**National HIV Surveillance System (NHSS)**

Attachment 4c.

Duplicate Review Technical Guidance

# Technical Guidance for HIV Surveillance Programs

## Duplicate Review

HIV Incidence and Case Surveillance Branch
Atlanta, Georgia

# Contents

# Background

HIV surveillance systems must provide a reliable measure of the number of persons in need of HIV prevention and care services at the local, state and national levels. An accurate HIV surveillance system is one that minimizes the degree to which it overcounts or undercounts reported cases of HIV infection and maximizes the reliability with which data for a given person are linked over time. Failing to properly link an incoming surveillance report to an existing case leads to overcounting and incomplete case information; incorrectly linking an incoming surveillance report to an existing case may lead to undercounting and data contamination.

Because doctors, hospitals, laboratories, and other reporting entities may be required to report all diagnoses of HIV infection, duplicate case reports within a state (intrastate) or between states (interstate) may not be identified during routine case entry into the surveillance database. To prevent overcounting and undercounting of cases, identification of potential intrastate and interstate duplicate case reports, merging case reports that have been deemed to be duplicates at all levels, and providing duplicate review resolution to CDC (i.e., Same as or Different than) must be carried out on a regular basis.

Within a state, surveillance software and routine surveillance practices are used to identify and eliminate duplicate case reports. These processes can use personally identifiable information (PII) and other useful information maintained at a state or local level. At the national level, CDC does not receive PII (e.g., name, Social Security Number) so duplicate case reports cannot be identified with the same degree of accuracy. Thus, CDC requires all surveillance areas to perform both intrastate and interstate review and de-duplication on a routine basis and ensure that each person in the surveillance database is given one unique state-assigned case number (stateno).

# Intrastate Duplicate Review

The prerequisites (structural requirements), best practices (process standards), and outcome standards for intrastate duplicate review are described next, followed by more in-depth guidance on specific topics.

## Structural Requirements

1. Case, laboratory, and other reports received on a person
2. HIV Surveillance System Software, eHARS

3. eHARS Data Entry Quick Reference Guide[1]
4. eHARS Technical Reference Guide[2]
5. Data processing policies, procedures and tools for record linkage (see file Record Linkage)
6. Procedures for evaluating accuracy of HIV surveillance systems (see file Evaluation and Data Quality)
7. Variables to ascertain potential intrastate duplicate case reports:

    - The eHARS report "Intrastate Duplicate Cases Based on CDC Matching String" is available under Operational in the eHARS REPORT module index. The report identifies and generates a list of potential duplicate case reports within a jurisdiction's eHARS using CDC matching strings. Cases are first matched using the following Person View variables: last name soundex (last_name_sndx), date of birth (dob), sex at birth (birth_sex) and state of residence at HIV diagnosis (rsh_state_cd). Country of residence at HIV diagnosis (rsh_country_cd) is used if rsh_state_cd is 'FC – Foreign Country'. If no match is found then cases are matched on last name soundex, date of birth, sex at birth and state or country of residence at AIDS diagnosis (rsa_state_cd / rsa_country_cd). Cases that match on the CDC string but have previously been confirmed by the jurisdiction as different persons are excluded from the list.

    - In addition to running the above eHARS report, jurisdictions are encouraged to perform more in-depth duplicate reviews using information that are readily available at the state level, e.g., first name (first_name), last name (last_name), middle name (middle_name), first and last name soundex (first_name_sndx, last_name_sndx), date of birth (dob), sex at birth (birth_sex), race/ethnicity (race), full Social Security Number (ssn), death date (dod). When these values are identical, other variables may be used to determine if the cases are duplicates. Examples of such variables include: medical record number (medrecno); inmate identification number (prisno); date of diagnosis of HIV infection (hiv_dx_dt); and date of diagnosis of stage 3 HIV infection (aids_dxx_dt).

## Process Standards

### 1. Frequency of Procedure

- Monthly run eHARS canned report Intrastate Duplicate Cases Based on CDC Matching String, merge duplicate case reports and update eHARS status.

---

[1] All health department HIV Surveillance personnel who are United States citizens are eligible to access the HIV Incidence and Case Surveillance Branch (HICSB) workspace on CDC SharePoint at https://partner.cdc.gov/sites/NCHHSTP/HICSB/default.aspx. If you have questions or problems with access, please contact your assigned CDC epidemiologist through the HIV Incidence and Case Surveillance Branch's main number (404) 639-2050.

