Integrating data systems to improve HIV care engagement in King County, WA

Tigran Avoundjian

A dissertation submitted in partial

fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Julia C. Dombrowski, Chair

Matthew R. Golden

Brandon L. Guthrie

Janet G. Baseman

Program Authorized to Offer Degree:

Public Health - Epidemiology

University of Washington

Abstract

Integrating data systems to improve HIV care engagement in King County, WA

Tigran Avoundjian

Chair of the Supervisory Committee:

Dr. Julia C. Dombrowski

Associate Professor, Medicine

Adjunct Professor, Epidemiology

Continuous engagement in HIV care and treatment is crucial for the health of persons living with HIV (PLWH) and for preventing HIV transmission to others. However, in the United States (US), care engagement, or retention in care, represents the biggest drop off in the HIV care continuum, which maps out the care process from HIV testing and diagnosis, linkage to and retention in HIV care, and ultimately achievement of viral suppression. Many health departments in the US use HIV surveillance data to facilitate HIV care engagement activities, a process known as data to care. While Data to Care programs have had some success, their effectiveness is hindered by the completeness and timeliness of HIV surveillance data. A novel approach to Data to Care uses real-time data exchange between HIV surveillance with external data sources, such as emergency department (ED) and inpatient (IP) hospitalization data and jail booking rosters, to improve the signal of Data to Care investigations, and provide a setting and an opportunity to re-engage PLWH in HIV care. Since real-time data exchange involves

linking data sources that don’t often have a shared unique person identifier, these programs should also consider the accuracy of the record linkage algorithms they utilize, in order to maximize their reach and efficiency.

We investigated the effect of the use of real-time data exchange on HIV care engagement outcomes in two settings: emergency department and inpatient hospitals and in jails. First, we evaluated the impact of an existing ED/hospital-based health information exchange on HIV care outcomes. We compared the proportion of patients that had a viral load test in the 3 months and viral suppression in the 6 months after an alert-eligible ED visit/inpatient admission in the pre-intervention (01/20/13-01/20/15) and post-intervention (07/20/15-07/20/17) periods. To assess whether our pre/post results could be due to secular trends, we compared the difference between patients with an alert-eligible ED visit/IP admission to patients who had a visit outside of the alert window in both the pre-intervention and post-intervention periods.

Next, we developed a new automated, real-time data exchange between public health HIV surveillance and county jail data to identify incarcerated PLWH and facilitate post-incarceration HIV care engagement efforts. A team of public health relinkage specialists and jail release planners used this data exchange to guide case conferences about patients who were virally unsuppressed or out-of-care and jointly developed a plan for re-engagement in care and treatment. We compared viral load testing within 3 months and viral suppression within 6 months after release from jail among PLWH released in the post-intervention period (04/01/18-11/01/18) to those released in the pre-intervention period (10/01/16-10/01/17) using Cox proportional hazards models.

Finally, we compared the performance of record linkage algorithms commonly used by data exchanges commonly used in public health practice. We compared five deterministic algorithms and two probabilistic record linkage algorithms using simulations and a real-world scenario. We simulated pairs of datasets while varying the number of erroneous fields per record and overlap between these datasets. We matched datasets using each algorithm and calculated their recall (sensitivity) and precision (positive predictive value). In a real-world scenario, HIV and STD surveillance data from King County, WA were matched to identify PLWH who had a syphilis diagnosis. We used manual review to define a gold standard and calculate recall and precision.

In our evaluation of an ED/hospital-based health information exchange, patients in the post-intervention period were 1.08 times more likely to have a viral load test within 3 months after an ED visit/IP admission (95% CI: 0.97, 1.20) and 1.50 times more likely to achieve viral suppression within 6 months after an ED visit/IP admission (95% CI: 1.27, 1.76). However, there was a similar pre/post increase in both HIV care engagement (DID: 1.00, 95% CI: 0.84, 1.18) and viral suppression (DID: 1.01, 95% CI: 0.84, 1.20) among patients with visits outside of the alert window.

After implementation of a real-time data exchange between HIV surveillance and jail booking data coupled with HIV care coordination between health department and jail release planners, viral load testing within 3 months after release from jail increased by 35% (95% CI: 0.84, 2.18) and viral suppression within 6 months after release from jail increased by 37% (95% CI: 0.82, 2.30), but these differences were not statistically significant.

In our simulation study, we found that probabilistic algorithms maintained a high recall at nearly all data quality levels, while being comparable to deterministic algorithms in terms of precision. Deterministic algorithms typically failed to identify matches in scenarios with low data quality. In the real-world scenario, probabilistic algorithms had the lowest trade-off between recall and precision.

The results of this dissertation indicate that ED/hospital-based data exchange provides substantial opportunities to interact with PLWH who are poorly engaged in HIV care. However, the observed increase in HIV re-engagement and viral suppression after implementation of this data exchange may reflect secular trends resulting from diverse interventions of which this program was only one. Real-time health information exchange with emergency departments and hospitals can identify PLWH who are inadequately engaged with care and facilitate D2C efforts, but more efforts are needed to improve the effectiveness of reengagement interventions linked to real-time D2C. Implementation of a real-time data exchange between HIV surveillance and jail booking rosters resulted in a trend towards improved post-incarceration HIV care outcomes for incarcerated PLWH who are virally unsuppressed/out-of-care in King County. Real-time data exchange between health departments and county jails is a promising strategy for identifying incarcerated PLWH to support care coordination and improving post-incarceration HIV care engagement. Finally, in our simulation study on record linkage algorithms, we found that probabilistic algorithms maximize the number of true matches identified, while still maintaining high precision. Public health activities that rely on the integration of multiple data sources to target intervention delivery should utilize probabilistic algorithms to reduce gaps in the coverage of interventions and maximize their reach.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

Continuous engagement in HIV care and treatment is crucial for the health of persons living with HIV (PLWH) and for preventing HIV transmission to others. Improving HIV care engagement and retention is a key area for improvement in the United States National HIV/AIDS strategy, and an integral part of HIV prevention activities for US health department HIV programs.[1,2] However, retention in care remains the biggest drop off in the HIV care continuum in the United States.[3] In 2015, although about 86% of PLWH in the United States had been diagnosed with HIV, only 49% were retained in HIV care, and 51% were virally suppressed. Improving retention in HIV care would reduce HIV-related mortality and HIV transmission.[4–6]

The effectiveness of interventions aimed at reducing barriers to care engagement has been limited by the lack of clear mechanisms to identify and target outreach to PLWH most in need of care re-engagement assistance. Many health departments use HIV surveillance data to identify out-of-care and virally unsuppressed PLWH and connect them with resources to improve their HIV care engagement, a process known as "Data to Care" (D2C).[7] Most health departments conducting D2C use surveillance data to generate lists of people who appear to be out of care and attempt to contact these persons.[8–11] To date, the effectiveness of D2C has been hindered by issues of data quality, completeness, and timeliness of HIV surveillance data.[12–16] Since surveillance-based lists of out-of-care PLWH are generated annually or semi-annually, these data cannot be used to detect abrupt changes in HIV care status due to important life events (e.g., incarceration). In addition, ascertainment of HIV care status using HIV surveillance data may be distorted by migration or laboratory reporting delays.[12,14] Furthermore, many patients may re-engage in HIV care before ever being contacted by the health department.[17]

One strategy to improve the efficiency and effectiveness of D2C programs is to integrate HIV surveillance data with external data sources, such as data from emergency department (ED) and inpatient hospitalizations and from jails. PLWH who are poorly engaged in HIV care or at high risk of falling out of care have high rates of ED utilization. ED utilization has also been shown to be associated with known risk factors of poor HIV care engagement, such as unstable housing and substance/alcohol use. The emergency department encounter and inpatient hospitalization may present an opportunity for health department staff conducting HIV care relinkage work to engage PLWH who they could not otherwise contact.

PLWH also have high rates of incarceration. About 1 in 7 PLWH pass through a correctional facility ever year, and rates of incarceration among PLWH is three to five times higher than the rate in the general population. Incarceration disrupts continuous HIV care engagement, and incarcerated PLWH experience lower rates of retention in care and viral suppression after release from prison or jail compared to the general HIV-positive population. People of color, transgender individuals, and men who have sex with men are disproportionately incarcerated and disproportionately affected by the HIV epidemic. Thus, incarceration may also compound disparities in HIV care delivery among these vulnerable populations. The delivery of existing post-incarceration HIV care interventions may be more challenging in the jail setting than in the prison setting because jail stays are shorter and jail volumes are higher. Because of the shorter length of stay, many PLWH incarcerated in jail may not disclose their HIV status, despite experiencing a disruption in HIV care. Real-time data exchange between HIV surveillance and jails could facilitate identification of individuals who are poorly engaged in HIV care before release from jail and ensure that all incarcerated PLWH have an HIV care plan after release from jail, which would prevent disruptions in continuous care engagement.

Integrating data systems to improve HIV care engagement requires linking data sources that typically do not have a shared unique person identifier. Thus, these types of data exchanges require the use of record linkage algorithms to approximately identify matching records for the same individual between data sources. The choice of record linkage algorithm could affect the efficiency and the reach of interventions. While deterministic algorithms, which use exact matching on specific variables or a set of matching rules to identify matched record pairs, typically have low false positive rates, they may miss a great deal of true matches, especially when data quality is poor.[18] In contrast, probabilistic algorithms, which use statistical methods to identify the optimal set of matches, identify a greater number of true matches, but may also have higher false positive rates. [19,20] HIV/STD programs tend to favor deterministic algorithms because of their typically low rates of false positive matches.[18,21,22] Since a major concern of working with HIV data is inadvertent disclosure of HIV status, minimizing false matches is crucial to preserving individual privacy. However, deterministic algorithms may be overly conservative, resulting in a large number of missed true matches. Missed matches represent missed opportunities to deliver public health interventions to individuals who need them, and depending on their distribution in the population, could magnify health inequities. Probabilistic algorithms could potentially offer increased sensitivity compared to deterministic algorithms, while still identifying a small number of false matches.

This dissertation evaluates the impact of two interventions that integrate external data sources with HIV surveillance data on HIV care outcomes in King County, WA. First, we evaluated the impact of an existing ED- and hospital-based health information exchange on HIV care engagement and viral suppression among virally unsuppressed PLWH seen at a UW Medicine ED/inpatient hospital (Chapter 2). Next, we developed a novel data exchange between two King County jails and Public Health Seattle King County's HIV surveillance data and evaluated its impact on post-incarceration HIV care re-engagement and viral suppression (Chapter 3). Finally, we conducted a simulation study to compare the recall, precision, and computational performance of seven matching algorithms commonly used to link large public health surveillance databases (Chapter 4). We conclude by summarizing the findings of these studies and discussing their implications (Chapter 5).

Chapter 2: Evaluation of an emergency department and hospital-based data exchange to improve HIV care engagement and viral suppression

Abstract:

Background: Emergency department (ED) visits and inpatient (IP) admissions may provide an opportunity to provide poorly engaged people living with HIV services to improve HIV care engagement and viral suppression. We developed a real-time data exchange between Public Health Seattle and King County (PHSKC) and three UW Medicine EDs and hospitals to identify virally unsuppressed PLWH and relink them to HIV care.

Objective: To evaluate the impact of the PHSKC-UW D2C program on HIV care re-engagement and viral suppression after an ED visit or inpatient admission.

Methods: We used a pre/post-intervention design to assess the impact of the intervention on care re-engagement in the 3 months and viral suppression in the 6 months after an eligible visit. Alerts were triggered for patients who had an ED visit or IP admission between 8AM and 6PM on weekdays and whose most recent viral load before the visit was ≥200 copies/mL. Care re-engagement was defined as ≥1 viral load result within 3 months and viral suppression as a viral load <200 within 6 months after an eligible visit. The post-intervention period included all alert-eligible encounters in the two years after the alert system was implemented. The pre-intervention period included all alert-eligible encounters in the 7-30 months prior to the implementation of the system, excluding the 6 months prior to implementation. To assess whether our pre/post results could be due to secular trends, we compared the difference between patients with a visit in the 8AM-6PM on weekdays window to patients with a visit outside of the alert window in both the pre-intervention and post-intervention period.

Results: There were 647 alert-eligible visits in the post-intervention period and 862 visits in the pre-intervention period, of which 30% were inpatient hospitalizations. Patients in the intervention period were 1.08 times more likely than patients in the pre-intervention period to have a viral load test within 3 months after an ED visit/IP admission (95% CI: 0.97, 1.20) and 1.50 times more likely to achieve viral suppression within 6 months after an ED visit/IP admission (95% CI: 1.27, 1.76). However, there was a similar pre/post increase in both HIV care engagement (DID: 1.00, 95% CI: 0.84, 1.18) and viral suppression (DID: 1.01, 95% CI: 0.84, 1.20) among pateints with visits outside of the alert window.

Conclusions: The PHSKC-UW D2C program was associated with improved HIV care re-engagement and viral suppression. However, our results may reflect secular trends resulting from diverse interventions, of which this program was only one. Real-time health information exchange with emergency departments and hospitals can identify PLWH who are inadequately engaged with care and facilitate D2C efforts, but more efforts are needed to improve the effectiveness of reengagement interventions linked to real-time D2C.

Introduction

Continuous engagement in HIV care and treatment is crucial for the health of persons living with HIV (PLWH) and for preventing HIV transmission to others. Since 2015, the World Health Organization and the United States National HIV/AIDS Strategy have emphasized the goals of improving HIV care engagement and treatment with antiretroviral therapy (ART) and adopted "treatment-as-prevention" as a key component of HIV prevention.[1,2] However, retention in care remains the biggest drop off in the HIV care continuum in the United States.[3] In 2015, although about 86% of PLWH in the United States had been diagnosed with HIV, only

49% were retained in HIV care, and 51% were virally suppressed. Improving retention in HIV care would reduce HIV-related mortality and HIV transmission.[4–6]

The effectiveness of interventions aimed at reducing barriers to care engagement has been limited by the lack of clear mechanisms to identify and target outreach to PLWH most in need of care re-engagement assistance. Since 2014, the Centers of Disease Control and Prevention (CDC) has encouraged state and local health departments to implement surveillance-based "Data to Care" (D2C) programs,[7] which use HIV surveillance data to identify out-of-care and virally unsuppressed PLWH and connect them with resources to improve their HIV care engagement. Most health departments conducting D2C use surveillance data to generate lists of people who appear to be out of care and attempt to contact these persons.[8–11] To date, the effectiveness of D2C has been hindered by issues of data quality, completeness, and timeliness of HIV surveillance data.[12–16] Since surveillance-based lists of out-of-care PLWH are generated annually or semi-annually, these data cannot be used to detect abrupt changes in HIV care status due to important life events (e.g., incarceration). In addition, ascertainment of HIV care status using HIV surveillance data may be distorted by migration or laboratory reporting delays. In King County, Washington 47% of PLWH who had one or more 12-month gap in laboratory reporting from 2006-2010 had moved out of the area, while only 38% had remained in King County.[12] In a similar investigation of "out-of-care" cases involving six US health departments (Washington, Wyoming, Alaska, Montana, Idaho, and Oregon), between 43% and 90% of PLWH with no CD4/viral load test reported for more than 12 months during 2012-2014 in each state had an explanation other than care disengagement for their gap in

CD4/viral load reporting.[14] Furthermore, many patients may re-engage in HIV care before ever being contacted by the health department.[17]

Integrating emergency department (ED) records with HIV surveillance may be a promising strategy to prioritize D2C investigations to serve vulnerable populations that are most in need of relinkage assistance.[23] PLWH who are poorly engaged in HIV care or at high risk of falling out of care have high rates of ED utilization.[24–26] ED utilization has been shown to be associated with known risk factors of poor HIV care engagement, such as unstable housing and substance/alcohol use.[24,25] The emergency department encounter may present an opportunity for health department staff conducting HIV care relinkage work to engage PLWH who they could not otherwise contact. In this way, integrating HIV surveillance data with ED data in real-time could improve the efficiency and quality of HIV D2C interventions.

In 2015, Public Health Seattle and King County (PHSKC), in partnership with University of Washington (UW) Medicine, implemented a real-time data exchange in which health department staff are sent a short messaging service (SMS) notification when an out-of-care PLWH is seen at an emergency department or admitted as an inpatient at one of three UW Medicine hospitals: Harborview Medical Center (HMC), the University of Washington Medical Center (UWMC), and Northwest Hospital & Medical Center (NWH). HMC is a county hospital that includes the largest HIV clinic in the area, the Ryan White Part C-funded Madison Clinic. The public health relinkage team is located on the HMC campus, across the street from the HMC ED. After receiving an alert, health department staff attempt to contact the patient while they are still in the ED or inpatient hospital or very soon after they are discharged. This type of "real-time D2C" offers a way to prioritize re-engagement opportunities presented by D2C programs. However, the effectiveness of this program, and real-time D2C programs more broadly, is not known. The goal of this project was to determine whether the PHSKC-UW D2C program improves HIV care engagement and viral suppression among poorly engaged PLWH who present to the ED or inpatient hospital.

Methods

Data Exchange

The PHSKC-UW Medicine real-time data exchange scans the UW Medicine Enterprise Data Warehouse (EDW) every five minutes to identify patients presenting in the ED or inpatient hospital. If a patient has evidence of any previous HIV positive laboratory test and a most recent viral load >200 copies/mL within the UW Medicine system, the data exchange sends a short messaging system (SMS) notification to the PHSKC HIV care relinkage team that does not contain identifying information. The relinkage team receives the patient's identifying information using a SQL Server Reporting Services (SSRS) report. The SSRS report is updated in real-time, and patient information remains on the report until the patient is discharged. After reviewing the notification and patient information, the relinkage team contacts the nurse caring for the patient to check on the patient's status and, when possible, meet with the patient while they are in the ED or hospital to discuss HIV care re-engagement, identify barriers to HIV care, and assist with making a follow-up appointment and linking the patient to supportive services. The HIV relinkage team maintained a brief record of patient interactions in an Excel spreadsheet, however they did not systematically track all alerts. The intervention is tailored to the needs of each patient. During the period of this analysis, the data exchange sent SMS notifications only for patients registered in the ED or admitted to the hospital between 8 AM and 6 PM on Monday through Friday.

The PHSKC-UW Medicine real-time data exchange was implemented as a public health program intended to improve patient outcomes and the HIV care continuum in King County, and was determined not to be research by the University of Washington Institutional Review Board.

