:

- The nucleotide sequence is from a specimen that was drawn within three months after the date of collection of the diagnostic specimen (i.e., the HIV-positive specimen that led to the report in eHARS)

and

- The person has no evidence of prior ARV use (as determined by ARV history use data).

As deemed appropriate, CDC will provide a local dataset and accompanying SAS program for HIV surveillance programs to conduct local data analyses. Alternatively, programs that have staff with knowledge of Perl, Java, and XML can choose to process the sequence data by installing and using the Sierra Web service (or a similar service meeting security and confidentiality guidelines). Staff should also have advanced SAS skills to locally develop their own SAS programs to read in outputs from processed sequences, apply surveillance mutation lists, and analyze the data.

Developing and disseminating data reports and presentations

As appropriate, results of national data analyses will be presented at conferences and published in peer-reviewed journals. HIV surveillance programs conducting MHS should also disseminate the results of local analyses through surveillance reports and presentations to assist HIV treatment, prevention, and program planning and evaluation.

OUTCOME STANDARDS

Outcome standards described in the *Introduction to Policies and Procedures*, *Data Quality*, and *Reporting* chapters of *Technical Guidance for HIV Surveillance Programs, Vol. I: Policies and Procedures* should be applied to MHS. Meeting the surveillance standards for case ascertainment, data quality, timeliness, and completeness are essential to the success of MHS. The outcome standards for MHS relate only to persons who resided in the jurisdiction at the time of diagnosis.

For each calendar year:

- ≥ 50% of newly diagnosed persons will have an initial HIV nucleotide sequence (i.e., obtained from a specimen collected for HIV genotype [resistance] testing within 3 calendar months following HIV diagnosis) reported to the National HIV Surveillance System within 12 months following diagnosis.
- ≥ 85% of newly diagnosed persons with an initial HIV nucleotide sequence will have ARV use data reported to the National HIV Surveillance System within 12 months following diagnosis.

REFERENCES

1. Centers for Disease Control and Prevention. Vital Signs: HIV Prevention through Care and Treatment — United States. *MMWR* 2011;60:1618-1623. Accessed July 29, 2013.
2. Kim D, Ziebell R, Saduvala N, Kline R, Ocfemia MCB, Prejean J, Zhang X, Pearson M, Hall IH. Trend in Transmitted HIV-1 ARV Drug Resistance-associated Mutations: 10 HIV Surveillance Areas, US, 2007–2010. Presented at: 20th Conference on Retroviruses and Opportunistic Infections; 2013 Mar 3–6; Atlanta, GA. Available at http://www.retroconference.org/2013b/Abstracts/47795.htm. Accessed July 29, 2013.
3. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services. Available at http://aidsinfo.nih.gov/contentfiles/lvguidelines/adultandadolescentgl.pdf. Accessed July 29, 2013.
4. Hirsch MS, Günthard HF, Schapiro JM, Brun-Vézinet F, Clotet B, Hammer SM, et al. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society–USA Panel. Clin Infect Dis. 2008 Jul 15;47(2):266-85.
5. National Institute of Standards and Technology. Federal information processing standards publication 140-2: Announcing the standard for security requirements for cryptographic modules. Federal Information Processing Standards Publication 140-2, May 25, 2001. Available at http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf. Accessed July 29, 2013.

SUGGESTED READINGS

- Wheeler W, Ziebell R, Zabina H, Pieniazek D, Prejean J, Bodnard U, et al. Prevalence of transmitted drug resistance-associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, U.S.–2006. AIDS. 2010 May 15;24(8):1203-12.
- Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med. 2011 Aug 11;365(6):493–505.
- Pieniazek D, Ziebell R, Kline R, Prejean J, Hall I. Evaluation of Using Dried Fluid Spots for HIV-1 Drug Resistance Genotyping in the United States. Presented at: 2011 National HIV Prevention Conference; 2011 Aug 14–17; Atlanta, GA. Available at http://www.2011nhpc.org/archivepdf/2011%20NHPC%20Final%20Abstract%20Book.pdf.
- Pieniazek D, Ziebell R, Ocfemia CB, Saduvala N, Kline R, Prejean J, Kim D, Hall HI. Phylogenetic Surveillance of HIV-1 non-B Subtypes among HIV-1 New Diagnoses, United States, 2006-2008. Presented at: 2011 National HIV Prevention Conference; 2011 Aug 14–17; Atlanta, GA. Available at http://www.2011nhpc.org/archivepdf/2011%20NHPC%20Final%20Abstract%20Book.pdf.
- Ocfemia MCB, Kim D, Ziebell R, Prejean J, Saduvala N, Pieniazek D, Heneine W, Kline, Hall IH. Prevalence and Trends of Transmitted Drug Resistance-associated Mutations by Duration of Infection among Persons Newly Diagnosed with HIV-1 Infection: 5 States and 3 Municipalities, US, 2006 to 2009. Available at http://www.retroconference.org/2012b/PDFs/730.pdf. Accessed 2012 Mar 12.
- Pieniazek, D, Ziebell R, Ocfemia MCB, Saduvala N, Kline R, Prejean J, Kim D, Hall HI. Phylogenetic surveillance of HIV-1 non-B subtypes among U.S.-born and foreign-born individuals living in the United States revealed similar viral genetic diversity but different race and male

transmission patterns. Poster session presented at: 19th Conference on Retroviruses and Opportunistic Infections; 2012 Mar 5–8; Seattle, WA. Available at http://www.retroconference.org/2012b/PDFs/516.pdf.