[2] See footnote 1, immediately above. eHARS technical documentations available at SharePoint.

- Jurisdictions should also perform more in-depth intrastate duplicate review using exact and fuzzy (i.e., inexact) matching methods.

**2. Records that Represent the Same Person**

- Case reports that have been confirmed to be duplicates should be merged. When merging, retain the STATENO belonging to the case that was first entered into eHARS (the case with the earlier Person View enter_dt).

- eHARS contains a Transfer Document feature which can be found under Document and Case Maintenance in the ADMIN module index. Transfer Document allows the user to merge duplicate case reports by entering the appropriate state and state-assigned case number (stateno), eHARS unique identifier (ehars_uid) or document unique identifier (document_uid) of the source case (i.e., the case with the later Person View enter_dt), and the appropriate state and state-assigned case number (stateno) or eHARS unique identifier (ehars_uid) of the target case (i.e., the case with the earlier Person View enter_dt).

- The Adult Case Report Form (ACRF) and the Pediatric Case Report Form (PCRF) documents in eHARS contain a Duplicate Review tab that allows the user to enter duplicate status information regarding two reported cases of HIV infection. Jurisdictions may utilize the Duplicate Review tab to maintain a log of cases (e.g., STATENOs) that have been merged with another case within the jurisdiction's eHARS. To do this, the surveillance staff needs to enter an ACRF or PCRF document for the target case and, under the Duplicate Review tab, select duplicate status as '1 – Same as', select the jurisdiction's name for site and enter the STATENO of the source case as the state ID number.

**3. Records that Represent Different Persons**

- When a pair of case reports in the "Intrastate Duplicate Cases Based on CDC Matching String" report has been determined to represent two different persons, the jurisdiction should notify CDC by entering an ACRF or PCRF document into eHARS for at least one of the cases and updating the duplicate status under the Duplicate Review tab to "2 – Different than" and entering the jurisdiction's name for site and the STATENO of the other case in the pair as the state ID number.

### Outcome Standard

- ≤1% of cases for a report year have duplicate case reports, assessed 12 months after the report year. Duplication rates should be calculated using methods shown in the file Evaluation and Data Quality.

## Interstate Duplicate Review

The same HIV-infected person may be reported multiple times to public health departments in different states. Interstate duplicate case reports can result from persons moving or receiving care in different states over time and being reported to multiple state health departments in

accordance with local reporting requirements. Interstate duplicate review is designed to ensure that a case of HIV infection is counted only once in the National HIV Surveillance System (NHSS). The potential for duplicate reporting in the NHSS may increase as persons with HIV infection remain healthier longer due to advances in the clinical treatment of HIV infection and increased laboratory-driven surveillance. Therefore, routine interstate case de-duplication activities are critical to ensure accurate case counts at the national level. In 1986 and 2001, respectively, the Council of State and Territorial Epidemiologists (CSTE) passed resolutions for state-to-state reciprocal notification processes for AIDS and HIV case reporting to encourage resolution of duplicate case counting (See http://c.ymcdn.com/sites/www.cste.org/resource/resmgr/PS/1986-17.pdf and http://c.ymcdn.com/sites/www.cste.org/resource/resmgr/PS1/2001-ID-04.pdf). HIV surveillance program staff should communicate with other states to resolve potential duplicates using guidance outlined below in accordance with CSTE position statements and detailed procedural guidance disseminated by CDC.

Potential interstate duplicate case reports may be identified in three ways. Before entering a new case into eHARS, surveillance staff may contact the CDC Division of HIV/AIDS Prevention (DHAP) Helpdesk (1-877-659-7725) to determine if the case has been reported by another jurisdiction. (Using the Secure Online Soundex Match application, state and local health departments are able to search for potential duplicates instead of calling the Helpdesk.) There are three potential outcomes:

1. The case has been reported by another jurisdiction. In this situation, surveillance staff should still enter the case into eHARS and ensure that data elements in the Duplicate Review tab are appropriately populated (i.e., select '1 – Same as' for 'duplicate status', the name of the other jurisdiction for 'site', and the other jurisdiction's STATENO for the case for 'state ID number')

2. The case has not been reported by another jurisdiction, but the person's last name soundex, date of birth and sex at birth match those of a case reported by another jurisdiction. In this situation, when entering the case into eHARS, surveillance staff should also ensure that data elements in the Duplicate Review tab are populated (i.e., select '2 – Different than' for 'duplicate status', the name of the other jurisdiction for 'site', and the other jurisdiction's STATENO for the case for 'state ID number')

3. The case has not been reported by another jurisdiction and no match on last name soundex, date of birth and sex at birth are found by the CDC DHAP Helpdesk, then data elements in the Duplicate Review tab should be left blank when entering the case into eHARS.