Evaluation Design

We used a pre/post design to evaluate the impact of the PHSKC-UW Medicine real-time data exchange on HIV care engagement and viral suppression. We defined HIV care engagement as a viral load test within 3 months after an eligible ED visit/IP admission, and viral suppression as a viral load less than 200 copies/mL in the 6 months after an eligible visit. The post-intervention period was defined as the two years after the data exchange was implemented (07/20/2015 to 07/20/2017). The pre-intervention period was defined as the two years before the data exchange was implemented with a 6-month washout period (01/20/2013 to 01/20/2015) to ensure that patients in the pre-intervention period could not have either of the study outcomes in the post-intervention period.

Data Sources

Since the alerts were not maintained in records after the patients' discharge, we used EDW data to identify ED visits/IP admissions that would have triggered an SMS alert in both the pre- and post-intervention periods. Alert eligible visits were defined as visits for patients who 1) were in the hospital on a weekday between 8 AM and 6 PM, 2) had any previous positive HIV laboratory test and 3) a viral load >= 200 copies/mL on their most recent viral load prior to their visit. Since health record data from NWH was not added to the EDW until 2016 (i.e., after the end of the pre-intervention period), we restricted our study population to individuals with visits from HMC and UWMC only.

To obtain data on HIV care outcomes following an eligible visit, we linked patient data from EDW and PHSKC's electronic HIV/AIDS reporting system (eHARS) using a probabilistic record linkage algorithm. First name, last name, gender, race/ethnicity, date of birth, and social security number were included as matching fields. The R package, fastLink, was used to conduct the match.[27] Patients with an alert-eligible visit in EDW who did not have a match in eHARS (N = 27) were manually reviewed to identify a matching eHARS record. After the manual review process, all patients with an alert-eligible visit had a matching record in eHARS. Matches identified by the matching algorithm were manually reviewed to identify any potential mismatches; no mismatches were identified.

We obtained data on the following covariates using the EDW: age (at the time of the eligible visit), gender, race/ethnicity, visit type (ED visit or inpatient admission), and facility (UWMC or HMC). We used each patient's self-reported history of injection drug use at the time of HIV diagnosis from eHARS because this information is not systematically collected in or readily extractable from the medical record.

Statistical analysis

We compared patient characteristics between the pre-intervention and post-intervention periods using t-tests and chi-square tests. Separate models were used to estimate the effect of the intervention on HIV care re-engagement within 3 months and viral suppression within 6 months. Since patients could have multiple ED visits/IP admissions, we used generalized estimating equations (GEE) with a log link function and Poisson distribution to cluster visits for the same patient. We used an autoregressive-1 working correlation structure, which assumes that observations closer to each other in time are more correlated than those further apart in time, and used the sandwich estimator to estimate robust standard errors and 95% confidence intervals. Models were adjusted for age, gender (cis-gender male, cis-gender female, transgender), race/ethnicity (Hispanic, Black, White, other), UW Medicine facility (UWMC or HMC), self-reported injection drug use at the time of HIV infection, visit type, visit month, and the number of visits in the prior 6 months.

As defined *a priori* in our analysis plan, we conducted a sensitivity analysis using a difference-in-difference approach to assess whether our primary results could be due to secular trends in HIV care engagement and viral suppression in King County during the study period. We compared patients who had an alert-eligible visit (i.e., between 8 AM and 6 PM Monday through Friday) to patients who had visits outside of the alert window (i.e., alert-ineligible visits). We then estimated the difference in this difference between the pre- and post-intervention periods. We decided to conduct a second sensitivity analysis *post hoc* of all patients who were included in the HIV care relinkage team log and thus had confirmation that the relinkage team conducted an intervention.

Results

We identified a total of 1844 ED visits/IP admissions among 479 PLWH that met criteria for being out of care/virally unsuppressed at a UW Medicine facility during our study period (Figure 2.1). Of the 1070 visits identified during the pre-intervention period, 862 (81%) visits for 276 patients occurred during the alert window, and 208 visits for 122 patients occurred outside of the alert window. Of the 774 visits identified in the post-intervention period, 647 (84%) visits for 242 patients occurred during the alert window, and 127 visits for 77 patients occurred outside of the alert window. The HIV relinkage team recorded their actions taken in response to 208 of the 647 (32%) alert-eligible visits in the post-intervention period, leading to contact with 145 of the 242 (60%) out-of-care patients with at least one alert-eligible ED visit/IP admission during the post-intervention period. Of the 1509 alert-eligible visits in the study period, 446 (30%) were inpatient admissions. A majority of alert-eligible visits identified were for patients whose most recent viral load test was within the year before their visit and who had a viral load of at least 1000 copies/mL (75% in the pre-intervention period, 69% in the post-intervention period). Overall, 90% of alert-eligible visits occurred at HMC.

Patient characteristics did not differ between the pre-intervention and post-intervention periods (Table 2.1). Patients had an average of 3.7 visits (SD = 4.4 visits) in the pre-intervention period, and 3.1 (SD = 4.5) in the post-intervention period, and there was high variability in the number of visits per patients in both periods. The majority of patients were over 40 years of age (61%) and cisgender male (81%). 45% of patients were non-Hispanic White, 32% were non-Hispanic Black, and 12% were Hispanic/Latinx. About 39% of patients had a history of injection drug use at the time of their HIV diagnosis.

The proportion of patients who had a viral load test in the 3 months after an alert-eligible visit increased from 60% to 71% between the pre-intervention and post-intervention periods (Table 2.2). Compared to patients in the pre-intervention period, patients in the post-intervention period were 1.08 (95% CI: 0.97, 1.20) times more likely to have a viral load test within 3 months of an alert-eligible ED visit/IP admission in the post-intervention period, after adjusting for age, gender, race/ethnicity, visit type, visit facility, history of injection drug use, visit month, and the number of visits in the previous 6 months, but this was not statistically significant. Similarly, viral suppression within 6 months after an alert-eligible ED visit/IP admission increased between the pre-intervention and post-intervention period (40% to 56%). Patients were 1.50 (95% CI: 1.27, 1.76) times more likely to achieve viral suppression in the 6 months after an eligible visit in the post-intervention period compared to patients in the pre-intervention period, after adjusting for age, gender, race/ethnicity, visit type, visit facility, history of injection drug use, visit month, and the number of visits in the previous 6 months.

Among patients who met the criteria for unsuppressed viral load but had a visit outside of the time window for alerts, 60% in the pre-intervention period and 65% in the post-intervention period had a viral load test within 3 months after an ED visit/IP admission (Table 2.3). In addition, viral suppression within 6 months after an ED visit/IP admission increased among patients who had visits outside of the alert window from 35% in the pre-intervention period to 47% in the post-intervention period. There was no difference in viral load testing within 3 months after an ED visit/IP admission between patients who had a visit within the alert window and patients who had a visit outside of the alert window in both the pre-intervention period (aRR: 1.00; 95% CI: 0.89, 1.11) or in the post-intervention period (aRR: 1.00; 95% CI: 0.87, 1.14). Although viral load testing after ED visit/IP admission increased in both groups between the pre-intervention and post-intervention groups, there was no significant difference in difference in this increase (Figure 2.2; DID: 1.00, 95% CI: 0.84, 1.18). Similarly, there was no significant difference in viral suppression between patients who had a visit within the alert window and patients who had a visit outside of the alert window in both the pre-intervention period (aRR: 1.08; 95% CI: 0.99, 1.19) and post-intervention period (aRR: 1.09; 95% CI: 0.93, 1.27). Although viral suppression increased between the pre-intervention and post-intervention period among both of these groups of patients, there was no significant difference in difference in these increases (DID: 1.01; 95% CI: 0.84, 1.20).

There was no statistically significant difference in viral load testing within 3 months (RR: 1.06, 95% CI: 0.94, 1.25) and viral suppression within 6 months (RR: 0.85, 95% CI: 0.71, 1.02) after an eligible visit between patients who were contacted by the HIV relinkage team and patients who were not contacted by the HIV relinkage team.

Discussion

In a pragmatic evaluation of a real-time ED-based D2C program, we found that ED visits/IP admissions provide opportunities to interact with a substantial number of poorly engaged PLWH. While we found no difference in viral load testing in the 3 months after ED visits/IP admissions between the pre- and post-intervention periods, patients with ED visits/IP admissions in the post-intervention period were 50% more likely to achieve viral suppression in the 6 months after their visit compared to patients in the pre-intervention period. However, we found a similar increase in viral suppression among patients with ED visits/IP admissions outside of the alert window, suggesting that the observed pre/post effect could be due to secular trends. Similarly, we found no difference in HIV care engagement or viral suppression between patients with confirmed contact with the HIV relinkage team and those with no contact, which further indicates that the pre/post increases we observed were not directly attributable to the intervention.

During our study period, the proportion of all PLWH who were virally suppressed increased from 74% in 2013 to 84% in 2018 (KC HIV surveillance report). Additionally, in the post-intervention period, King County implemented a number of interventions designed to improve care engagement and viral suppression, including the opening of a low-barrier clinic for difficult to engage patients, and clinic- and community-based peer navigation and case management programs. While the data exchange may have worked in tandem with these competing interventions, we cannot attribute the observed pre/post increase to the data exchange alone. Patients with ED visits/IP admissions within and outside of the alert window may have benefited from this increase in HIV care engagement services, which may have diminished the independent impact of the data exchange. Insofar as other areas might not have competing interventions, the impact of a data exchange on HIV care re-engagement could be greater.

Nonetheless, data exchange between public health department HIV programs and ED/IP hospitals has the potential to reach a substantial proportion of the population of out-of-care PLWH. In King County in 2015, surveillance data indicate that 1223 PLWH were out-of-care or virally unsuppressed. The PHSKC-UW Medicine real-time data exchange identified 242 unique patients in the post-intervention period, representing about 20% of the out-of-care or virally unsuppressed PLWH population in 2015. Expanding our data exchange to include other ED/IP hospitals could expand the reach of our intervention even further. Nationally, PLWH had about 633 ED visits per 1000 person-years between 2000 and 2010, which was 40% higher than the rate among non-HIV-infected persons (433 visits per 1000), and groups that are disproportionately out of care have higher rates of ED utilization.[26,28,29] Real-time data exchange between health departments and EDs could have leverage ED visits as an opportunity to improve HIV care engagement and viral suppression and address disparities in the US HIV care continuum.

This study, to our knowledge, is the first real-world evaluation of the use of real-time data exchange between ED and health departments to improve HIV relinkage to care. While this strategy has been described in other settings, our study is the first to assess the impact of such a data exchange on both HIV care re-engagement and viral suppression and the first to use a comparison group in its evaluation. In 2009, the Louisiana Office of Public Health implemented the Louisiana Public Health Information Exchange (LaPHIE).[23,30] LaPHIE used HIV surveillance and EMR data from Louisiana State University Health Care Services Division facilities to send alerts to ED providers in order to improve HIV care engagement for out-of-care PLWH. While LaPHIE relied on ED providers to re-engage patients back to HIV care, the PHSKC-UW Medicine real-time D2C program used health department HIV relinkage specialists. Among patients identified through LaPHIE, 82% had at least one viral load or CD4 test in the 18 months after an ED visit. In 2016, Ridgeway et al developed a data exchange between the University of Chicago Medicine (UCM) system and the Chicago Department of Health (CDPH), in which CPDH HIV care navigators were notified when out-of-care PLWH were seen at a UCM ED/IP setting.[31] This data exchange identified 56 out-of-care patients seen at a UCM ED/IP setting from July 2016 through August 2017, of whom 66% were re-engaged in HIV care before January 31, 2018. Although the LaPHIE and UCM interventions were able to re-engage high proportions of PLWH seen in the ED/IP setting, without a comparison group, these findings may reflect the effect of competing interventions, and not necessarily the independent effect of the data exchange. Our study found similar proportions of viral load testing after an ED/IP visit to previous studies (in a shorter time period), but did not find a significant difference in relinkage when compared to both retrospective controls and in a difference-in-differences comparison. Further research is needed to confirm our findings and assess the impact of real-time data exchange between health departments and ED/IP hospitals in a setting with fewer competing interventions.

Our study had several limitations. First, although we used data on all alert-eligible ED visits/IP admissions in our evaluation, we had limited data on the reasons why the HIV relinkage team did not respond to some alerts. The HIV relinkage team did record their responses to some alerts, which we used to evaluate the impact of having contact with the HIV relinkage team on HIV care outcomes. However, the HIV relinkage team may have recorded their response to alerts for patients that had relatively high barriers to care and needed additional follow-up to re-engage in care and achieve viral suppression. These patients may have been systematically harder to re-engage in HIV care than patients who had an alert-eligible visit, but no documented contact with the HIV relinkage team. Comparing these groups would result in a spuriously low relative risk estimate. Indeed, we found that patients with contact with the HIV relinkage team were less likely (although not statistically significant) to achieve viral suppression within 6 months compared to patients who did not have contact with the HIV relinkage team, which may be a result of the described selection bias.

Second, we used HIV surveillance data to ascertain viral load testing and viral suppression after an ED visit/IP admission and assumed that absence of viral load testing indicated that the patient had not re-engaged in HIV care or achieved viral suppression. If patients had viral load tests that were not captured by PHSKC's HIV surveillance data, they would have been misclassified as not having re-engaged in HIV care or not achieving viral suppression, which would have biased our effect estimates towards the null. PHSKC's HIV laboratory reporting is estimated to be 95% complete, indicating that this may not be a major source of bias in our study.

Real-time data exchange between health departments and ED/IP hospitals can be used to identify large numbers of persons who are virally unsuppressed and out of care. This type of data exchange can be used to prioritize surveillance-based D2C programs, and potentially improve the efficiency of health department HIV care re-engagement activities. Although our evaluation findings are mixed, they indicate that the PHSKC-UW Medicine real-time D2C program may work in tandem with multiple, diverse interventions to improve HIV care engagement and viral suppression. Additional efforts are needed to improve the effectiveness of reengagement interventions that are linked to real-time D2C.

Tables

Table 2.1: Characteristics of PLWH with an ED visit/IP Admission at UW Medicine, 2013-2017

	Pre-Intervention N (%)	Post-Intervention N (%)	p-value
Patient Characteristics			
Number of patients	276 (100)	242 (100)	
Number of visits per patient (Mean (SD))	3.7 (4.4)	3.1 (4.5)	0.11
Age			0.18
19-29	35 (13)	32 (13)	
30-39	78 (28)	56 (23)	
40-49	75 (27)	86 (36)	
50 or older	88 (32)	68 (28)	
Gender			0.07
Cisgender Male	223 (81)	199 (82)	
Cisgender Female	52 (19)	37 (15)	
Transgender	1 (<1)	6 (3)	
Race/Ethnicity			0.98
Hispanic/Latinx	34 (12)	31 (13)	
Black	91 (33)	76 (31)	
White	121 (44)	109 (45)	
Other	30 (11)	26 (11)	
History of Injection Drug Use	105 (38)	94 (39)	0.86
Visit Characteristics			
Alert-eligible visits	862 (100)	647 (100)	
Inpatient Admissions	247 (29)	199 (31)	0.39
HIV Care Status at Visit			0.03
Most recent VL >= 1000 copies/mL and no VL in past year	105 (12)	83 (13)	
Most Recent VL >= 1000 copies/mL in past year	646 (75)	447 (69)	
Most recent VL 200-1000 copies/mL and no VL in past year	14 (2)	13 (2)	
Most Recent VL 200-1000 copies/mL in past year	97 (11)	104 (16)	
Facility			0.35
Harborview Medical Center	779 (90)	575 (89)	
UW Medical Center	83 (10)	72 (11)	

Table 2.2: Comparison of HIV care outcomes after ED visit/IP admission in post-intervention and pre-intervention periods

		Viral load test within 3 months after visit			Viral suppression within 6 months after visit		
Study Period	**Number of visits**	**N (%)**	**Unadjusted RR (95% CI)**	**Adjusted RR (95% CI)[1]**	**N (%)**	**Unadjusted RR (95% CI)**	**Adjusted RR (95% CI)[1]**
Pre-intervention	862	515 (60)	Ref.	Ref.	356 (41)	Ref.	Ref.
Post-intervention	647	461 (71)	1.09 (0.98, 1.22)	1.08 (0.97, 1.20)	376 (58)	1.48 (1.26, 1.74)	1.50 (1.27, 1.76)

[1]Adjusted for age, gender, race/ethnicity, history of injection drug use (at time of HIV diagnosis), visit type, facility, visit month

Table 2.3: Association between study period, alert eligibility, and HIV care re-engagement outcomes after ED visit/IP admission

			Viral load test within 3 months after visit		Viral suppression within 6 months after visit	
Study Period	**Alert Eligibility**	**Number of visits**	**N (%)**	**aRR (95% CI)[1]**	**N (%)**	**aRR (95% CI)[1]**
Pre-Intervention	Outside Alert Window	208	125 (60)	Ref.	73 (35)	Ref.
	Within Alert Window	862	515 (60)	1.00 (0.89, 1.11)	356 (41)	1.08 (0.99, 1.19)
Post-Intervention	Outside Alert Window	127	83 (65)	Ref.	60 (47)	Ref.
	Within Alert Window	647	461 (71)	1.00 (0.87, 1.14)	376 (58)	1.09 (0.93, 1.27)
Difference-in-difference (post vs. pre)		-	-	1.00 (0.84, 1.18)	-	1.01 (0.84, 1.20)

[1]Adjusted for age, gender, race/ethnicity, history of injection drug use (at time of HIV diagnosis), visit type, facility, visit month