APPENDIX A

History and Timeline of MHS

2002: CDC funds four surveillance areas to conduct the Antiretroviral Drug Resistance Testing (ARVDRT) pilot project, which demonstrates the feasibility of using remnant sera from diagnostic testing to conduct HIV genotypic resistance testing.

2004: Through a four-year cooperative agreement, 22 CDC-funded HIV Incidence Surveillance (HIS) areas opt to also conduct Variant, Atypical, and Resistant HIV Surveillance (VARHS) using a protocol developed under ARVDRT.

CDC determines that VARHS is a non-research disease surveillance activity that does not require institutional review board approval or informed consent.

2007: CDC publishes first VARHS data in abstract presented at the 2007 Conference on Retroviruses and Opportunistic Infections.

2008: Through a five-year cooperative agreement, CDC funds 11 surveillance areas to conduct VARHS using the protocol developed in 2004.

CDC establishes a contract with the Stanford Clinical Virology Laboratory at Stanford University Medical Center (Palo Alto, CA) to conduct genotypic resistance testing for VARHS.

2009: CDC adapts the WHO global HIV-1 surveillance mutation list developed for interpreting transmitted drug resistance-associated mutations (TDRMs) for all HIV-1 subtypes and finalizes the CDC HIV-1 surveillance mutation list for evaluating the prevalence of TDRMs for subtype B, the predominant HIV-1 subtype in the United States.

2010: CDC publishes first VARHS analysis in peer-reviewed journal. (*AIDS*. 2010 May 15; 24(8):1203-12.)

2011: CDC does not renew the contract with the Stanford Laboratory and discontinues support of genotypic resistance testing. VARHS transitions to a program reliant on VARHS-funded areas to collect nucleotide sequence data from commercial and private laboratories.

2012: CDC expands HIV surveillance to include molecular epidemiology approaches to describe the burden of HIV infection and supplement other HIV surveillance data. Molecular HIV Surveillance (MHS) reflects these broader goals and objectives.

MHS includes the collection of (1) ARV history use data for all new diagnoses of HIV infection and (2) nucleotide sequence data for all persons diagnosed with HIV infection, regardless of when the person was diagnosed or when the specimen (from which the sequence was obtained) was collected.

APPENDIX B

Glossary of Terms (unless otherwise noted, terms are limited to HIV)

Antiretroviral drug: a medication used to treat HIV infection and prevent HIV transmission.

Antiretroviral therapy: a combination of antiretroviral drugs administered consistently and therapeutically to treat HIV. It does not cure HIV infection, but controls HIV reproduction and transmission and boosts the immune system of the person in treatment.

Circulating recombinant form (CRF): an HIV strain with a mosaic structure of the genome consisting of two or more distinct subtypes that have been identified in at least three individuals who do not have direct, epidemiologically-linked infections. As of May 2012, 51 CRFs have been identified globally.

Drug resistance: the ability of HIV to continue to reproduce in the presence of antiretroviral drugs. Drug resistance can be acquired (induced) after exposure to antiretroviral drugs or transmitted from an HIV-infected person.

Drug resistance testing: testing the blood plasma of HIV-infected individuals to identify HIV drug resistance. Commonly used drug resistance tests include genotypic resistance testing (genotyping), phenotypic drug resistance testing (phenotyping), and virtual phenotypic drug resistance testing.

Enhanced HIV/AIDS Reporting System (eHARS): a browser-based HIV surveillance system deployed at state and local health departments. The data are collected in documents such as case reports, lab reports and death certificates. The health departments submit de-identified data electronically on a monthly basis to CDC's national database through a secure data network.

Electronic laboratory-based reporting (ELR): the electronic transfer of public health data from clinical laboratories to public health agencies in pre-established formats that do not require extensive human manipulation to add to a database. HL7 messaging is an example ELR format.

FASTA file: a standard, text-based format for nucleotide sequences. Each nucleotide sequence is preceded by a line starting with > and a description or name. Base pairs or amino acids are represented using single-letter codes.

> *Example:*
> >HumanATGGCACATGCAGCGCAAGTAGGTCTACAAGACGCTACTTCCCCTATCATAG
> AAGAGCTTATCACCTTTCATGATCACGCCCTCATAATCATTTTCCTTATCTGCTTCCTA

Genotypic resistance testing: testing the blood plasma of HIV-infected individuals to detect the presence of mutations associated with drug resistance. Genotypic resistance assays compare the nucleotide sequences (e.g., the protease and reverse transcriptase genes of the *pol* region) of the infected person with a wild-type strain.

