Without the availability of a unified health record, there are often gaps between clinical research data abstracted from medical records and surveillance data collected on HIV-infected persons.

Further, the completeness of surveillance data is often limited by the complexity of reporting, underreporting, timeliness, and completeness of data.

The DC Cohort study (DCC), an observational longitudinal cohort study of HIV-infected persons in care at 13 clinical sites, conducts routine linkages to the DC Department of Health (DOH) surveillance databases.

We sought to assess differences between data collected through the DCC and DOH data.

Most DCC participants (n=6,166; 88.3%) were male, 217 previously undiagnosed STD diagnoses were captured post linkage.

For purposes of this study, we defined linkage as matches between DCC and DOH data that were confirmed to be correct by a human reviewer.

To address discrepancies between DCC and DOH data, final assignment of linkage was made by humans.

Most DCC participants (n=6,166; 88.3%) matched to the DC DOH surveillance database; of whom 143 (2.3%) participants were enrolled at >1 DCC site (Figure 2).

Participants who did not match (12.3%) were either non DC residents or likely not reported to the DC DOH (Figure 2).

Post linkage, an additional 15,201 and 32,852 CD4 and VL lab reports were added to the DCC and DC DOH databases, respectively (Figure 2).

Significant differences were observed with respect to mode of transmission, availability of HIV diagnosis date, and AIDS diagnoses across DCC, DC DOH, and post linkage data (all p<0.001) (Table 1).

Median duration of HIV infection increased from 6.5 to 8.0 years post linkage (p<0.001) (Table 1).

217 previously undiagnosed STD diagnoses were captured post linkage (Table 1).

32% of participants had labs from another site, suggestive of receipt of HIV care from more than one clinical site (Figure 3).

The proportion of persons retained in care and on ART improved post linkage (Figure 4).

Most DCC participants (n=6,166; 88.3%) matched to the DC DOH surveillance database; of whom 143 (2.3%) participants were enrolled at >1 DCC site (Figure 2).

Participants who did not match (12.3%) were either non DC residents or likely not reported to the DC DOH (Figure 2).

Post linkage, an additional 15,201 and 32,852 CD4 and VL lab reports were added to the DCC and DC DOH databases, respectively (Figure 2).

Significant differences were observed with respect to mode of transmission, availability of HIV diagnosis date, and AIDS diagnoses across DCC, DC DOH, and post linkage data (all p<0.001) (Table 1).

Median duration of HIV infection increased from 6.5 to 8.0 years post linkage (p<0.001) (Table 1).

217 previously undiagnosed STD diagnoses were captured post linkage (Table 1).

32% of participants had labs from another site, suggestive of receipt of HIV care from more than one clinical site (Figure 3).

The proportion of persons retained in care and on ART improved post linkage (Figure 4).

### Results

#### Methods

**Linkage Methods**

- DCC participant data was sent to the DC DOH from participating sites and overlapping variables of interest were sent from the DCC Data and Statistics Coordinating Center (DSCC) (Figure 1).

- DCC data for participants enrolled between January 1, 2011 and June 15, 2015 were electronically linked to case surveillance data reported to the DC DOH using an 11-key algorithm based on patient first and last name, date of birth, and SSN.

- Surveillance laboratory data (i.e., CD4 and viral load (VL)) were linked to DCC data based on the sample test date, result, and type.

- To address discrepancies between DCC and DOH data, final assignment of case demographic and risk characteristics was based on a hierarchical decision matrix.

**Analysis Methods**

- Descriptive statistics characterizing the analytic population pre- and post-linkage by demographic and HIV clinical characteristics were computed.

- Differences in pre- and post linkage data were assessed using Chi-square and Wilcoxon tests.

- Sources of laboratory reporting and HHS continuum of care outcomes were measured.

---

### Introduction

Increasing the Completeness of HIV Public Health Surveillance and Clinical Research Databases: Linkage of District of Columbia Surveillance Data with DC Cohort Study Data-Washington, DC

Amanda D. Castel1, Jenevieve Opoku2, Lindsey Powers Hap3, Michael Kharfen2, Alan E. Greenberg1, Naji Younes4, on Behalf of the DC Cohort Executive Committee

1Georgetown University Milken Institute School of Public Health, Washington, DC; 2District of Columbia Department of Health

---

### Results (cont’d)

- Linkage of public health surveillance data with clinical cohort data was mutually beneficial and improved the completeness of both systems.

- For the clinical cohort, linkage assisted in identifying study co-enrollment, more accurate assessment of care patterns, improved accuracy of HIV diagnosis dates, and previously missed STD and OI diagnoses.

- For surveillance, linkage allowed for improved accuracy with respect to risk factor ascertainment, reportable laboratory indicators, and resulted in improved viral suppression rates.

- Routine linkages such as this are useful to researchers and public health surveillance programs and can assist with more accurately characterizing co-morbidities and HIV continuum of care patterns among HIV-infected persons at the clinical and population-level.