Figures

Figure 2.1: Study participants flow diagram

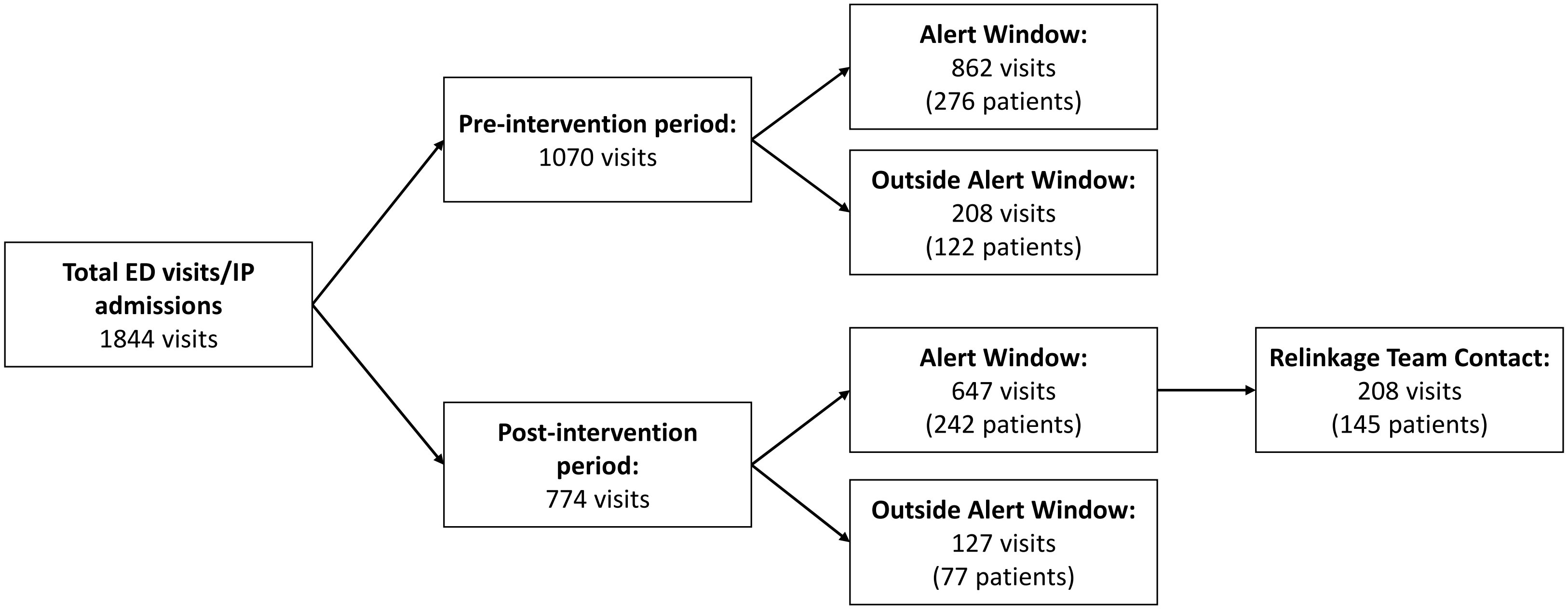
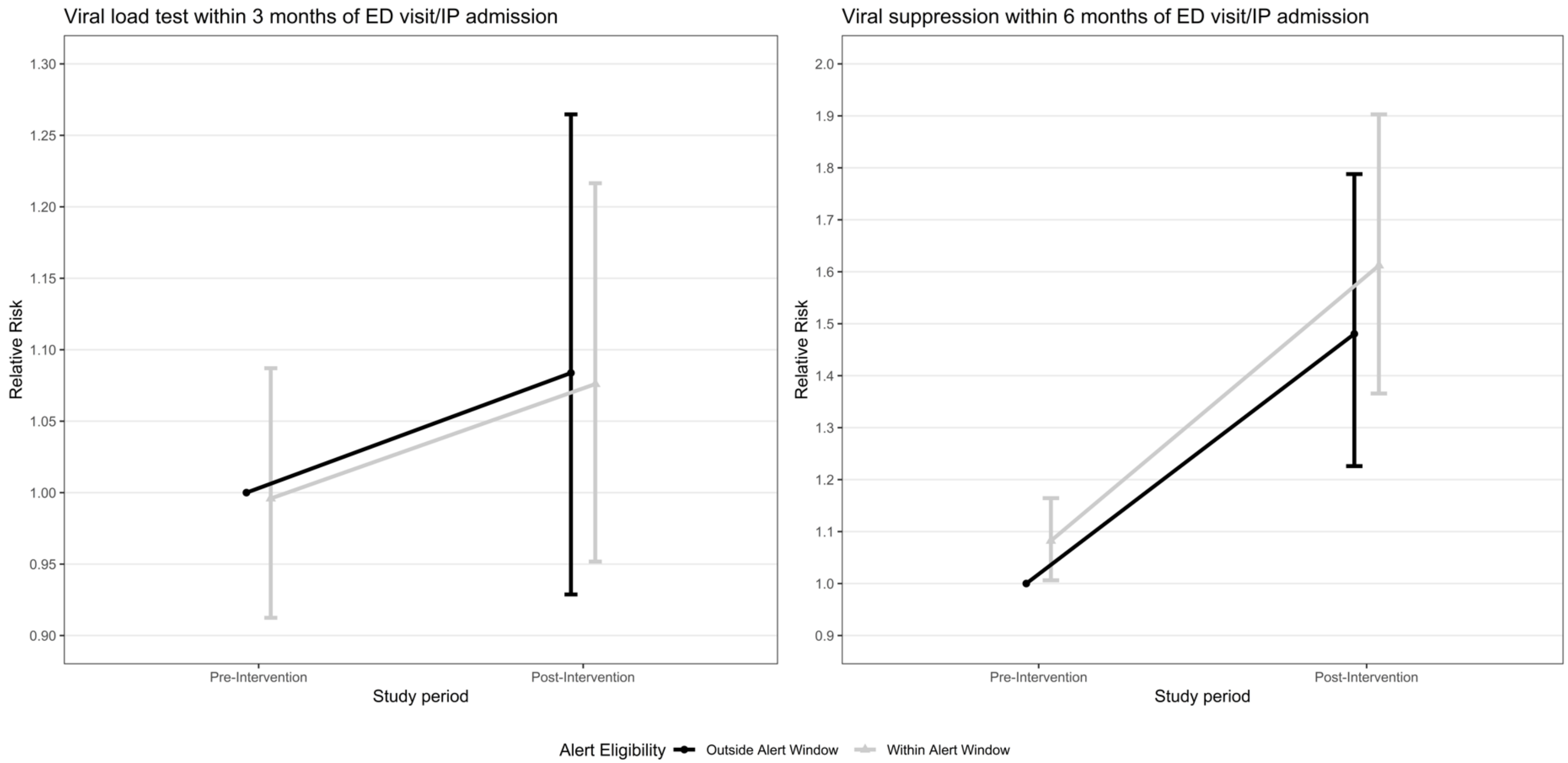


Figure 2.2: Difference-in-differences model results



Chapter 3: Integrating HIV surveillance and jail booking data to facilitate post-incarceration HIV care re-engagement

Abstract

Objective: To evaluate the impact of using an automated, real-time data exchange between public health HIV surveillance and county jail data to identify incarcerated persons living with HIV (PLWH) and facilitate post-incarceration HIV care engagement efforts.

Methods: We developed an automated data exchange (JBLink) that matched Public Health Seattle King County (PHSKC) HIV surveillance data with jail booking rosters from two King County jails. A team of public health relinkage specialists and jail release planners used the data to guide case conferences about patients who were out-of-care or virally suppressed and jointly developed a plan for re-engagement in care and treatment. We compared viral load testing within 3 months and viral suppression within 6 months after release from jail among virally unsuppressed or out-of-care PLWH released in the post-intervention period (04/01/18-11/01/18) to those released in the pre-intervention period (10/01/16-10/01/17) using Cox proportional hazards models.

Results: We identified 153 bookings for virally unsuppressed or out-of-care PLWH released from jail in King County in the pre-intervention period and 80 bookings in the post-intervention period. Viral load testing within 3 months post-incarceration increased from 43% in the pre-intervention period to 58% in the post-intervention period, and viral suppression within 6 months post-incarceration increased from 36% to 44%. Incarcerated PLWH released in the post-intervention period were 1.35 times more likely to have a viral load test within 3 months after

release (95% CI: 0.84, 2.18), and 1.37 times more likely to be virally suppressed within 6 months after release (95% CI: 0.82, 2.30), but these differences were not statistically significant.

Conclusions: Implementation of JBLink resulted in a trend towards improved post-incarceration HIV care outcomes for virally unsuppressed/out-of-care PLWH in King County. Real-time data exchange between health departments and county jails is a promising strategy for identifying incarcerated PLWH to support care coordination and improving post-incarceration HIV care engagement.

Introduction

Persons living with HIV (PLWH) in the United States are disproportionately incarcerated. Rates of incarceration among PLWH are between three and five times the rate in the general US population, and approximately 1 in 7 PLWH pass through correctional facilities every year.[32,33] Incarceration can disrupt continuous HIV care engagement and PLWH who have been incarcerated experience lower rates of retention in care and viral suppression after release from prison or jail compared to the general HIV-positive population.[34–36] Furthermore, incarceration disproportionately affects groups that experience disparities in HIV care outcomes, such as Black and Hispanic/Latino men (particularly men who have sex with men), and transgender persons, and may compound disparities in HIV outcomes among these groups.[37–43] HIV care engagement services and interventions specifically aimed at improving post-incarceration HIV care are critical to ensuring that all PLWH have access to services to facilitate reaching viral suppression and may also reduce disparities in HIV health outcomes.

Existing efforts to improve post-incarceration HIV care engagement have utilized a variety of approaches, including peer navigation, care coordination, and case management, and have had some promising results.[44–47] These efforts have focused primarily on the prison setting, where individuals have longer stays, and typically receive HIV care while incarcerated. The delivery of these interventions in the jail setting may be more challenging because stays are shorter and volumes are higher. Because of the shorter length of stay, many PLWH incarcerated in jail may not disclose their HIV status. As a result, these individuals would be missed by efforts to improve HIV care engagement after release from jail, even though they may still experience a disruption in their HIV care engagement. One strategy to improve HIV care engagement after incarceration in the jail setting is to integrate jail booking data with HIV surveillance in order to identify incarcerated PLWH. Real-time data exchange between jails and HIV surveillance programs may improve the effectiveness of post-incarceration HIV care re-engagement efforts in the jail setting by ensuring that all PLWH released from jail are linked to services they need to stay engaged in HIV care. Many health departments in the United States already use HIV surveillance data to identify out of care PLWH to provide HIV care engagement assistance, a program known as "Data to Care (D2C)."[9,10,48–50] However, the efficiency and effectiveness of D2C programs that rely on HIV surveillance data alone have been limited due to the lack of real-time data.[9,12,14,51,52] Real-time data exchange between HIV surveillance programs and county jails could aid health department D2C efforts by making them more efficient and concentrating on a marginalized population of PLWH.

In April 2018, Public Health Seattle King County (PHSKC) and the King County Department of Adult and Juvenile Detention (DAJD) developed an automated daily data exchange (called "JBLink") to identify incarcerated PLWH and provide assistance to facilitate HIV care engagement after release from jail. Every day, HIV surveillance data are matched with jail booking rosters from two King County jails (King County Jail [KCJ] and Maleng Regional Justice Center [RJC]), and health department HIV program staff are sent a SMS notification when there a newly booked PLWH is identified. Health department staff work with jail release planners to assess needs and barriers to HIV care, develop a post-release care plan for all incarcerated PLWH who are virally unsuppressed or poorly engaged in care, and provide services to ensure engagement in HIV care post-incarceration. The goal of our study is to describe the implementation of JBLink, and evaluate its impact on post-release HIV care engagement and viral suppression among incarcerated PLWH in King County.

Methods

Data Exchange

In partnership with King County DAJD, we developed an automated data exchange to match HIV surveillance data and jail booking rosters from KCJ/RJC. Each day, DAJD uploads a dataset containing 1) all new bookings in the past day, 2) all individuals currently incarcerated, and 3) all individuals released in the past 36 hours, to a secure PHSKC server. The jail booking rosters are matched with HIV surveillance data using a probabilistic record linkage algorithm.[27] Incarcerated PLWH identified by the match are exported to the JBLink database, and a SMS alert is sent to health department staff notifying them of the number of new matches.

After receiving a text alert, health department staff conduct a record review to confirm that the identified matches are true matches, determine each match patient's current HIV care status, their HIV care history, and their HIV care needs. Each patient is assigned an intervention/service intensity based on their current and previous HIV care history and additional case-by-case assessment (Table 3.1). We developed these intervention/service intensity tiers to better tailor re-engagement interventions to the needs of each individual. The tiers and associated intervention activities were developed and iteratively refined through discussions with health department staff. Importantly, patients in the “investigation only” tier were determined to be engaged in HIV care, on ART, virally suppressed and not in need of any intervention after record review. Health department staff then work with King County Jail Health Services release planners in the context of weekly case conferences and additional as needed conversations to develop an individualized post-release care plan and connect individuals with services to ensure HIV care engagement after release from jail. Health department staff systematically documented the intervention services provided to each individual in the JBLink tracking database using both free-text descriptions of the post-release plan and a pre-defined list of common actions. These actions included appointment scheduling assistance, discussion with jail case managers or medical case managers in the community, and recruitment into the MAX clinic (PHSKC’s high intensity, low threshold, incentivized care clinic). In addition, PHSKC’s HIV/STD program began having weekly case conferences with Jail Health Services and community partners to discuss release plans of known PLWH incarcerated at KCJ and RJC. When possible, health department staff discussed post-release plans for individuals identified through JBLink during these case conferences, and coordinated their efforts with release planners and community partners.

As a continuous quality improvement activity, health department staff, including two members of the study team (TA and JD), had biweekly meetings to review case patients identified through JBLink and the actions taken to re-engage them in HIV care. These meetings were used to identify and address any barriers to relinkage to care that arise over the course of the study.

Evaluation Design and Study Population

We used a pre/post design to evaluate the impact of JBLink on post-release HIV care engagement and viral suppression. We defined HIV care engagement as a viral load test within 3 months after release from jail, and viral suppression as a viral load less than 200 copies/mL in the 6 months after release from jail. We included all PLWH who were released from KCJ/RJC in the seven months after JBLink was implemented (04/01/2018 to 10/31/2018). The pre-intervention period was defined as the year (10/01/2016 to 10/01/2017) before the data exchange was implemented with a 6-month washout period to ensure that individuals in the pre-intervention period could not have either of the study outcomes in the post-intervention period. While health department staff conducted extensive record review to identify case patients who were confirmed in care and on ART (and subsequently not provided any intervention), no similar record review was conducted for case patients in the pre-intervention period. Since a majority of these patients also had a viral load less than 200 in the year before their booking in HIV surveillance data, we restricted the study population to PLWH whose last viral load was at least 200 copies/mL or had no viral load tests in the year prior to their booking date. In addition to our pre-post comparison, we also examined the characteristics, intervention actions provided, and HIV care outcomes among *all* PLWH identified by JBLink in the post-intervention period.

JBLink was developed and implemented as a public health program intended to improve patient outcomes and the HIV care continuum in King County. This project was determined not to be research by the University of Washington Institutional Review Board.

Data Sources

To construct the pre-intervention group, we obtained a dataset containing all persons incarcerated in KCJ and RJC between 10/01/2016 and 10/01/2017. We matched this dataset with HIV surveillance data to identify all PLWH released from these jails from 10/01/2016 and 10/01/2017 using the same probabilistic matching algorithm that we used in JBLink. We manually reviewed matches to identify any false matches, which we excluded from the study population. We included all PLWH identified through JBLink who were released from jail between 04/01/2018 and 10/31/2018 in the post-intervention group.

We used PHSKC HIV viral load laboratory reporting data to define HIV care re-engagement and viral suppression. If patients did not have a viral load reported in the follow-up period after release, they were assumed to not have re-engaged in care and to still be virally unsuppressed. A combination of HIV surveillance and DAJD data were used to define the following covariates: age (at the time of booking), gender (from HIV surveillance), race/ethnicity (from HIV surveillance), history of injection drug use (at the time of HIV infection), number of prior bookings since the start of each study period, the length of time incarcerated, release type (transfer of custody or otherwise), HIV care status at booking (last viral load $\geq$ 200 copies/mL in the past year, last viral load < 200 and no lab in the past year, last viral load $\geq$ 200 and no viral load in the past year), and the release date type (weekday/weekend).

Statistical analysis

We compared patient characteristics between the pre- and post-intervention periods using t-tests and chi-square tests. Kaplan-Meier curves and Cox proportional hazards models were used to estimate the effect of the intervention on viral load testing within 3 months and viral suppression within 6 months after release from jail. For each individual, we defined entry into the cohort as the date of release from jail. For viral load testing, individuals contributed follow-up time until their first viral load test reported in HIV surveillance data and were censored if they had a subsequent jail booking in KCJ/RJC or at 3 months after release from jail. For viral suppression, individuals contributed follow-up time until their first viral load test with a viral load less than 200 copies/mL and were censored if they had a subsequent jail booking in KCJ/RJC or at 6 months after release from jail. Individuals were allowed to re-enter the cohort if they had any subsequent incarcerations and were released in the study period. Kaplan-Meier curves were used to estimate the proportion of patients who had a viral load test within 3 months after release and viral suppression within 6 months after release. A Cox proportional hazards model was used to test if the hazard of testing/suppression was different between the pre- and post-intervention periods. Models were adjusted for age, race/ethnicity, gender, history of injection drug use, age, number of previous bookings, time incarcerated, release type (transfer of custody or otherwise), and release date type (weekday/weekend). To account for multiple bookings in the Cox models, we used the sandwich estimator to calculate cluster robust standard errors and 95% CI, clustering on the individual.

Since intervention intensity and process outcomes were not available for PLWH booked in the pre-intervention period, we descriptively investigated process outcomes in the post-intervention period only. Among all PLWH identified by JBLink in the post-intervention period, we compared the demographic characteristics of PLWH who were virally suppressed at booking (i.e., last viral load < 200 in the year prior to booking) to those that were virally unsuppressed/out of care at booking (i.e., last viral load $\geq$ 200 and/or no viral load in the past year), as well as the intervention services these groups received. We used Kaplan-Meier curves (as described above) to estimate the cumulative incidence of viral load testing within 3 months post-release and viral suppression within 6 months post-release by HIV care status, and intervention/service intensity.

Results

We identified 608 bookings for 342 PLWH who were released from KCJ and RJC during the study period (Figure 3.1). Of the 357 bookings in the pre-intervention period, 153 (43%) bookings for 98 people were among individuals who were virally unsuppressed and/or had no viral load tests in the year prior to booking; in the post-intervention period, 80 of the 251 (32%) bookings for 62 PLWH were among individuals who were virally unsuppressed and/or had no viral load tests in the year prior to booking.

There was no difference in age, race/ethnicity, gender, and history of injection drug use between PLWH who were virally unsuppressed and/or who had no viral load test in the past year in the pre-intervention and post-intervention periods (Table 3.2). Overall, 5% of the study population were Hispanic/Latinx, 36% were non-Hispanic Black, and 41% were non-Hispanic White. A majority of the study population was cisgender male (85%), among whom 75% were men who have sex with men (MSM). In total, about 42% of jail bookings were for PLWH with a history of injection drug use at the time of their HIV diagnosis. At booking, PLWH in the pre-intervention period were more likely to have a most recent viral load greater than 200 copies/mL and no viral load test in the year before booking than PLWH in the post-intervention period (14% vs. 1%). The median length of incarceration was 4 days in the pre-intervention period (IQR: 0, 14 days) and 3 days in the post-intervention period (IQR: 0, 15 days). The maximum number of bookings per person was greater in the pre-intervention period (10 bookings) than in the post-intervention period (6 bookings), but the median and IQR was the same in both periods (median: 1; IQR: 1, 2).