Health Information Exchange (HIE): the mobilization of healthcare information electronically across organizations within a region, community or hospital system. HIE provides the capability to electronically move clinical information among disparate health care information systems while maintaining the

meaning of the information being exchanged. The goal of HIE is to facilitate access to and retrieval of clinical data to provide safer, more timely, efficient, effective, patient-centered care.

HL7 messaging: a standard method of electronically transmitting laboratory results.

Molecular epidemiology: the application of molecular biology techniques (e.g. genetic sequencing, phylogenetic analyses of viral sequences) for the detection, characterization, and transmission of HIV to study the distribution and determinants of disease occurrence and health-related events in the human population.

Mutation: a genetic change that results in a viral strain that is different from the wild-type HIV strain. Mutations can occur naturally or in the presence of antiretroviral drugs. A mutation is described by a combination of letters and numbers (e.g., M41L). The first letter (M) represents the amino acid in the wild-type strain, the number (41) represents the amino acid position in the gene, and the last letter (L) represents the mutation.

Nucleotides: molecules that make up the structural basis of nucleic acids, such as DNA or RNA. Each nucleotide consists of a phosphate group, a sugar (ribose in RNA), and a set of nucleotide bases: adenine (A), cytosine (C), guanine (G), and thymine (T). Three nucleotides make up a codon, which represent a single amino acid, the building blocks of proteins.

Nucleotide sequence: the genetic code of nucleic acids (i.e., DNA and RNA).

Phylogenetic analysis: the process of studying the relationship between HIV strains through analysis of nucleotide sequences. Phylogenetic methods are used to detect closely related HIV strains (i.e., clusters) and graphically display them through phylogenetic trees.

Polymorphism: a genetic mutation that occurs naturally in the absence of antiretroviral drugs.

Recombinant form: a hybrid HIV strain created when two or more HIV strains of different subtypes are combined. See circulating recombinant form and unique recombinant form.

Secure data network (SDN): method of transmitting data across defined, secure boundaries.

Secure file transfer protocol (SFTP): a type of protocol that provides a set of rules that govern the syntax, semantics, and synchronization of communication across computer networks. SFTP encrypts both commands and data, requires certification on the sending or receiving end, and allows the secure transmission of passwords and sensitive information.

Transmitted HIV drug resistance: the transmission of a drug-resistant HIV strain from an infected person to an uninfected person, resulting in HIV drug resistance in the newly infected, drug-naïve person. The presence of mutations associated with transmitted drug resistance (i.e., transmitted drug-resistance–associated mutations, TDRMs) can be detected through genotypic resistance testing.

Transmitted drug resistance-associated mutations: HIV mutations that confer antiretroviral drug resistance.

Unique recombinant form (URF): a hybrid strain that is the result of recombination of two or more HIV subtypes that has not been identified elsewhere.

Viral load: an estimate of the amount of virus in an infected person's blood. For clinical HIV management, it is typically expressed as the number of HIV RNA copies calculated per milliliter of blood plasma.

Wild-type strain: an HIV strain that has not been exposed to antiretroviral drugs.

APPENDIX C

Acronyms

ACRF	Adult HIV Confidential Case Report Form
AES	Advanced Encryption Standard
AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral therapy
ARV	Antiretroviral
ARVDRT	Antiretroviral Drug Resistance Testing
ASCII	American Standard Code for Information Interchange
CDC	Centers for Disease Control and Prevention
CD4	Cluster of differentiation 4
CLIA	Clinical Laboratory Improvement Amendments
CRF	Circulating Recombinant Form
CROI	Conference on Retroviruses and Opportunistic Infections
DFS	Dried fluid spots
DHAP	Division of HIV/AIDS Prevention
eHARS	Enhanced HIV/AIDS Reporting System
ELR	Enhancing Laboratory Reporting, Electronic laboratory reporting
HICSB	HIV Incidence and Case Surveillance Branch
HIS	HIV Incidence Surveillance
HIV	Human Immunodeficiency Virus
IAS-USA	International Antiviral Society-USA
INSTI	Integrase strand-transfer inhibitor
IRB	Institutional Review Board
LOINC	Logical Observation Identifiers Names and Codes
MHS	Molecular HIV Surveillance
NCHHSTP	National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention
NEDSS	National Electronic Disease Surveillance System
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside reverse transcriptase inhibitor
ORP	Overall responsible party
PHIN	Public Health Information Network
PI	Protease inhibitor
S&C	Security and confidentiality
SDN	Secure data network
SFTP	Secure file transfer protocol
TDRM	Transmitted drug resistance-associated mutation
TTH	Testing and treatment history
URF	Unique recombinant form
VARHS	Variant, Atypical, Resistant HIV Surveillance
VL	Viral load
VPN	Virtual private network

The *Molecular HIV Surveillance* chapter of the *Technical Guidance for HIV Surveillance Programs* was revised and reviewed by staff within the CDC HIV and Incidence Case Surveillance Branch and the following health departments: Rory Angulo (Connecticut), Mary-Grace Brandt (Michigan), Mariama Gondo (Florida), Dan Gordon (New York), Tom Jaenicke (Washington), Michelle Porter (Texas), and Lucia Torian (New York City).