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. Castel[1], Jenevieve Opoku[2] , Lindsey Powers Happ[1], Michael Kharfen[2], Alan E. Greenberg[1], Naji Younes[1], on Behalf of the DC Cohort Executive Committee

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

Introduction

- 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 DC Cohort study and routinely collected DC DOH HIV/AIDS surveillance data.

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.

Results

Figure 1. Linkage Algorithm for DC Cohort and DC DOH Data

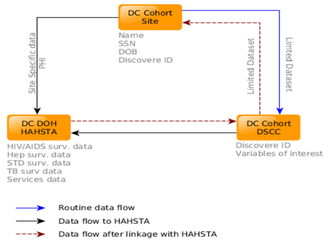


Figure 2. Results of Linkage of DC Cohort and DC DOH Surveillance Data (N= 7,031)

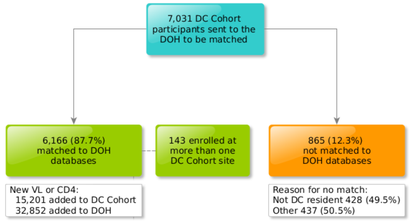


Table 1. Characteristics of DC Cohort Participants Matching to the DC DOH Database, N= 6,166

Characteristic	DCC Pre-Linkage	DOH Pre-Linkage	Post- Linkage	p-value (DCC vs. post linkage)	p-value (DOH vs. post linkage)
Demographics	N (%)	N (%)	N (%)		
Race/ethnicity				0.998	0.141
Black non-Hispanic	4,620 (78.8%)	4,690 (78.2%)	4,624 (78.9%)		
White non-Hispanic	846 (14.4%)	848 (14.1%)	845 (14.4%)		
Other	394 (6.7%)	458 (7.6%)	393 (6.7%)		
DC resident	4,646 (76.8%)	4,623 (78.1%)	4,646 (76.8%)	1.00	0.092
Known risk factor for HIV transmission	4,824 (78.2%)	5,260 (85.3%)	5,864 (95.1%)	*<0.001*	*<0.001*
Mode of transmission				*<0.001*	*<0.001*
MSM/IDU	78 (1.6)	213 (4.0%)	100 (1.7%)		
MSM	2,257 (46.8%)	2,403 (45.7%)	2,951 (50.3%)		
Heterosexual	1,790 (37.1%)	1,654 (31.4%)	1,723 (29.4%)		
Perinatal	198 (4.1%)	140 (2.7%)	191 (3.3%)		
Other	501 (10.4%)	850 (16.2%)	899 (15.3%)		
Clinical indicators					
HIV diagnosis date available	6,045 (98.0%)	5,971 (96.8%)	6,047 (98.1%)	0.896	*<0.001*
Median duration of HIV infection (IQR)	6.5 (12.0, 19.2)	6.5 (10.8, 16.7)	8.0 (13.5, 20.0)	*<0.001*	*<0.001*
AIDS diagnosis	2,515 (40.8%)	3,537 (57.4%)	3,808 (61.8%)	*<0.001*	*<0.001*
OI at AIDS diagnosis	508 (22.0%)	1,034 (29.2%)	1,045 (28.1%)	*<0.001*	0.286
Individuals with ≥1 STD diagnosis (GC, chlamydia, syphilis) on or after enrollment	431 (7.0%)	419 (6.8%)	710 (11.5%)	*<0.001*	*<0.001*
Ever virally suppressed (VL<200 copies/ml at or after enrollment)	4,663 (75.6%)	2,587 (42.0%)	4,978 (80.7%)	*<0.001*	*<0.001*

P-values in italics are statistically significant at p<0.05.

Figure 3. Sources of HIV Laboratory Data among DC Cohort Participants

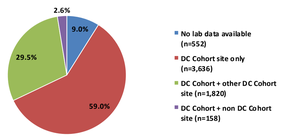
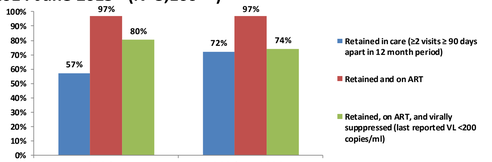


Figure 4. HIV Care Continuum Measures by Data Source, June 2014-June 2015* (N=5,160)**



*Each step in the care continuum is dependent on the prior step. "On ART" defined as ever virally suppressed for the DOH data .
** Excludes participants who died, withdrew, or had less than 1 year of follow up by June 2015.

Results (cont'd)

- Most DCC participants (n=6,166; 88%) 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).

Conclusions

- 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.

Acknowledgements: This work was supported by the National Institute of Allergy and Infectious Diseases at the National Institutes of Health [UO1 AI69503-03S2]. DC Cohort Executive Committee Members: DC Cohort Data in this analysis were collected by the DC Cohort investigators and research staff located at: Cerner Corporation (Jeffrey Binkley, Harlen Hays, Thilakvathy Subramanian, Rachel Hart, Dana Bryant); Children's National Medical Center Adolescent (Lawrence D'Angelo) and Pediatric (Natella Rakhmanina) clinics; Family and Medical Counseling Service (Angela Wood); Georgetown University (Princy Kumar, Mary Young); George Washington Medical Faculty Associates (David Parenti); George Washington University Department of Epidemiology and Biostatistics (James Peterson, Lindsey Powers Happ, Maria Jaurretche, Arpi Terzian); Howard University Adult (Maya Bryant, Saumil Doshi) and Pediatric (Sohail Rana) clinics; La Clinica Del Pueblo (Ricardo Fernandez); MetroHealth (Annick Hebou); National Institutes of Health (Carl Dieffenbach, Henry Masur); Unity Health Care (Stephen Abbott); Veterans Affairs Medical Center (Debra Benator); Washington Hospital Center (Maria Elena Ruiz); Whitman-Walker Health (Rick Elion, Deborah Goldstein). We would also like to acknowledge the Research Assistants at all of the participating sites, the DC Cohort Community Advisory Board, and staff of the DC Department of Health HIV/AIDS, Hepatitis, STD, TB Administration Strategic Information Division.