The proportion of PLWH who had a viral load test within 3 months after release from jail increased from 43% in the pre-intervention period to the 58% in the post-intervention period (Figure 3.2; Table 3.3). PLWH released from jail in the post-intervention period were 1.35 (95% CI: 0.84, 2.18) times more likely to have a viral load test within 3 months after release from jail compared to PLWH who were released from jail in the pre-intervention period after adjusting for confounding, but this finding was not statistically significant. Similarly, the proportion of PLWH who achieved viral suppression in the 6 months after release from jail increased from the pre-intervention period (36%) to the post-intervention period (44%). After adjusting for confounders, PLWH released from jail in the post-intervention period were 1.37 (95% CI: 0.82, 2.30) times more likely to achieve viral suppression than PLWH released from jail in the pre-intervention period, but this finding was also not statistically significant.

Among all bookings in the post-intervention period only, a greater proportion of virally unsuppressed/out of care PLWH were categorized as medium or high intervention/service intensity (48%) compared to virally suppressed PLWH (3%; Table 3.4). As expected, PLWH who were virally unsuppressed/out of care at booking were more likely to need health department intervention than virally suppressed PLWH. For both groups, the most common intervention services provided was referral to jail (41%) or community (10%) case manager. 20% of virally unsuppressed/out of care PLWH were referred to the MAX clinic, compared to only one virally suppressed PLWH.

In the post-intervention period, a similar proportion of PLWH who were virally unsuppressed/out of care at booking and PLWH who were virally suppressed at booking had a viral load test within 3 months after release from jail (58% and 56%, respectively; Table 3.5). The proportion of virally unsuppressed/ PLWH with a viral load test within 3 months after release was lower in the high (22%) and medium (56%) intensity tiers than in the low (68%), minimal (62%), and investigation only (100%) tiers. Virally unsuppressed/out of care PLWH had a lower cumulative incidence of viral suppression in the 6 months after release from jail than virally suppressed PLWH (44% vs. 69%). Among virally unsuppressed/out of care PLWH, the proportion of individuals who achieved viral suppression within 6 months after release was lower in the high (22%) and medium (37%) intensity tiers than in the low (45%), minimal (77%) and investigation only (49%) tiers.

Discussion

After the implementation of automated data matching between HIV surveillance and jail booking data that facilitated care coordination between jail release planners and public health relinkage specialists, we observed a trend toward increased viral load testing in the 3 months after release from jail (35% increase) and viral suppression in the 6 months after release from jail (37% increase). Although these differences were not statistically significant, real-time data exchange between health department HIV surveillance programs and county jail booking rosters along with care coordination may be a promising strategy for identifying incarcerated PLWH and providing resources to improving HIV care engagement after release from jail.

This study demonstrates how public health departments can utilize partnerships with other governmental agencies to improve the efficiency and reach of public health prevention programs. To develop JBLink, the PHSKC HIV/STD program first collaborated with the King County DAJD to strategize how to achieve the shared goal of improving continuity of care of incarcerated PLWH after release from jail, which was a major facilitator in developing post-release care plans for incarcerated PLWH and ensuring that they were linked to all necessary services post-release. The weekly case conferences between health department staff, jail release planners, and community partners were an important part of JBLink's success. These calls provided an opportunity for multiple groups to coordinate their efforts for improving the health and wellness of incarcerated PLWH after release from jail. These conference calls have been expanded to now include an HIV physician and case manager that primarily serve PLWH incarcerated and/or recently released from Washington state prison. The effectiveness of JBLink may be improved by refining the intervention and services provided to incarcerated PLWH after they are identified by the data exchange. Previous studies have considered multiple strategies for improving post-incarceration engagement (and re-engagement) in HIV care, including case management, outreach, peer navigation, and care coordination, which have had some promising results.[43–45,53–55] Integrating real-time data exchange with jails with these approaches could improve their effectiveness. JBLink could increase the reach of these interventions and provide an efficient mechanism to identify incarcerated PLWH. Similarly, identifying evidence-based interventions and delivering them based on each individual's needs could result in further improvement of HIV care outcomes of incarcerated PLWH identified through JBLink. PLWH who have been incarcerated often face complex challenges (e.g., substance use, housing stability, stigma, the post-release environment, and skills needed to engage in HIV care) that may also be barriers to HIV care engagement.[34,55–59] In our study, we observed a large degree of heterogeneity in HIV care outcomes by intervention/service intensity in the post-intervention period. Individuals in the high and medium tiers have complex life circumstances that may not have been completely addressed by our intervention. Achieving viral suppression for these individuals requires a coordinated multi-agency response that not only addresses barriers to HIV care, but all of their challenges, such as housing, substance use, and mental health treatment. Integrating services that address these challenges into JBLink could further improve its effectiveness as an HIV care engagement intervention. Further research is needed to explore the integration of JBLink with other re-engagement interventions and to develop guidelines for the design of individualized post-release care plans based on barriers to care and service need.

This study has several limitations. First, our study may not have detected a significant difference between the pre-intervention and post-intervention periods due to our small sample size. While our observed effect estimates indicate that both viral load testing and viral suppression increased between the pre- and post-intervention periods, our estimates lacked precision. Further research in a setting with a larger sample size is needed to more precisely estimate the effect of our intervention. Second, although health department staff conducted extensive record review to assign an intervention/service intensity category to incarcerated PLWH identified by JBLink, we did not conduct such a record review for individuals in the pre-intervention group. As our descriptive findings indicate, there was a large degree of heterogeneity in the proportion of incarcerated PLWH who were re-engaged in HIV care and who achieved viral suppression by intervention/service intensity. Since intervention/service intensity was primarily assigned based on HIV care status at the time of booking, we adjusted our pre/post comparisons for HIV care status (i.e., whether last viral load was at least 200 copies/mL and whether last viral load was in the past year). However, we lacked the sample size to assess whether HIV care status was an effect modifier of JBLink's impact on post-incarceration HIV care outcomes.

Third, we ascertained post-incarceration HIV care re-engagement and viral suppression using laboratory reporting in HIV surveillance data. We assumed that individuals who did not have a viral load reported after release from jail were not engaged in HIV care and remained virally unsuppressed. Individuals who re-engaged in HIV care or resumed ART without having a viral load test would have been misclassified as not having re-engaged in HIV care. However, this misclassification is likely non-differential in the pre- and post-intervention periods, and would have resulted in an underestimate of JBLink's impact on HIV care outcomes.

Real-time data exchange between health departments and jails is a promising strategy for identifying PLWH who are poorly engaged in HIV care or at high risk of falling out of HIV care. The implementation of PHSKC's JBLink data exchange, coupled with increased collaboration between PHSKC, jail health services, and community partners, resulted in an increase in both care engagement and viral suppression, although this difference was not significant due to small sample size. Further study is needed to expand and improve on the

JBLink model and better integrate data exchange into evidence-based interventions for improving post-incarceration HIV care engagement.

Tables

Table 3.1: Intervention/service intensity categories

Intervention/service intensity	Need	Example intervention
High	Not taking ART or virally unsuppressed; virally suppressed only when institutionalized; not completing regular visits; previous efforts to re-engage unsuccessful	MAX Clinic recruitment
Medium	No labs for 12 months; ART interruption; not virally suppressed, but attending regular care appointments; no previous attempts to re-engage in care	Care navigation, referral to support services, brief counseling, plan for follow-up (Dombrowski 2018)
Low	Not yet virally suppressed, but moving toward suppression based on lab reports or resumption of ART after discontinuation	Appointment scheduling assistance
Minimal	Last VL suppressed but no labs in the past year, no known ART interruption	Appointment scheduling assistance
Investigation Only	Last VL suppressed, confirmed in care and on ART	None

Table 3.2: Characteristics of PLWH released from jail, by study period

	Pre-Intervention (N = 153) N (%)	**Post-Intervention (N = 80) N (%)**	**p**
Age			0.68
20-29	40 (26)	26 (33)	
30-39	40 (26)	22 (27)	
40-49	45 (30)	19 (24)	
50 or older	28 (18)	13 (16)	
Race/Ethnicity			0.98
Hispanic/Latinx	7 (5)	4 (5)	
Black	54 (35)	30 (38)	
White	67 (44)	33 (41)	
Other	25 (16)	13 (16)	
Gender (%)			0.34
Cisgender Male	130 (85)	69 (86)	
Cisgender Female	19 (12)	11 (14)	
Transgender	4 (3)	0 (0)	
History of Injection Drug Use	65 (43)	32 (40)	0.82
HIV Risk Factor			0.96
MSM	58 (38)	32 (40)	
IDU	25 (16)	12 (15)	
MSM & IDU	40 (26)	20 (25)	
Heterosexual	10 (7)	7 (9)	
Other	20 (13)	9 (11)	
HIV care status prior to booking			0.001
Most recent VL >= 200 copies/mL more than 1 year ago	22 (14)	1 (1)	
Most recent VL >= 200 copies/mL in the past year	88 (58)	59 (74)	
No VL in the past year	43 (28)	20 (25)	
Length of Booking (days)[1]	4 (0, 209)	3 (0, 238)	0.57
Number of prior bookings[1,2]	0 (0, 9)	0 (0, 3)	<0.01

[1]Median (range)

[2]Since start of pre/post-intervention period

Table 3.3: Comparison of HIV care outcomes after release from jail in pre-intervention and post-intervention periods

	Pre-Intervention (N = 153)		Post-Intervention (N = 80)		Hazard Ratio (95% CI)	
	Events	**CI[1]**	**Events**	**CI[1]**	**Unadjusted**	**Adjusted[2]**
VL test within 3 months after release	58	43%	41	58%	1.40 (0.92, 2.13)	1.35 (0.84, 2.18)
VL suppression within 6 months after release	43	36%	30	44%	1.40 (0.87, 2.23)	1.37 (0.82, 2.30)

[1]Cumulative incidence estimated using Kaplan-Meier survival curves

[2]Adjusted for age, race/ethnicity, gender, history of injection drug use, age, number of previous bookings, time incarcerated, release type (transfer of custody or otherwise), HIV care status at booking, release date type (weekday/weekend)

Table 3.4: Characteristics and process outcomes of all incarcerated PLWH in the post-intervention period, by HIV care status

	Overall (N = 251) N (%)	VL < 200 in past year (N = 171) N (%)	VL >= 200 and/or no VL in past year (N = 80) N (%)	p-value
Age				0.001
20-29	45 (18)	19 (11)	26 (32)	
30-39	81 (32)	59 (35)	22 (28)	
40-49	69 (27)	50 (29)	19 (24)	
50 or older	56 (22)	43 (25)	13 (16)	
Race/Ethnicity				0.095
Hispanic/Latinx	22 (9)	18 (11)	4 (5)	
Black	85 (34)	55 (32)	30 (38)	
White	117 (47)	84 (49)	33 (41)	
Other	27 (11)	14 (8)	13 (16)	
Gender				0.372
Cisgender Male	222 (88)	153 (89)	69 (86)	
Cisgender Female	27 (11)	16 (9)	11 (14)	
Transgender	2 (1)	2 (1)	0 (0)	
History of Injection Drug Use	101 (40)	69 (40)	32 (40)	1.000
HIV Risk Factor				0.814
MSM	106 (42)	74 (43)	32 (40)	
IDU	31 (12)	19 (11)	12 (15)	
MSM & IDU	70 (28)	50 (29)	20 (25)	
Heterosexual	19 (8)	12 (7)	7 (9)	
Other	25 (10)	16 (9)	9 (11)	
Length of Booking (days)	3.0 [1.0, 14.5]	3.0 [1.0, 13.5]	3.0 [1.0, 15.2]	0.837
Intervention Intensity				<0.001
Investigation Only	105 (42)	94 (55)	11 (14)	
Minimal	71 (28)	55 (32)	16 (20)	
Low	31 (12)	16 (9)	15 (19)	
Medium	27 (11)	4 (2)	23 (29)	
High	17 (7)	2 (1)	15 (19)	
Final Disposition				<0.001
Needs intervention	45 (18)	13 (8)	32 (40)	

Confirmed in care and on ART	151 (60)	126 (74)	25 (31)	
Transfer of custody	7 (3)	2 (1)	5 (6)	
Missed	23 (9)	8 (5)	15 (19)	
No action taken	25 (10)	22 (13)	3 (4)	
Actions Taken				
Jail Case Manager Referral	103 (41)	48 (28)	55 (69)	<0.001
Community Case Manager Referral	25 (10)	14 (8)	11 (14)	0.252
MAX Clinic Referral	17 (7)	1 (1)	16 (20)	<0.001
Appointment Scheduling Assistance	4 (2)	3 (2)	1 (1)	1.000
Other Action Taken	17 (7)	7 (4)	10 (12)	0.028

Table 3.5: HIV Care Outcomes in post-intervention period by HIV care status before booking and intervention intensity

			VL test within 3 months post-release		VL suppression within 6 months post-release	
HIV Care Status before booking	**Intervention Intensity**	**Number of bookings (%)**	**N**	**CI[1]**	**N**	**CI[1]**
VL >= 200 and/or no VL in past year	Overall	80 (100)	41	58%	30	44%
	High	15 (19)	3	22%	3	22%
	Medium	23 (28)	12	56%	8	37%
	Low	15 (19)	9	68%	6	45%
	Minimal	16 (20)	9	62%	9	77%
	Investigation only	11 (14)	8	100%	4	49%
VL < 200 in past year	Overall	171 (100)	81	56%	95	69%
	High	2 (1)	1	50%	1	50%
	Medium	4 (2)	3	100%	2	67%
	Low	16 (10)	6	45%	9	68%
	Minimal	55 (32)	22	50%	30	74%
	Investigation only	94 (55)	49	60%	53	67%

[1]Cumulative incidence estimated using Kaplan-Meier survival curves

Figures

Figure 3.1: Study participants flow diagram

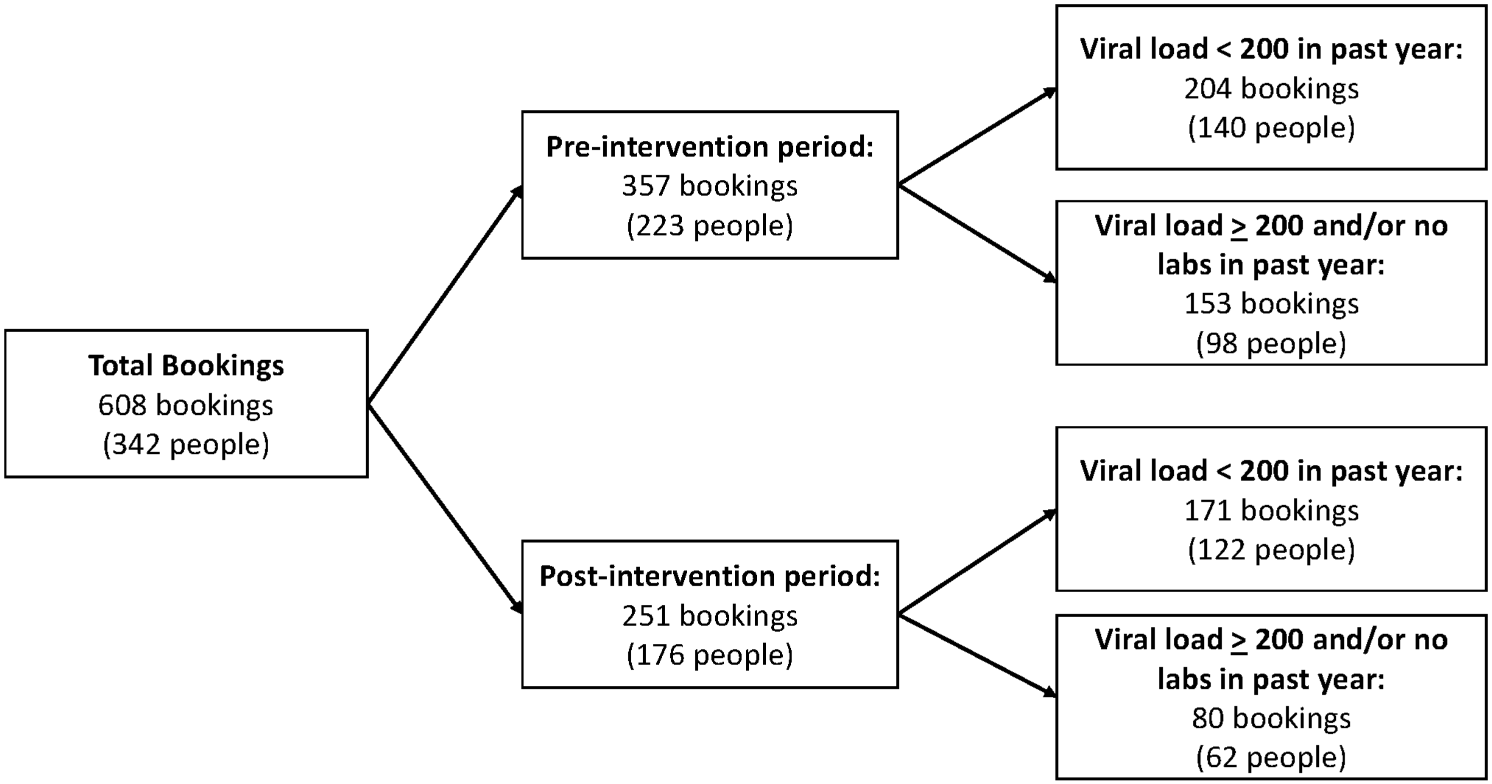
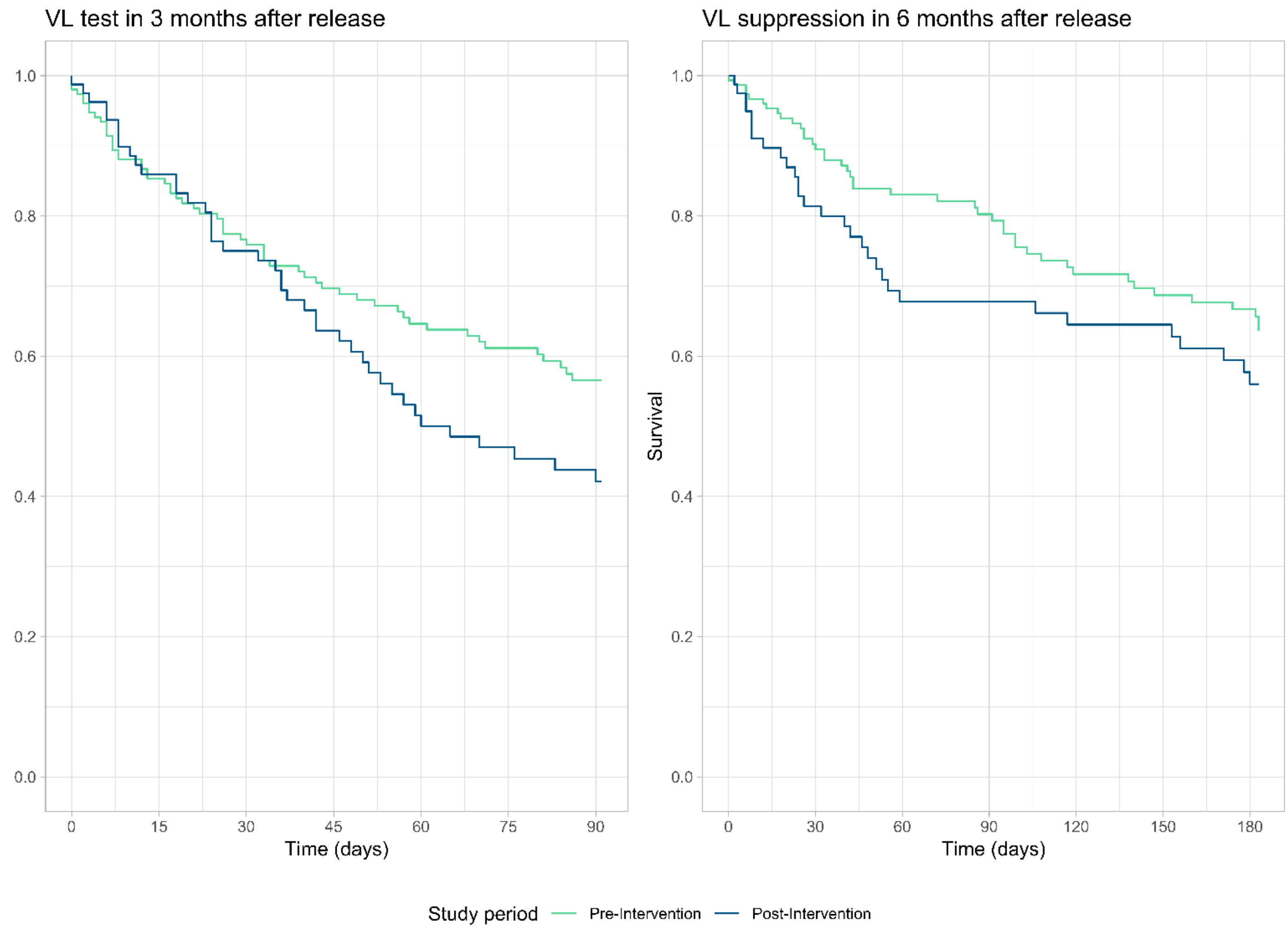