The second way that potential interstate duplicate case reports are identified is through duplicate review reports that CDC distributes to local and state health departments; the semi-annual Routine Interstate Duplicate Review (RIDR) reports and the Cumulative Interstate Duplicate Review (CIDR) report that was distributed in January 2018. RIDR/CIDR reports are generated after data transmitted to CDC by local and state health departments have been consolidated. It is highly encouraged that jurisdictions proactively contact the CDC DHAP Helpdesk for more timely identification of potential duplicates to help reduce the number of potential interstate duplicate pairs in their semi-annual RIDR reports.

The third way that potential interstate duplicate case reports can be identified is through the use of a secure data sharing tool. Through grant PS 18-1805 Georgetown University is funded to provide a secure data sharing tool with matching algorithm to all 59-funded state and local health departments. The secure data sharing tool will assess case pairs using information available at the local level that is not available at the national level (e.g., Social Security Number, last name, etc.), and will generate a report indicating the matching level for each potential duplicate (e.g., exact, extremely high, etc.). Therefore, it has the ability to more efficiently identify "exact" matches compared to standard RIDR/CIDR methods and may also find matches not detected through RIDR/CIDR. However, accuracy of the matches should be determined before entering the information into eHARS. Accuracy can be determined by selecting a subset of matches at various matching levels and discussing them further with the other jurisdictions to determine if they are true matches. This will establish a threshold where matches can be assumed to be true matches. For details on Georgetown's secure data sharing tool and requirements for participation, please contact Georgetown University, PS18_1805@georgetown.edu.

## National Data Processing and RIDR/CIDR Report Generation

To prevent overcounting of cases at the national level, on a quarterly basis, CDC de-duplicates the national HIV surveillance database as part of National Data Processing. The de-duplication process involves 1) identifying duplicate case reports and 2) combining duplicate case reports and selecting one report state code (report_state_cd) and the corresponding STATENO (stateno) for the case. Duplicate case reports are identified using the CDC match string as well as the eHARS duplicate review data. Cases are first linked by last name soundex (last_name_sndx), date of birth (dob), sex at birth (birth_sex), and state of residence at HIV diagnosis (rsh_state_cd). Country of residence at HIV diagnosis (rsh_country_cd) is used if rsh_state_cd is 'FC – Foreign Country'. If no match is found, the process substitutes state and country of residence at stage 3 HIV infection diagnosis (rsa_state_cd / rsa_country_cd) for state and country of residence at HIV diagnosis. Moreover, case reports are regarded as duplicates if they do not agree on the CDC match string but one or more jurisdiction's duplicate review data indicate that the cases are the "1 – Same as"; case reports are regarded as for different persons if they match on the CDC match string but one or more jurisdiction's duplicate review data indicate that the cases are "2 – Different than".

RIDR reports are generated using data from the eHARS consolidated database on a semiannual basis. The list is generated by identifying cases that match on last name soundex (last_name_sndx), date of birth (dob) and sex at birth (birth_sex), but have not been confirmed as the same or different persons by the local and state health departments. These potential interstate duplicate case reports are distributed to local and state health departments for resolution. In RIDR reports, at least one case in the duplicate had to be reported during the six months prior to the generation of the report. In order to resolve older duplicates in the national dataset, in January 2018, CDC also generated the CIDR report. CIDR reports contain all unresolved duplicates regardless of when they were reported to CDC.

The prerequisites (structural requirements), best practices (process standards), and outcome standard for interstate duplicate review are described next.

## Structural Requirements

1. Link to CSTE 1986-17 AIDS Case Reporting: Reciprocal Notification (http://c.ymcdn.com/sites/www.cste.org/resource/resmgr/PS/1986-17.pdf).
2. Link to CSTE 2001-ID-04 Reciprocal (Interstate) Notification of HIV Cases (http://c.ymcdn.com/sites/www.cste.org/resource/resmgr/PS1/2001-ID-04.pdf).
3. HIV Surveillance System Software, eHARS.
4. Variables used for CDC matching string last name soundex (last_name_sndx), date of birth (dob), sex at birth (birth_sex), and state of residence at diagnosis (rsh_state_cd or rsa_state_cd) or, if a non-US resident at time of diagnosis, country of residence at diagnosis (rsh_country_cd / rsa_country_cd).
5. Standard procedure for processing of CDC's Cumulative and Routine Interstate Duplicate Review reports (see SharePoint/Case Surveillance/RIDR/Instructions for Processing CDC's Duplicate Review Report_YYYYMM).
6. Case Residency Assignment Policies and Procedures (see file Date and Place of Residence).
7. Procedures for evaluating accuracy of integrated HIV surveillance systems (see file Evaluation and Data Quality).
8. Access to Secure Access Management Services; Current Digital Certificate.