Figure 3.2: Kaplan-Meier survival curves for time to viral load test and viral suppression after release from jail, by study period



Chapter 4: Record linkage for public health action: A comparison of matching algorithms

Abstract:

Background: Many public health departments use record linkage between surveillance data and external data sources to inform public health interventions. Our goal is to compare the performance of record linkage algorithms commonly used in public health practice.

Methods: We compared five deterministic and two probabilistic record linkage algorithms using simulations and a real-world scenario. We simulated pairs of datasets while varying their error rates and overlap. We matched the datasets using each algorithm and calculated their recall (sensitivity) and precision (positive predictive value). In a real-world scenario, HIV and STD surveillance data from King County, Washington were matched to identify people living with HIV who had a syphilis diagnosis in 2017. We used manual review to define a gold standard and calculate recall and precision.

Results: Probabilistic algorithms maintained a high recall at nearly all data quality levels, while being comparable to deterministic algorithms in terms of precision. Deterministic algorithms typically failed to identify matches in scenarios with low data quality. In the real-world scenario, probabilistic algorithms had the lowest trade-off between recall and precision.

Conclusions: Probabilistic record linkage algorithms maximize the number of true matches identified, reducing gaps in the coverage of interventions and maximizing the reach of public health action.

Introduction

A central goal of public health surveillance is to provide continuous and systematically collected health-related data to inform public health practice and guide interventions to improve individual and population health. For example, health departments in the United States use HIV surveillance data [52,60–63] to identify persons living with HIV (PLWH) who are not engaged in HIV care in order to provide assistance and services to facilitate care engagement – a strategy known as Data to Care.[10,11,64–68] In this way, surveillance data are used to improve both HIV care and prevention as well as to reduce inequities in access and utilization of HIV care resources to improve the well-being of vulnerable populations with HIV.

When used in isolation from other sources of information, public health surveillance can be inefficient and ineffective. In the case of Data to Care, many people living with HIV who appear to be out of care in HIV surveillance data because they have not had a recent HIV viral load or CD4 test have actually moved out of the jurisdiction and engaged in HIV care elsewhere.[12–14] Thus, Data to Care strategies that rely entirely on HIV surveillance data involve time-consuming individual case investigations to determine whether persons are truly out of care even though that information is often readily available in other data sources, such as Ryan White-funded care programs, STD surveillance, electronic health records, or HIV surveillance systems in other jurisdictions. The Centers for Disease Control (CDC) is supporting efforts to match surveillance data between jurisdictions through programs like the “black box” system, in which HIV surveillance data from multiple jurisdictions are matched to identify people living with HIV who have moved from one jurisdiction to another.[50] In addition, several health departments are seeking to improve real-time record linkage between HIV and STD surveillance in order to provide HIV care relinkage services as part of STD partner services.[69,70]

Despite the widespread use of record linkage techniques throughout public health, little information is available to guide this process from the perspective of algorithm accuracy and the implications of missing true matches and identifying false matches. There are two primary approaches to record linkage: deterministic algorithms and probabilistic algorithms.[18,21,22] Deterministic algorithms use exact matching on specific variables or a set of matching rules to identify matched record pairs.[18] In contrast, probabilistic algorithms use statistical methods to identify the optimal set of matches, which often involves estimating and thresholding the probability that two records are a match.[18,71,72] Probabilistic algorithms typically have higher recall than deterministic algorithms, especially when linking databases that have high rates of data quality errors.[19,20] However, probabilistic algorithms also tend to be more computationally complex than deterministic algorithms, and may require more computing resources to implement in practice.[18,22]

U.S. Health department HIV/STD programs typically use deterministic record linkage algorithms to improve the quality of HIV surveillance data and its use in Data to Care.[50,73] HIV/STD programs tend to favor deterministic algorithms because they are not computationally complex, and can be executed in a short amount of time.[18,21,22] Because they are rule-based, deterministic algorithms are intuitive to understand and easy to implement and modify. Additionally (and perhaps more importantly), deterministic algorithms typically have low rates of false positive matches. Since a major concern of working with HIV data is inadvertent disclosure of HIV status, minimizing false matches is crucial to preserving individual privacy.

However, although deterministic algorithms may be highly specific, they may be overly conservative in identifying matches, leading to large numbers of missed matches. Missed matches represent missed opportunities to deliver public health interventions to individuals who need them, and depending on their distribution in the population, could magnify health inequities. Probabilistic algorithms could potentially offer increased sensitivity compared to deterministic algorithms, while only identifying a small number of false matches.

The performance of deterministic compared to probabilistic algorithms in the context of public health record linkage is unknown. The goal of this study was to compare the validity and computation time of record linkage algorithms often used in HIV/STD programs in order to better define the trade-offs between these algorithms in a variety of record linkage scenarios.

Methods

We compared deterministic and probabilistic record linkage algorithms using two approaches. First, we compared the recall, precision, and computation time of different algorithms using paired simulated datasets, varying the quality of the data and overlap between datasets (i.e. the proportion of true matches in each pair of datasets). Second, we conducted a "real-world" matching scenario involving public health surveillance data from Public Health Seattle & King County (PHSKC) to assess whether our simulation findings were generalizable to record linkage involving real datasets, where the exact error rate and overlap is difficult to assess.

Matching scenario

As the premise for our simulation study, we considered the scenario of linking records between HIV and STD surveillance data in order to identify syphilis cases reported in the past year among persons living with HIV. Such record linkage is conducted by many health departments in the United States as a way to integrate HIV care engagement activities into syphilis partner services. We assumed that both HIV and STD surveillance data contain the following shared fields that can be used for record linkage: first name, last name, date of birth (year, month, and day), gender, and race.

Matching algorithms

We compared seven algorithms used to conduct record linkage involving public health surveillance data: exact matching, four deterministic, and two probabilistic algorithms (Table 4.1). The exact matching algorithm identifies matched pairs of records between two datasets using an exact match on first name, last name, and year of birth. This was chosen as a "base case" algorithm, since it uses the simplest rule set to match two datasets. The four deterministic algorithms ("Stenger," "Ocampo 1," "Ocampo 2," and "Bosh") define rule sets for identifying a match using a combination of first name, last name, date of birth, gender, and race (Table 1).[69,73,74] These were chosen because they have been recently cited in the literature as matching algorithms used to conduct record linkage involving HIV surveillance data. Notably, the Ocampo algorithms have been used by the CDC to match interstate HIV surveillance data.[74]

The two probabilistic algorithms are "fastLink" and "beta record linkage." Fastlink is an implementation of the traditional Fellegi-Sunter approach to record linkage.[27,71] This approach uses comparisons of the shared fields between two datasets (in this case, first name, last name, year of birth, month of birth, day of birth, gender, and race) to compute the conditional probability that each record pair is a match. Record pairs are classified as "matches" or "non-matches" based on thresholding these conditional probabilities. Beta record linkage (BRL) is similar to the Fellegi-Sunter approach, but uses a Bayesian implementation to explore the space of plausible matching configurations between the datafiles.[72] By using a Bayesian approach, BRL allows for quantifying uncertainty on the matching decisions and finds the optimal set of matches by minimizing the expected misclassification errors based on a loss function.

Simulation study

Simulations were used to compare the accuracy of the selected record linkage algorithms in scenarios varying dataset size, overlap, and measurement error. GeCo, a Python-based program that creates realistic datasets of personal information, was used to generate pairs of datasets based on STD surveillance data from PHSKC's partner services data system, known as Public Health Information Management System (PHIMS).[75] In each simulation, we generated two datasets containing records of 2000 individuals each. A number of individuals were included in both datasets, which we refer to as the *overlap* between the datasets. We considered scenarios where 5%, 10%, 25%, and 50% of individuals overlapped. To generate each pair of datasets, we used the distribution of values for each field from PHIMS. Using PHIMS, we created frequency tables for first and last names, year of birth, gender (male, female, transgender male, transgender female), and race/ethnicity (Asian, Black, Hispanic/Latinx, Native American/Alaska Native, Native Hawaiian/Other Pacific Islander, White, other, multiple race). We created a joint frequency table for month and day of birth, giving an equal sampling weight for each day of the year. For each individual, a value was sampled from each frequency table to generate a number of clean records, which were then "corrupted" to create the datasets. For each pair of datasets, the first dataset consisted of "clean" records, and the second dataset consisted of "corrupted" records. Each corrupted record has a fixed number of erroneous fields that are selected at random. For each dataset size and overlap scenario, we generated datasets containing 1, 2, 3, 4, and 5 erroneous fields per record. The types of errors introduced into each field are selected at random from a set of possibilities that vary from field to field (Table 4.2). The types of errors are edits (insertions, deletions, substitutions and transpositions of characters in a string), keyboard (typing errors based on a QWERTY keyboard layout), phonetic (using a list of predefined phonetic rules), value swap (an entire value is swapped with another value selected from a pre-defined list of possible values), and missing values. The probability of missing values was determined by the frequency of missing values for each field in PHSKC's STD surveillance data. The probabilities of the remaining error types were defined based on the default probabilities provided by GeCo.

We matched each pair of datasets using each record linkage algorithm and measured their *recall* (i.e., sensitivity; the proportion of true matches identified by the algorithm), and *precision* (i.e., positive predictive value; the proportion of algorithm matches that were true matches). Each matching scenario was simulated 100 times, and we calculated the mean and standard deviation of recall and precision for each algorithm across these replicates. In addition, we measured the computational performance of each algorithm in terms of their average runtime. We ran each matching algorithm 20 times while fixing the overlap between the two datasets (50% of the individuals in second dataset overlap with the first dataset) and the number of erroneous fields (1 erroneous field per record) and varying the size of the second dataset (10%, 25%, 50%, 100% of the first dataset). We then calculated the mean and standard deviation of computation time for each algorithm.

Real-world matching scenario

In our “real-world” matching scenario, we linked PHSKC HIV (eHARS) and STD (PHIMS) surveillance data to identify PLWH who had a syphilis diagnosis in 2017. Because there is no shared unique identifier between PHIMS and eHARS, we did not have a gold standard against which we could compare each matching algorithm’s performance. Thus, we constructed a gold standard based on a composite of all seven algorithms. If all of the algorithms identified a pair of records as a match, we considered it a true match. If none of the algorithms identified a pair of records as a match, it was considered a true non-match. When there was a lack of consensus between the record pairs, we manually reviewed the records to determine whether they were a true match or non-match. As in the simulations described above, we calculated the precision and recall of each algorithm. In addition, we measured the “value and error added” by each algorithm beyond exact matching, which we considered as the baseline algorithm. We measured “value added” as the number of additional true matches and “error added” as the additional false matches identified by each algorithm over and beyond exact matching.

Dataset generation and corruption was done using GeCo and Python 2.7. All other analyses were done using R version 3.5.2. Python and R programs used to perform simulations, the real-world match, and measure computational performance are provided as Supplemental Material.

Results

Simulations

The selected deterministic algorithms had a lower recall than the selected probabilistic algorithms, regardless of overlap or number of erroneous fields per record (Figure 4.1; Table 4.3). The exact algorithm had a recall of between 56% (5% overlap) and 57% (50% overlap) when there was one erroneous field per record, and its recall decreased as the number of erroneous fields per record increased. The exact matching algorithm's precision was between 99% and 100% when there were three or fewer erroneous fields per record (Table 4.4). The Stenger, Ocampo 1, and Ocampo 2 algorithms had similar recall and precision, but had lower recall than the exact match. When there was only one erroneous field, the Stenger and Ocampo 1 algorithms both had a recall of 30%, while the Ocampo 2 algorithm had a recall of 39%, regardless of dataset size and overlap. The precision for all three algorithms was 100% when there was only one erroneous field per record. All three algorithms failed to identify any matches when there were at least three erroneous fields. The Bosh algorithm had the highest recall of the five deterministic algorithms. When there was one erroneous field per record, the Bosh algorithm's recall ranged between 74% (5% overlap) and 75% (50% overlap). However, its recall decreased to less than 20% in scenarios with at least 3 erroneous fields per record. The precision for the Bosh algorithm was high across all scenarios (between 88% and 100%).

FastLink and BRL had better recall than the deterministic algorithms. In the one erroneous field per record scenario, both fastLink and BRL had about 100% recall, regardless of dataset overlap. In the three erroneous field scenario, fastLink's recall ranged between 73% (5% overlap) and 85% (50% overlap), while BRL's recall ranged between 94% and 99%. In the five erroneous field scenario, fastLink's recall was between 8% and 27%, while BRL's recall was between 74% and 92%. The precision of both algorithms was high across all scenarios (fastLink: 97% to 100%; BRL: 85% to 100%).

Computational performance

The exact, Ocampo, and Stenger algorithms took an average of about 0.01 seconds to compute, even when the datasets being compared contained 2000 records (Figure 4.2). The Bosh algorithm took between 2 seconds and 18 seconds to compute, depending on the dataset size. The two probabilistic algorithms took a longer time to compute than all of the deterministic algorithms. FastLink took an average of between 2.3 minutes and 4 minutes to compute. On average, BRL performed faster than fastLink when the second dataset contained 200 records (1.5 minutes vs. 2.3 minutes), but was the slowest algorithm in every other scenario. BRL on average took between 3.6 minutes (second dataset N = 500) and 14.1 minutes (second dataset N = 2000) in the remaining scenarios.

Real-world matching scenario

In the real-world match of PHSKC's STD and HIV surveillance data, the exact matching algorithm identified 256 true matches, and one mismatch (Figure 4.3). Compared to this algorithm, the Stenger and Ocampo 1 algorithms identified two fewer true matches and did not have any mismatches. The Ocampo 2 algorithm identified three more matches than the exact matching algorithm, and also had no mismatches. The Bosh algorithm identified 36 additional true matches, but also identified 20 additional false matches. Both fastLink and BRL identified 53 additional true matches. However, fastLink had 33 additional false matches, while BRL only had two additional false matches.

Compared to our combined gold standard, all of the deterministic algorithms had lower recall compared to the probabilistic algorithms (Figure 4.4). The recall of the exact, Stenger, Ocampo 1, and Ocampo 2 algorithms ranged between 82% and 84%. The recall for the Bosh algorithm was about 94%, and the recall of fastLink and BRL was 100%. The precision of the deterministic algorithms (except for Bosh) was overall higher than the precision of the probabilistic algorithms. The Stenger, Ocampo 1, and Ocampo 2 algorithms had 100% precision, while the exact algorithm had 99.6% precision. The precision of the Bosh algorithm was about 93%, and the precision for fastLink was about 90%. BRL had a precision of 99%, which was the lowest trade-off between recall and precision.

Discussion

Using simulations, we found that the probabilistic algorithms we evaluated had substantially better recall than our selected deterministic algorithms, while the deterministic algorithms had higher precision. However, in scenarios with three or more erroneous fields per record nearly all of the deterministic algorithms (except for Bosh) failed to identify any matches, which diminishes their utility in record linkage scenarios where data quality is poor. In contrast, both BRL and fastLink offered high recall in our simulations without sacrificing much in terms of precision. In addition, in our "real-world" comparison, BRL had the highest recall with only a minimal sacrifice in precision and was the best performing algorithm overall.

Our findings suggest that while deterministic algorithms offer a high degree of precision, they are highly sensitive to data quality issues, and may miss a substantial number of matches even in situations where there is only one erroneous field per record. The recall of deterministic algorithms can be improved by implementing more matching rules (as in the case of the Bosh algorithm), but this also results in lower precision. Furthermore, even with additional match keys, deterministic algorithms still do not reach the level of recall offered by probabilistic algorithms.

In the context of public health action, choosing a record linkage algorithm that prioritizes identification of true matches is critical to preventing gaps in the provision of public health interventions to those who are most in need of assistance. Choosing overly conservative record linkage algorithms that prioritize precision over recall could increase gaps among these groups in public health prevention delivery and may amplify disparities among marginalized populations. Previous studies have demonstrated that imperfect record linkage algorithms may disproportionately miss women, older individuals, and persons of minoritized races/ethnicities and lower socioeconomic status.[76–79] The use of probabilistic record linkage methods (such as BRL and fastLink) or more complex deterministic algorithms (such as the Bosh algorithm) would result in a large increase in the reach of public health interventions relying on the linkage of data systems, which offsets small decreases in match precision.

A disadvantage of probabilistic algorithms is their computational complexity. While the computational time of the deterministic algorithms is generally under 1 second, both probabilistic methods took minutes to compute. For applications that require near-instant record linkage (e.g., real-time linkage between public health surveillance and electronic medical records), probabilistic algorithms may not be practical because of their slow computation time. However, when record linkage is done on a daily or less-frequent basis, the increased computation time of fastLink and BRL is less problematic. In addition, because of their increased computational complexity, BRL and fastLink require more memory and processing power than the deterministic algorithms. Both BRL and fastLink required over 4 GB of RAM and a 64-bit version of R, which may be a limitation of using these algorithms in resource-limited settings. However, 64-bit computing and 4 or more GB of RAM are becoming increasingly common, suggesting that these barriers would be less problematic in the future. As of May 2019, the estimated minimum cost of a new business desktop with these specifications starts at about $400.

Another advantage of deterministic algorithms is that these are easier to implement in different programming languages. Matching rules used by the deterministic algorithms we evaluated are relatively intuitive and translatable to multiple programming languages. Although fastLink has thorough documentation and support, modifications to the algorithm requires an understanding of the Fellegi-Sunter record linkage methodology and the R programming language.[27] Modifications to BRL are particularly challenging, as there is currently limited documentation on the method.[72] In addition, much of the BRL algorithm is implemented in the C programming language, an additional prerequisite to making modifications to the algorithm. To address these barriers, we have provided R programs for each algorithm in a “Load, Clean, Func, Do” (LCFD) framework, a portable and flexible organizational structure for developing R projects, to implement them in practice (Supplemental Material).[80]

Our study has several limitations. First, in our simulations, we assumed a uniform error rate across all records in each matching scenario. Since our probabilistic algorithms use information from across all records, this may have misrepresented how well they perform when linking datasets that contain a wide range of erroneous fields per record, including records that have zero erroneous fields. Indeed, in our real-world match scenario, in which record quality was more variable, BRL had much higher precision than in our simulations, suggesting that it is able to leverage information from record pairs that have high data quality to make decisions about record pairs that have poor data quality.

Second, the Bosh and Ocampo algorithms both include matching keys that involve social security number (SSN), which is not available in PHSKC's STD surveillance database. This may have resulted in an underestimation of the performance of these algorithms. In the Bosh algorithm, SSN is used as additional criteria to reduce mismatches for matching keys that are very broad, and its inclusion may have resulted in improved precision. In the original Bosh study, 1.7% of true matches were identified using SSN alone, suggesting that if SSN was available, we would have observed a very slightly improved recall of the Bosh algorithm, although it probably would not have reached the levels of recall observed with the probabilistic algorithms.[73] In addition, if SSN had been available, it could have also been included in both probabilistic algorithms, which could have possibly improved their recall and precision as well.

Third, we have only considered deterministic and probabilistic algorithms that can be implemented in R, and have excluded algorithms that require third party software (e.g., the Link King and CDC's Link Plus) and novel record linkage methodologies (e.g., active, supervised, and unsupervised learning algorithms). Third party software for record linkage offer a point-and-click interface for implementing probabilistic (and deterministic) record linkage methodologies. Both the Link King and Link Plus, two popular applications for conducting record linkage involving public health surveillance databases, use the Fellegi-Sunter methodology for conducting probabilistic record linkage, which is the same methodology used by fastLink. Supervised learning and active learning-based algorithms may yield greater match quality than probabilistic or deterministic algorithms in cases where databases are to be linked prospectively or when training data are available (in the case of supervised learning).[21] These algorithms use data on record pairs that are known to be matches or non-matches to develop a predictive model that is used to classify record pairs in the databases that are being linked as matches or non-matches. Since these algorithms require a training dataset of known matches and non-matches (something neither the probabilistic nor deterministic algorithms we evaluated required), we chose to exclude them from our analysis. Further research is needed to assess the performance and utility of these techniques in conducting record linkage for public health action, as well as the feasibility of implementing them in practice.

Finally, for the probabilistic matching algorithms we evaluated, we only considered their default parameterizations. We chose to evaluate these algorithms using their default (or "out-of-the-box") implementations, as this would represent a baseline level of their performance. Modifying the parameters for fastLink and BRL, such as the string distance measure used to match string variables or the number of partial agreement levels, could improve their performance. Importantly, fastLink and BRL use different default methods to match string variables (e.g., first name and last name). This may partially explain why BRL had better recall than fastLink in our simulations and a lower trade-off between recall and precision in our real-world match.

In conclusion, public health interventions that involve record linkage of multiple data systems should carefully consider their choice of record linkage algorithm. This choice should be based not only on reducing false matches, but also on maximizing intervention coverage. Record linkage methodologies that do not seek to maximize true matches, especially in the context of imperfect data quality, limit the reach of public health interventions and could exacerbate existing health disparities. Probabilistic algorithms, such as BRL, can maximize the number of true matches identified without sacrificing precision and should be considered as a first choice when using record linkage for public health action.

Tables

Table 4.1: Record linkage algorithms

Algorithm	Match criteria	Source
Exact Match	Exact match on first name, last name AND year of birth	-
Stenger	Best record pairs with a score of 50+ based on the following criteria: • +20 points: first 3 letters of last name and 2 letters of first name • +15 points: exact match on full name • +15 points: match on birth year (+/- 2 years) • +5 points: exact match on year of birth • +10 points: exact match on month of birth • +5 points: exact match on day of birth	Public Health Seattle King County; Avoundjian et al. 2018
Ocampo 1	Record pairs that met the following criteria: • Exact: last name, first name, date of birth, race, AND sex, OR • Very high: last name, first name, date of birth AND sex, OR • High: last name, first name, date of birth AND (sex OR race)	Ocampo et al. 2016
Ocampo 2	Record pairs that matched in Ocampo 1 OR met the following criteria: • Medium high: last name, first name (Soundex), date of birth, sex	Ocampo et al. 2016
NYC algorithm	Records that met any of the following matching keys: • Full last name + first 6 letters of first name + full DOB • First letter of last name + letters 3-10 of last name + letters 2-9 of first name + full DOB • Letters 2-7 of last name + first 6 letters of last name + full DOB • First 2 letters of last name + first 3 letters of first name + full SSN + full DOB • Full last name + first 3 letters of first name + full DOB • Letters 3-5 of last name + first 3 letters of first name + full DOB • First 4 letters of last name + first 4 letters of first name + full DOB • First letter of last name + letters 3-10 of last name + letters 2-9 of first name + month and year of DOB • First 5 letters of last name + first 4 letters of first name + month and year of DOB • First letter of last name + letters 3-10 of last name + letters 2-9 of first name + (day OR month of DOB) + year of DOB, switching the first and last names in 1 data set • First 5 letters of last name + first 4 letters of first name + month and year of DOB, switching the first and last names in 1 data set	Bosh et al. 2018
fastLink (Fellegi-Sunter)	Calculates match/non-match weights using an expectation maximization algorithm and computes a match probability for each record pair. Pairs are classified as a match if their match probability is above 0.85. The following fields are used to estimate the match probability: • First name & last name: partial match using Jaro Winkler string distance, with three agreement levels[1] • Year of birth, month of birth, day of birth, gender & race: exact match	Enamorado et al. 2017
Beta Record Linkage	Uses a Gibbs sampler to sample plausible matching configurations and uses a loss function to identify the optimal set of matching pairs. The following fields are used by the algorithm: • First name & last name: partial match using Levenshtein string distance, with four agreement levels[2] • Year of birth, month of birth, day of birth, gender & race: exact match	Sadinle 2017

[1]FastLink's default agreement levels for partially matched fields: 0-0.87: not a match; 0.88-0.91: partial match; 0.92+: exact match

[2]Beta record linkage's default agreement levels for partially matched fields: 0-0.49: not a match; 0.5-0.74: probable non-match; 0.76-0.998: probable match; 0.99+: exact match

Table 4.2: Error types and probabilities

Field	Error Type[1]	Probability
First name	Edits	9.9%
	Keyboard	9.9%
	Phonetic	69.3%
	Value swap	9.9%
	Missing	1%
Last name	Edits	14.55%
	Keyboard	14.55%
	Phonetic	67.9%
	Missing	3%
Year, month, and day of birth	Edits[2]	99.93%
	Missing	0.07%
Gender, race/ethnicity	Value swap[3]	99.93%
	Missing	0.07%

[1]Error types: edits – insertion, deletion, substitution, or transposition of characters in a string; keyboard – typing errors based on QWERTY keyboard layout (values close to each other on a QWERTY keyboard more likely to be swapped than values farther apart); phonetic – character substitutions based on list of phonetically similar characters; value swap – swaps first name for a nickname, alternate spelling or alias based on a pre-defined list of values; missing – value swapped with blank value to represent missing data
[2]Edit errors restricted to substitution/transposition of numeric characters only
[3]Values swapped based on list of valid genders, race/ethnicities defined in frequency tables

Table 4.3: Simulations: Record linkage algorithm recall (Mean (SD))[1]

Erroneous fields[2]	Overlap[3]	Record Linkage Algorithm: Exact	Stenger	Ocampo 1	Ocampo 2	Bosh	FastLink	Beta Record Linkage
1	5%	0.56 (0.05)	0.30 (0.05)	0.30 (0.05)	0.39 (0.05)	0.74 (0.04)	0.99 (0.02)	1.00 (0)
	10%	0.57 (0.03)	0.3 (0.03)	0.3 (0.03)	0.39 (0.03)	0.74 (0.03)	1.00 (0.001)	1.00 (0)
	25%	0.56 (0.02)	0.3 (0.02)	0.3 (0.02)	0.39 (0.02)	0.74 (0.02)	1.00 (0.0002)	1.00 (0)
	50%	0.57 (0.02)	0.30 (0.01)	0.30 (0.01)	0.39 (0.02)	0.75 (0.02)	1.00 (0.0001)	1.00 (0.0001)
3	5%	0.11 (0.03)	0 (0)	0 (0)	0 (0)	0.18 (0.03)	0.73 (0.05)	0.94 (0.02)
	10%	0.11 (0.02)	0 (0)	0 (0)	0 (0)	0.18 (0.02)	0.76 (0.04)	0.96 (0.01)
	25%	0.11 (0.01)	0 (0)	0 (0)	0 (0)	0.17 (0.02)	0.81 (0.02)	0.97 (0.01)
	50%	0.11 (0.01)	0 (0)	0 (0)	0 (0)	0.17 (0.01)	0.85 (0.01)	0.99 (0.003)
5	5%	0 (0)	0 (0)	0 (0)	0 (0)	0.01 (0.01)	0.08 (0.04)	0.74 (0.05)
	10%	0 (0)	0 (0)	0 (0)	0 (0)	0.005 (0.005)	0.14 (0.04)	0.82 (0.04)
	25%	0 (0)	0 (0)	0 (0)	0 (0)	0.004 (0.003)	0.22 (0.02)	0.89 (0.01)
	50%	0 (0)	0 (0)	0 (0)	0 (0)	0.01 (0.002)	0.27 (0.02)	0.92 (0.01)

[1]Dataset size fixed to 2000 records for both datasets

[2]Number of erroneous fields per record

[3]Percent of records in second dataset that had a match in first dataset

Table 4.4: Simulations: Record linkage algorithm precision (mean (SD))[1]

Erroneous fields[2]	Overlap[3]	Exact	Stenger	Ocampo 1	Ocampo 2	Bosh	FastLink	Beta Record Linkage
		Record Linkage Algorithm						
1	5%	0.99 (0.01)	1.00 (0)	1.00 (0)	1.00 (0.003)	1.00 (0.004)	0.99 (0.01)	0.98 (0.02)
	10%	1.00 (0.01)	1.00 (0)	1.00 (0)	1 (0.001)	1 (0.0015)	1 (0.004)	0.99 (0.009)
	25%	1.00 (0.002)	1.00 (0)	1.00 (0)	1 (0.001)	1 (0.0035)	1 (0.002)	0.98 (0.007)
	50%	1.00 (0.001)	1.00 (0)	1.00 (0)	1.00 (0.0003)	0.99 (0.01)	1.00 (0.001)	0.97 (0.01)
3	5%	0.99 (0.04)	-	-	-	1.00 (0.01)	0.98 (0.02)	0.93 (0.03)
	10%	0.99 (0.04)	-	-	-	1.00 (0.01)	0.98 (0.02)	0.93 (0.028)
	25%	0.99 (0.02)	-	-	-	1.00 (0.01)	0.99 (0.01)	0.95 (0.018)
	50%	1.00 (0.004)	-	-	-	0.99 (0.01)	1.00 (0.002)	0.97 (0.01)
5	5%	-	-	-	-	0.98 (0.09)	0.97 (0.07)	0.85 (0.04)
	10%	-	-	-	-	0.99 (0.04)	0.98 (0.03)	0.89 (0.03)
	25%	-	-	-	-	0.93 (0.14)	0.99 (0.01)	0.95 (0.01)
	50%	-	-	-	-	0.88 (0.16)	1.00 (0.01)	0.97 (0.01)