## Process Standards

1. **Frequency of Procedure**

   Routine Interstate Duplicate Review must be performed semiannually. Cumulative Interstate Duplicate Review must be completed over the course of the PS 18-1802 funding cycle (2018-2022).

2. **Duplicate Review of Out-of-Jurisdiction Cases**

   States should maintain information on out-of-jurisdiction cases in eHARS. To determine if a pair in the RIDR/CIDR report represents the same person or different persons, contact the other state's surveillance coordinator (or his or her designees) to compare and collect additional information. Questions to ask to determine if pairs are the same or different persons might include:

   - Do the cases share the same name, including considerations of other available name types (e.g., alias)?
   - Does the Social Security Number prefix come from the other state?
   - Are there any comments that reference the other state?
   - Is there a death date match?

- Is there a current residence match?
- Is there an unusual mode of exposure?

3. **Records that Represent the Same Person**

   If, after discussion with the other state's surveillance coordinator (or his or her designees), the cases are deemed to represent the same person, case residency at diagnosis must be established for the pair. Use policies and procedures for state of residence at diagnosis to ensure that cases are counted appropriately (see file Date and Place of Residence). Once state of residence is established, jurisdictions should inform CDC of the duplicate review resolution by updating the data elements in the Duplicate Review tab of the ACRF or PCRF document (i.e., select "1-same as" for duplicate status, etc.) as well as the residency at diagnosis information (i.e., rsh_state_cd / rsh_country_cd and rsa_state_cd / rsa_country_cd [if the person has a diagnosis of stage 3 HIV infection]).

   In addition to updating the residence at diagnosis and information on the Duplicate Review tab, jurisdictions are encouraged to share with each other additional information about the case in accordance with their respective reporting and data sharing laws and regulations. Such information may include risk factors, AIDS-defining conditions, vital status, date of death, last negative test result, if nucleotide sequences are available, care status etc. In particular, surveillance staff should help each other to determine in which jurisdiction does the patient currently reside and enter the address information into the Identification tab of the ACRF or PCRF document in eHARS.

4. **Records that Represent Different Persons**

   If, after discussion with the other state's surveillance coordinator (or his or her designees), the cases are deemed to represent different persons, jurisdictions should inform CDC of the duplicate review resolution by updating the data elements in the Duplicate Review tab of the ACRF or PCRF document (i.e., select "2 – Different than" for duplicate status, etc.).

5. **Resolution of Potential Duplicates**

   100% of potential interstate duplicate pairs in the RIDR/CIDR reports should be resolved and duplicate status updated in eHARS in the following timeframes:

   - RIDR report released in January should be completed by June of the same year.
   - RIDR report released in July should be completed by December of the same year.
   - CIDR report released in January 2018 should be completed by December 2022, with at least 20% of duplicates being resolved each year.

Staff approved to release information about HIV cases to other jurisdictions can be found on the CSTE HIV/AIDS Contact Board Web site. Please contact the HIV surveillance support staff at CSTE for information on obtaining sign-on identifications and passwords to access the web site (http://www.cste.org/?page=HIVContact); the CSTE point of contact can be reached at 770-458-3811.

Contact the CDC's designated subject matter expert (SME) for RIDR.CIDR for any questions related to the RIDR/CIDR process. The RIDR/CIDR SME may be reached at the CDC HIV Incidence and Case Surveillance Branch's main number (404-639-2050) or through the CDC epidemiologist assigned to your jurisdiction for technical assistance support.

## Outcome Standard

- ≤2% of Routine Interstate Duplicate Review (RIDR) pairs remain unresolved at the end of each six month RIDR cycle, assessed at the end of each cycle.

- A minimum of 20% of case pairs on the CIDR list are to be resolved by the December data transfer each year of the funding cycle (2018-2022). At the end of PS18-1802 (December 2022), 100% of case pairs on the CIDR list should be investigated and resolved.