[1]Dataset size fixed to 2000 records for both datasets

[2]Number of erroneous fields per record

[3]Percent of records in second dataset that had a match in first dataset

Figures

Figure 4.1: Simulations: Record linkage algorithm recall and precision

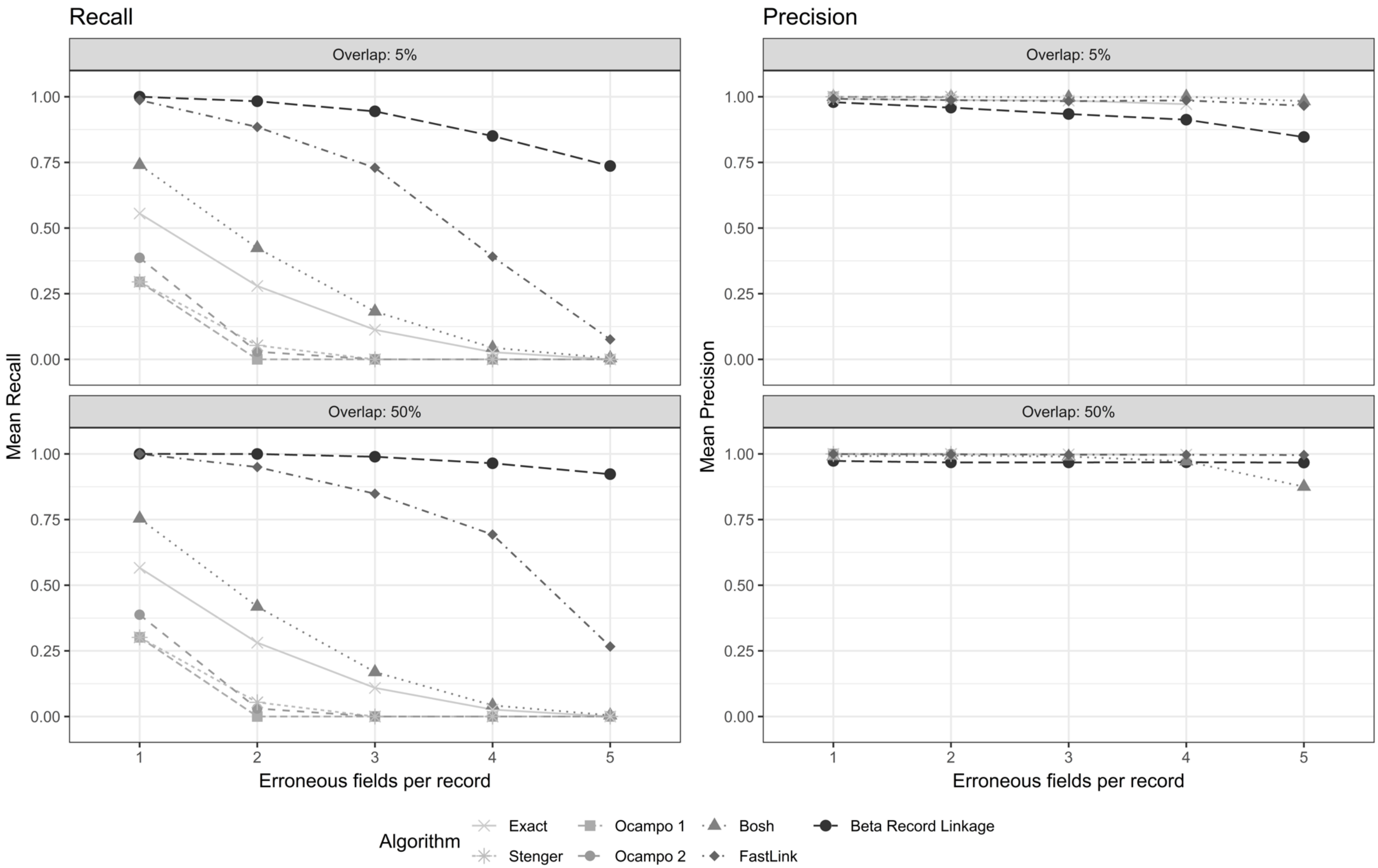


Figure 4.2: Record linkage algorithm computational performance

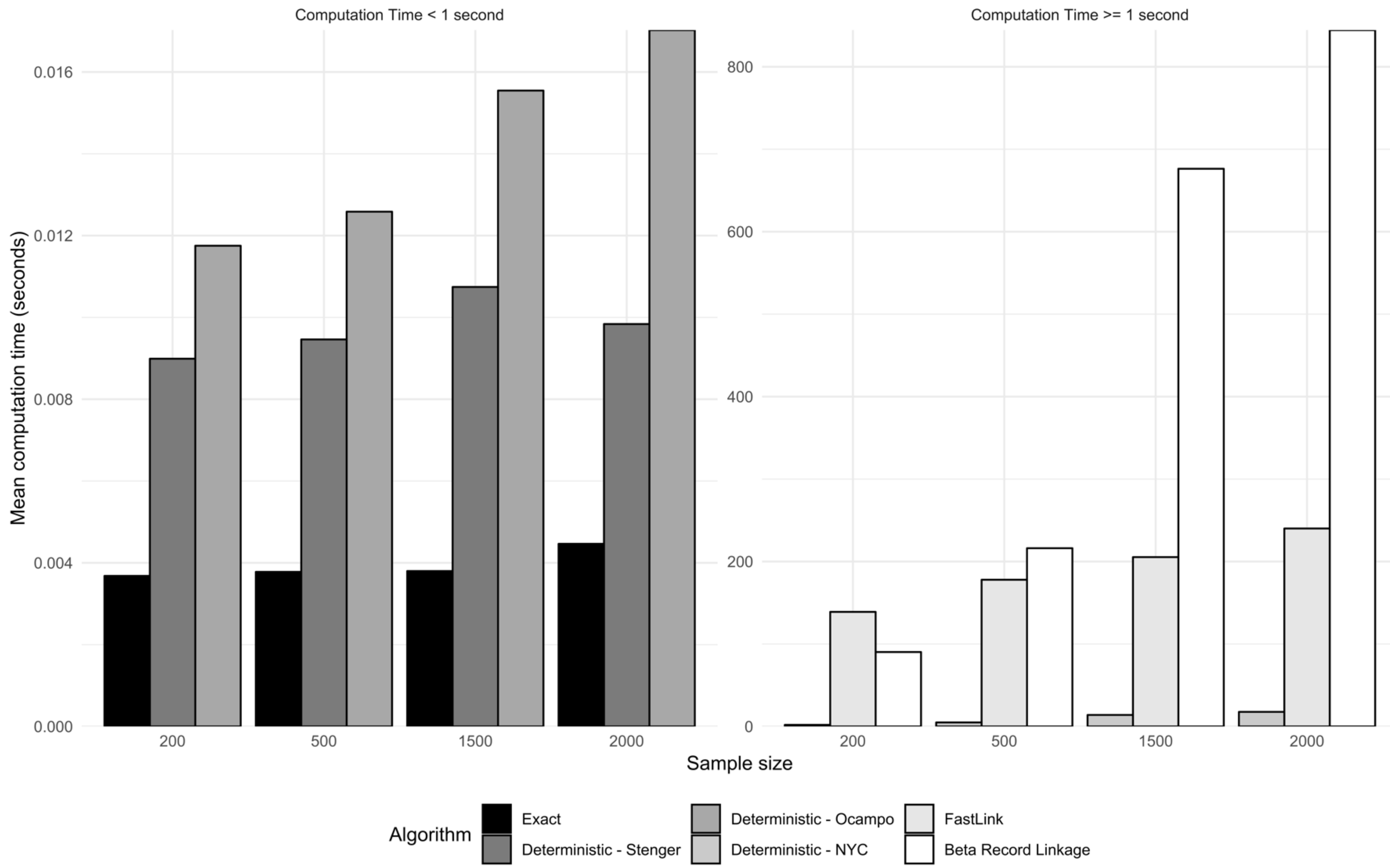


Figure 4.3: Real-world matching scenario: Value and error added over exact matching algorithm

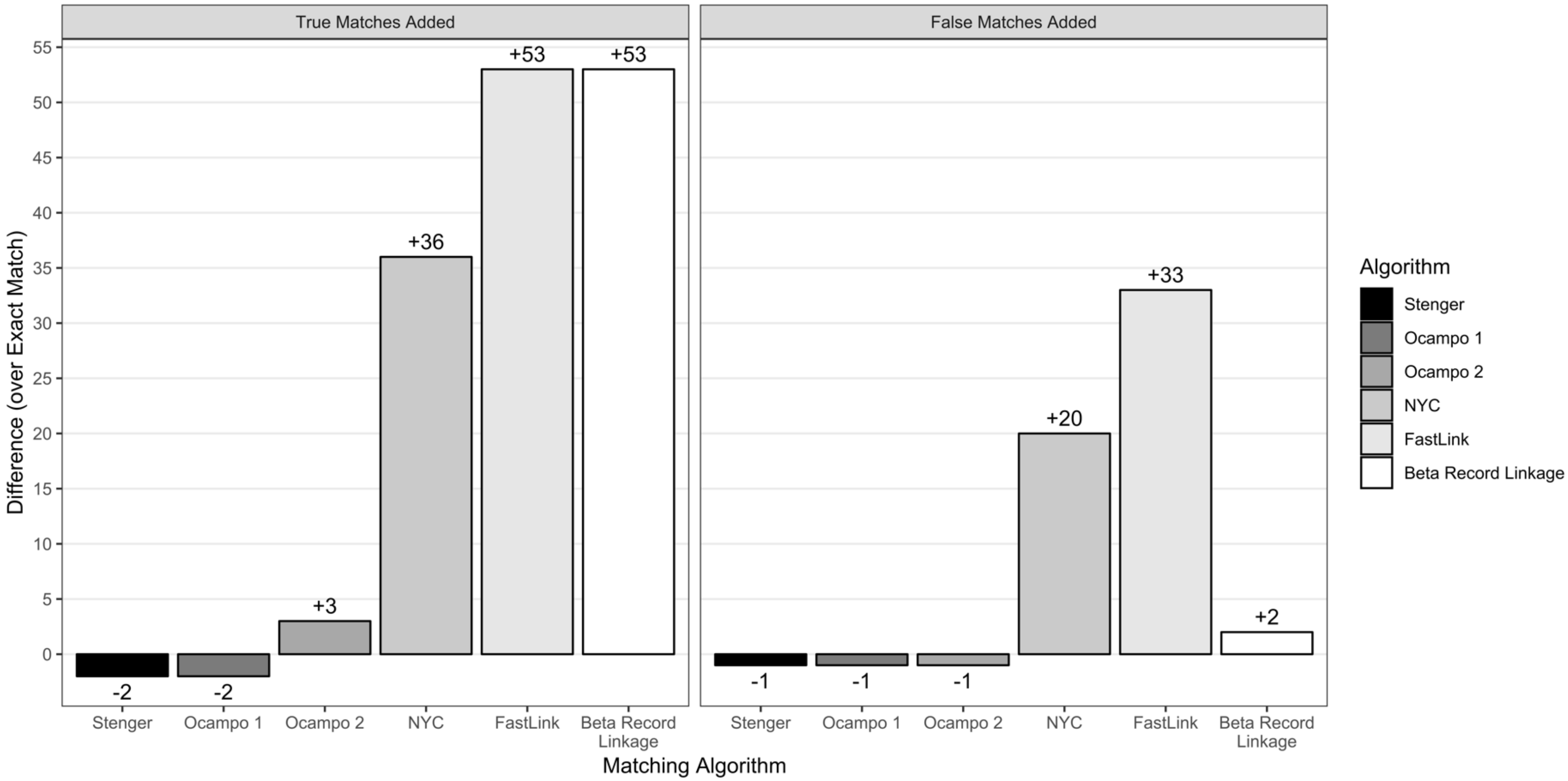
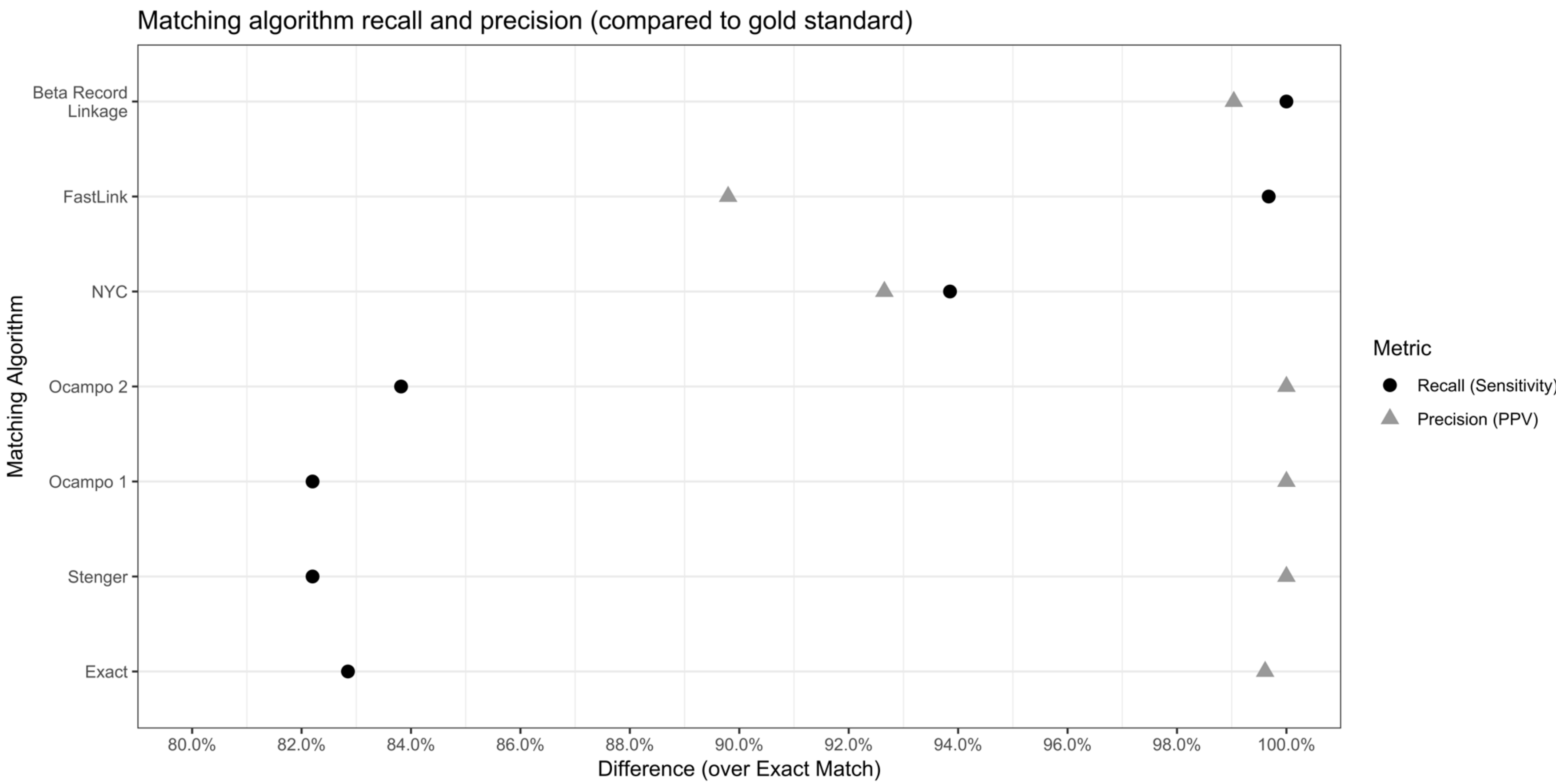


Figure 4.4: Matching algorithm recall and precision (compared to gold standard)



Chapter 5: Conclusion

In this dissertation, we evaluated two interventions that utilize integrated data systems to improve King County, WA, and examined the importance of the choice of record linkage methodology in the context of public health action. In a pre/post-intervention evaluation of an ED- and hospital-based health information exchange, we found that ED visits and inpatient hospitalizations provide opportunities to interact with a substantial number of poorly engaged PLWH. Although we observed a 50% increase in viral suppression within 6 months after an ED visit/IP admission after the data exchange was implemented, we observed a similar increase in viral suppression among patients with visits outside of the alert window, suggesting the observed effect could be due to secular trends and competing interventions. In our evaluation of an automated data exchange between HIV surveillance and jail booking data to facilitate care coordination between public health relinkage specialists and jail release planners, we observed a trend toward increased viral load testing in the 3 months after release from jail (35% increase) and viral suppression in the 6 months after release from jail (37% increase). Finally, in simulations and a real-world scenario, we found that probabilistic record linkage algorithms like BRL and fastLink have much higher recall than deterministic record linkage algorithms, while only having a small trade-off in precision.

Findings from these studies have several implications. Although data exchange between ED/inpatient hospitals and HIV surveillance data identified a large number of poorly engaged PLWH, its independent effect on HIV care engagement and viral suppression may have been overshadowed by competing interventions. During the post-intervention period of our intervention, King County implemented a number of interventions to improve care engagement and viral suppression, and patients with ED visits/IP admissions within and outside of the alert window may have benefited from this increase in HIV care engagement services, diminishing the independent impact of the data exchange. Nonetheless, data exchange between health department HIV programs and ED/IP hospitals could potentially reach a substantial proportion of the out-of-care PLWH and have a greater impact in areas that might not have competing interventions.

Data exchange between HIV surveillance and jail booking rosters is a promising strategy to facilitate post-incarceration HIV care engagement interventions. Our findings demonstrate how public health departments can utilize partnerships with other governmental agencies to improve the efficiency and reach of public health prevention programs. A major facilitator to the success of our jail-based data exchange was the collaboration between the PHSKC HIV/STD program and King County jail release planners. The weekly case conferences between health department staff, jail release planners, and community partners provided opportunities for multiple groups to coordinate their efforts for improving the health and wellness of incarcerated PLWH after release from jail. We observed a great deal of heterogeneity in HIV care outcomes by intervention intensity in the post-intervention period. Individuals in the high and medium tiers had a lower cumulative incidence of viral suppression than individuals in the low, minimal, and investigation only tiers. These individuals may have complex life circumstances that were not completely addressed by our intervention. Achieving viral suppression for individuals with high barriers to care requires a coordinated multi-agency response that addresses all of their life challenges, including housing, substance use, and mental health treatment. Integrating services that address these challenges into JBLink could further improve its effectiveness as an HIV care engagement intervention. Future research should focus on the development of guidelines for the design of individualized post-release care plans based on barriers to care and service need.

Finally, integrating data systems to improve the effectiveness and efficiency of HIV care engagement (and public health interventions, more broadly) requires record linkage of data systems that do not always contain a shared identifier. The choice of record linkage methods used to link data systems for public health action could have consequences on the reach and efficiency of these interventions. Record linkage algorithms that have high rates of false positive matches could result in breaches of individual privacy such as accidental disclosure of HIV status. On the other hand, record linkage algorithms that miss large numbers of true matches diminish the reach of interventions and could exacerbate existing health disparities. Many health department HIV/STD programs rely on deterministic algorithms because they have low rates of false positives. However, our findings indicate that the use of deterministic algorithms may be overly conservative, and result in a high degree of missed true matches. In contrast, in simulations and a real-world matching scenario, we found that probabilistic algorithms identify substantially higher numbers of true matches, while still having high precision. Although probabilistic algorithms are more computationally complex than deterministic algorithms, we found that these algorithms are not prohibitively slow, and could be practically implemented for routine record linkage activities.

In conclusion, the use of real-time data exchange between HIV surveillance and external data sources, such as ED/inpatient hospitalization data and jail booking data, is a promising strategy for identifying opportunities to provide HIV care re-engagement assistance and improving HIV care outcomes. To leverage these opportunities, these data exchanges need to be connected with comprehensive services that address barriers to HIV care and the complex life challenges that many poorly engaged PLWH experience. Improved collaboration and coordination between public health departments, medical providers, and other governmental agencies is needed to improve the effectiveness of real-time ED-, hospital-, and jail-based data exchanges that facilitate HIV care engagement. Finally, a greater emphasis is needed on maximizing true matches when using record linkage for public health action to improve the reach and efficiency of public health interventions, and reduce disparities in intervention delivery.

Acknowledgements

I would like to thank my dissertation chair and mentor, Julie Dombrowski, for her mentorship and support, and for opening doors to numerous professional opportunities in HIV/STI research. Thank you to my dissertation committee members, Matt Golden, Brandon Guthrie, Janet Baseman, and Jim Hughes, for their invaluable guidance and feedback. Thank you for all of your mentorship and advice both in shaping my dissertation and my career development.

I'd like to acknowledge the contribution of all the people who supported this work. Thank you to the PHSKC HIV/STD program, including Mark Fleming, Angela Nunez, Meena Ramchandani, Amy Bennett, Michelle Perry, Susan Buskin, Richard Lechtenberg, Dawn Spellman, Lindley Barbee, Roxanne Kerani, Sara Glick, Julia Hood, and Christina Thibault. Thank you to Mike West from King County Department of Adult and Juvenile Detention for your help to set up the jail-based data exchange described in Chapter 3. Additionally, thank you to King County IT, in particular Eric Leaf, for providing technical assistance in developing automated extracts of jail booking data. Thank you to Mauricio Sadinle for all of your guidance on record linkage algorithms.

A huge thank you to Christine Khosropour and David Katz, who have been my epidemiologist and personal role models, for all of their support, inspiration, and advice. I'd also like to thank my mentors and collaborators from other activities I've worked on during my doctoral training, including Ali Rowhani-Rahbar, Amanda Phipps, Noel Weiss, Steve Mooney, Marco Carone, Amanda Midboe, Lara Troszak, Taryn Erhardt, Kendra Johnson, David Peyton, and Christie Lewis.

I would like to acknowledge the ITHS TL1 training grant TL1 TR002318 (PI: Joie Whitney), which supported part of my dissertation work.

Finally, I would like to thank my family, friends, and most importantly my wife, for all of their support and encouragement.

References

1. WHO | Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV. *WHO*. 2017.

2. White House Office of National Aids Policy. *National HIV/AIDS Strategy for the United States: Updated to 2020*.; 2015.

3. Centers for for Disease Control Division of HIV/AIDS Prevention. *Understanding the HIV Care Continuum*.; 2018.

4. Yehia BR, Stewart L, Momplaisir F, et al. Barriers and facilitators to patient retention in HIV care. *BMC Infect Dis*. 2015;15(1):246. doi:10.1186/s12879-015-0990-0

5. Metsch LR, Pereyra M, Messinger S, et al. HIV Transmission Risk Behaviors among HIV-Infected Persons Who Are Successfully Linked to Care. *Clin Infect Dis*. 2008;47(4):577-584. doi:10.1086/590153

6. Mugavero MJ, Lin H-Y, Willig JH, et al. Missed visits and mortality among patients establishing initial outpatient HIV treatment. *Clin Infect Dis*. 2009;48(2):248-256. doi:10.1086/595705

7. Centers for Disease Control and Prevention. Data to Care. https://effectiveinterventions.cdc.gov/en/HighImpactPrevention/PublicHealthStrategies/DatatoCare.aspx

8. SWEENEY P, GARDNER LI, BUCHACZ K, et al. Shifting the Paradigm: Using HIV Surveillance Data as a Foundation for Improving HIV Care and Preventing HIV Infection. *Milbank Q*. 2013;91(3):558-603. doi:10.1111/milq.12018

9. Padilla M, Mattson CL, Scheer S, et al. Locating People Diagnosed With HIV for Public Health Action: Utility of HIV Case Surveillance and Other Data Sources. *Public Heal Methodol*. doi:10.1177/0033354918754541

10. Tesoriero JM, Johnson BL, Hart-Malloy R, et al. Improving Retention in HIV Care Through New York's Expanded Partner Services Data-to-Care Pilot. *J Public Health Manag Pract*. 2017;23(3):255-263. doi:10.1097/PHH.0000000000000483

11. Centers for Disease Control and Prevention. *Data to Care : Using HIV Surveillance Data to Support the HIV Care Continuum*.; 2017.

12. Buskin SE, Kent JB, Dombrowski JC, Golden MR. Migration distorts surveillance estimates of engagement in care : results of public health investigations of persons who appear to be out of HIV care. 2014;41(1):35-40. doi:10.1097/OLQ.0000000000000072

13. Dombrowski JC, Buskin SE, Bennett A, Thiede H, Golden MR. Use of Multiple Data Sources and Individual Case Investigation to Refine Surveillance-Based Estimates of the HIV Care Continuum. *J Acquir Immune Defic Syndr Novemb*. 2014;1(673):323-330. doi:10.1097/QAI.0000000000000302

14. Dombrowski JC, Bove J, Roscoe JC, et al. "Out of Care" HIV Case Investigations: A Collaborative Analysis Across 6 States in the Northwest US. *J Acquir Immune Defic Syndr*. 2017;74 Suppl 2:S81-S87. doi:10.1097/QAI.0000000000001237

15. Padilla M, Mattson CL, Scheer S, et al. Locating People Diagnosed With HIV for Public Health Action: Utility of HIV Case Surveillance and Other Data Sources. *Public Health Rep*.

2018;133(2):147-154. doi:10.1177/0033354918754541

16. Sweeney P, Hoyte T, Mulatu MS, et al. Implementing a Data to Care Strategy to Improve Health Outcomes for People With HIV: A Report From the Care and Prevention in the United States Demonstration Project. *Public Health Rep*. 2018;133(2_suppl):60S-74S. doi:10.1177/0033354918805987

17. Nosyk B, Lourenço L, Min JE, et al. Characterizing retention in HAART as a recurrent event process: insights into “cascade churn.” 2015;29(13). doi:10.1097/QAD.0000000000000746

18. Christen P. *Data Matching*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. doi:10.1007/978-3-642-31164-2

19. Waldenburger A, Nasseh D, Stausberg J. Detecting Duplicates at Hospital Admission: Comparison of Deterministic and Probabilistic Record Linkage. *Stud Health Technol Inform*. 2016;226:135-138.

20. Campbell KM, Deck D, Krupski A. Record linkage software in the public domain: a comparison of Link Plus, The Link King, and a `basic’ deterministic algorithm. *Health Informatics J*. 2008;14(1):5-15. doi:10.1177/1460458208088855

21. Elmagarmid AK, Ipeirotis PG, Verykios VS. Duplicate Record Detection: A Survey. *IEEE Trans Knowl Data Eng*. 2007;19(1).

22. Winkler WE. *Overview of Record Linkage and Current Research Directions*.; 2006.

23. Magnus M, Herwehe J, Gruber D, et al. Improved HIV-related outcomes associated with implementation of a novel public health information exchange. *Int J Med Inform*. 2012;81(10):e30-e38. doi:10.1016/j.ijmedinf.2012.06.005

24. Kerr J, Duffus WA, Stephens T. Relationship of HIV care engagement to emergency department utilization. *AIDS Care*. 2014;26(5):547-553. doi:10.1080/09540121.2013.844764

25. Fairbairn N, Milloy M-J, Zhang R, et al. Emergency department utilization among a cohort of HIV-positive injecting drug users in a Canadian setting. *J Emerg Med*. 2012;43(2):236-243. doi:10.1016/j.jemermed.2011.05.020

26. Hsieh Y-H, Kelen GD, Laeyendecker O, Kraus CK, Quinn TC, Rothman RE. HIV care continuum for HIV-infected emergency department patients in an inner-city academic emergency department. *Ann Emerg Med*. 2015;66(1):69-78. doi:10.1016/j.annemergmed.2015.01.001

27. Enamorado Benjamin Fifield Kosuke Imai T. Using a Probabilistic Model to Assist Merging of Large-scale Administrative Records *. 2017.

28. Mohareb AM, Rothman RE, Hsieh Y-H. Emergency department (ED) utilization by HIV-infected ED patients in the United States in 2009 and 2010 – a national estimation. *HIV Med*. 2013;14:605-613. doi:10.1111/hiv.12052

29. Josephs J, Fleishman J, Korthuis P, Moore R, Gebo K. Emergency department utilization among HIV-infected patients in a multisite multistate study. *HIV Med*. 2010;11:74-84. doi:10.1111/j.1468-1293.2009.00748.x

30. Herwehe J, Wilbright W, Abrams A, et al. Implementation of an innovative, integrated electronic medical record (EMR) and public health information exchange for HIV/AIDS. *J Am Med Informatics Assoc*. 2012;19(3):448-452. doi:10.1136/amiajnl-2011-000412

31. Ridgway JP, Almirol E, Schmitt J, Wesley-Madgett L, Pitrak D. A Clinical Informatics Approach to Reengagement in HIV Care in the Emergency Department. *J Public Health Manag Pract*. 2019;25(3):270-273. doi:10.1097/PHH.0000000000000844

32. Barskey AE, Surendera Babu A, Hernandez A, Espinoza L. Patterns and trends of newly diagnosed HIV infections among adults and adolescents in correctional and noncorrectional facilities, United States, 2008–2011. *Am J Public Health*. 2016;106(1):103-109. doi:10.2105/AJPH.2015.302868

33. Spaulding AC, Seals RM, Page MJ, Brzozowski AK, Rhodes W, Hammett TM. HIV/AIDS among Inmates of and Releasees from US Correctional Facilities, 2006: Declining Share of Epidemic but Persistent Public Health Opportunity. Litvintseva AP, ed. *PLoS One*. 2009;4(11):e7558. doi:10.1371/journal.pone.0007558

34. Ammon B, Iroh P, Tiruneh Y, et al. HIV Care After Jail: Low Rates of Engagement in a Vulnerable Population. *J Urban Heal*. 2018;95(4):488-498. doi:10.1007/s11524-018-0231-0

35. Iroh PA, Mayo H, Nijhawan AE. The HIV care cascade before, during, and after Incarceration: a systematic review and data synthesis. *Am J Public Health*. 2015;105(7):e5-16. doi:10.2105/AJPH.2015.302635

36. Eastment MC, Toren KG, Strick L, Buskin SE, Golden MR, Dombrowski JC. Jail booking as an occasion for HIV care reengagement: a surveillance-based study. *Am J Public Health*. 2017;107(5):717-723. doi:10.2105/AJPH.2017.303668

37. Rich JD, DiClemente R, Levy J, et al. Correctional facilities as partners in reducing HIV disparities. *J Acquir Immune Defic Syndr*. 2013;63 Suppl 1(0 1):S49-53. doi:10.1097/QAI.0b013e318292fe4c

38. Westergaard RP, Kirk GD, Richesson DR, Galai N, Mehta SH. Incarceration predicts virologic failure for HIV-infected injection drug users receiving antiretroviral therapy. *Clin Infect Dis*. 2011;53(7):725-731. doi:10.1093/cid/cir491

39. Milloy M-J, Montaner JSG, Wood E. Incarceration of people living with HIV/AIDS: implications for treatment-as-prevention. *Curr HIV/AIDS Rep*. 2014;11(3):308-316. doi:10.1007/s11904-014-0214-z

40. Brewer RA, Magnus M, Kuo I, Wang L, Liu T-Y, Mayer KH. Exploring the Relationship Between Incarceration and HIV Among Black Men Who Have Sex With Men in the United States. *JAIDS J Acquir Immune Defic Syndr*. 2014;65(2):218-225. doi:10.1097/01.qai.0000434953.65620.3d

41. Brewer RA, Magnus M, Kuo I, Wang L, Liu T-Y, Mayer KH. The high prevalence of incarceration history among Black men who have sex with men in the United States: associations and implications. *Am J Public Health*. 2014;104(3):448-454. doi:10.2105/AJPH.2013.301786

42. Millett GA, Peterson JL, Wolitski RJ, Stall R. Greater risk for HIV infection of black men who have sex with men: a critical literature review. *Am J Public Health*. 2006;96(6):1007-1019. doi:10.2105/AJPH.2005.066720

43. Cunningham WE, Weiss RE, Nakazono T, et al. Effectiveness of a Peer Navigation Intervention to Sustain Viral Suppression Among HIV-Positive Men and Transgender Women Released From Jail. *JAMA Intern Med*. 2018;178(4):542. doi:10.1001/jamainternmed.2018.0150

44. Bailey S, Gilmore K, Yerkes L, Rhodes A. Connecting Corrections and HIV Care: Building a Care Coordination Program for Recently Incarcerated Persons Living with HIV in Virginia. *AIDS Behav*.

2019;23(S1):25-31. doi:10.1007/s10461-017-2003-3

45. Spaulding AC, Drobeniuc A, Frew PM, et al. Jail, an unappreciated medical home: Assessing the feasibility of a strengths-based case management intervention to improve the care retention of HIV-infected persons once released from jail. Lim S, ed. *PLoS One*. 2018;13(3):e0191643. doi:10.1371/journal.pone.0191643

46. Koester KA, Morewitz M, Pearson C, et al. Patient Navigation Facilitates Medical and Social Services Engagement Among HIV-Infected Individuals Leaving Jail and Returning to the Community. *AIDS Patient Care STDS*. 2014;28(2):82-90. doi:10.1089/apc.2013.0279

47. Fuller SM, Koester KA, Maiorana A, et al. "I don't have to do this all by myself": Systems Navigation to Ensure Continuity of HIV Care for Persons Leaving Prison. *AIDS Behav*. 2019;23(S1):14-24. doi:10.1007/s10461-018-2050-4

48. Centers for Disease Control and Prevention Division of HIV/AIDS Prevention. *Data to Care: Using HIV Surveillance Data to Support the HIV Care Continuum*.; 2017.

49. Ford MA, Spicer CM, Care C on RDS for MH, Medicine I of. The Role of Health Information Technology and Data System Integration in the Collection of HIV Care Data. 2012.

50. Michelle Ocampo JF, Smart J, Allston A, et al. Improving HIV Surveillance Data for Public Health Action in Washington, DC: A Novel Multiorganizational Data-Sharing Method. *JMIR Public Heal SurveillJMIR Public Heal Surveill*. 2016;22(111). doi:10.2196/publichealth.5317

51. Dombrowski JC. A Cluster Randomized Trial of a Population-Based HIV Surveillance-Guided Outreach Intervention to Increase Engagement in HIV Care and Antiretroviral Use. *Sex Transm Dis*. 2018;(In Press).

52. Dombrowski JC, Buskin SE, Bennett A, Thiede H, Golden MR. Use of Multiple Data Sources and Individual Case Investigation to Refine Surveillance-Based Estimates of the HIV Care Continuum. *J Acquir Immune Defic Syndr Novemb*. 2014;1(673):323-330. doi:10.1097/QAI.0000000000000302

53. Rapp RC, Ciomcia R, Zaller N, Draine J, Ferguson A, Cagey R. The Role of Jails in Engaging PLWHA in Care: From Jail to Community. *AIDS Behav*. 2013;17(S2):89-99. doi:10.1007/s10461-012-0298-7

54. Rajabiun S, Tryon J, Feaster M, et al. The Influence of Housing Status on the HIV Continuum of Care: Results From a Multisite Study of Patient Navigation Models to Build a Medical Home for People Living With HIV Experiencing Homelessness. *Am J Public Health*. 2018;108(S7):S539-S545. doi:10.2105/AJPH.2018.304736

55. Harawa NT, Brewer R, Buckman V, et al. HIV, Sexually Transmitted Infection, and Substance Use Continuum of Care Interventions Among Criminal Justice-Involved Black Men Who Have Sex With Men: A Systematic Review. *Am J Public Health*. 2018;108(S4):e1-e9. doi:10.2105/AJPH.2018.304698

56. Dusetzina SB, Tyree S, Meyer A-M, Meyer A, Green L, Carpenter WR. An Overview of Record Linkage Methods. 2014.

57. Loeliger KB, Meyer JP, Desai MM, Ciarleglio MM, Gallagher C, Altice FL. Retention in HIV care during the 3 years following release from incarceration: A cohort study. Tsai AC, ed. *PLOS Med*. 2018;15(10):e1002667. doi:10.1371/journal.pmed.1002667

58. Lim S, Nash D, Hollod L, Harris TG, Lennon MC, Thorpe LE. Influence of Jail Incarceration and Homelessness Patterns on Engagement in HIV Care and HIV Viral Suppression among New York City Adults Living with HIV/AIDS. Lima VD, ed. *PLoS One*. 2015;10(11):e0141912. doi:10.1371/journal.pone.0141912

59. Haley DF, Golin CE, Farel CE, et al. Multilevel challenges to engagement in HIV care after prison release: a theory-informed qualitative study comparing prisoners' perspectives before and after community reentry. *BMC Public Health*. 2014;14(1):1253. doi:10.1186/1471-2458-14-1253

60. - National Center for HIV C, Hepatitis V, Prevention T, of HIV D, Prevention A. Selected National HIV Prevention and Care Outcomes in the United States. 2017.

61. Incidence H, Surveillance Branch Centers for Disease Control C, of HIV D, Prevention A. Monitoring Selected National HIV Prevention and Care Objectives by Using HIV Surveillance Data Confidential information, referrals, and educational material on HIV infection and AIDS. 21:1-888.

62. Wiewel EW, Braunstein SL, Xia Q, Shepard CW, Torian L V. Monitoring Outcomes for Newly Diagnosed and Prevalent HIV Cases Using a Care Continuum Created With New York City Surveillance Data. *JAIDS J Acquir Immune Defic Syndr*. 2015;68(2):217-226. doi:10.1097/QAI.0000000000000424

63. Lesko CR, Sampson LA, Miller WC, et al. Measuring the HIV Care Continuum Using Public Health Surveillance Data in the United States. 2015;70(5):489-494.

64. Dombrowski JC, Carey JW, Pitts N, et al. HIV provider and patient perspectives on the Development of a Health Department "Data to Care" Program: a qualitative study. *BMC Public Health*. 2016;16(1):491. doi:10.1186/s12889-016-3152-4

65. Hood JE, Katz DA, Bennett AB, et al. Integrating HIV Surveillance and Field Services: Data Quality and Care Continuum in King County, Washington, 2010-2015. *Am J Public Health*. 2017;107(12):1938-1943. doi:10.2105/AJPH.2017.304069

66. Evans D, Van Gorder D, Morin SF, Steward WT, Gaffney S, Charlebois ED. Acceptance of the use of HIV surveillance data for care engagement: national and local community perspectives. *J Acquir Immune Defic Syndr*. 2015;69 Suppl 1(0 1):S31-6. doi:10.1097/QAI.0000000000000573

67. Sweeney P, Gardner LI, Buchacz K, et al. Shifting the paradigm: using HIV surveillance data as a foundation for improving HIV care and preventing HIV infection. *Milbank Q*. 2013;91(3):558-603. doi:10.1111/milq.12018

68. Buchacz K, Chen M-J, Parisi MK, et al. Using HIV Surveillance Registry Data to Re-Link Persons to Care: The RSVP Project in San Francisco. Sullivan PS, ed. *PLoS One*. 2015;10(3):e0118923. doi:10.1371/journal.pone.0118923

69. Avoundjian T, Lewis C, Peyton D, et al. Integrating HIV and STD surveillance to direct partner services and improve HIV care engagement in Jackson, Mississippi. In: *STD Prevention Conference*. Washington, D.C.; 2018.

70. Tesoriero JM, Johnson BL, Hart-Malloy R, et al. Improving Retention in HIV Care Through New York's Expanded Partner Services Data-to-Care Pilot. *J Public Health Manag Pract*. 2017;23(3):255-263. doi:10.1097/PHH.0000000000000483

71. Fellegi IP, Sunter AB. A theory for record linkage. *J Am Stat Assoc*. 1969;64(328):1183.

doi:10.2307/2286061

72. Sadinle M. Bayesian Estimation of Bipartite Matchings for Record Linkage. *J Am Stat Assoc*. 2017;112(518):600-612. doi:10.1080/01621459.2016.1148612

73. Bosh KA, Coyle JR, Muriithi NW, et al. Linking HIV and Viral Hepatitis Surveillance Data: Evaluating a Standard, Deterministic Matching Algorithm Using Data From 6 US Health Jurisdictions. *Am J Epidemiol*. August 2018. doi:10.1093/aje/kwy161

74. Ocampo JMF, Smart JC, Allston A, et al. Improving HIV Surveillance Data for Public Health Action in Washington, DC: A Novel Multiorganizational Data-Sharing Method. *JMIR public Heal Surveill*. 2016;2(1):e3. doi:10.2196/publichealth.5317

75. Tran K-N, Vatsalan D, Christen P. GeCo - An online personal data Generator and Corruptor. In: *Conference on Information and Knowledge Management*. San Francisco, CA; 2013. doi:10.1145/2505515.2508207

76. Bohensky MA, Jolley D, Sundararajan V, et al. Data linkage: a powerful research tool with potential problems. *BMC Health Serv Res*. 2010;10:346. doi:10.1186/1472-6963-10-346

77. Zingmond DS, Ye Z, Ettner SL, Liu H. Linking hospital discharge and death records—accuracy and sources of bias. *J Clin Epidemiol*. 2004;57(1):21-29. doi:10.1016/S0895-4356(03)00250-6

78. Shaw C, Atkinson J, Blakely T. (Mis)classification of ethnicity on the New Zealand Cancer Registry: 1981-2004. *N Z Med J*. 2009;122(1294):10-22.

79. Lariscy JT. Differential record linkage by Hispanic ethnicity and age in linked mortality studies: implications for the epidemiologic paradox. *J Aging Health*. 2011;23(8):1263-1284. doi:10.1177/0898264311421369

80. Silverman N. *Package "MakeProject."*; 2015.