Statistical Methods for Addressing Missing Data in HIV/AIDS Surveillance Systems
Secondary Analysis of the HPTN 065 Trial

Sahar Zangeneh, PhD

Vaccine and Infectious Disease Division
Fred Hutchinson Cancer Research Center
Seattle WA, U.S.A

May 17, 2018
Outline

1. HPTN 065 Study and National HIV Surveillance
2. Missing Data in HIV Surveillance
3. Conclusion and Future work
Outline

1. HPTN 065 Study and National HIV Surveillance
2. Missing Data in HIV Surveillance
3. Conclusion and Future work
HPTN 065 Study

Assess feasibility of a community-focused enhanced test and link-to-care strategy

- Improve performance of the HIV care cascade
- Community randomized trial with randomization at care facilities
- Used surveillance data for the design and analysis
HPTN 065 Study – ctd

Intervention cities: Bronx NY and Washington DC (38 care facilities total)

- **Control Groups:** Enhanced community outreach programs (ECOP)
- **Treatment Groups:** ECOP plus financial incentives

Non-intervention cities: Chicago, Philadelphia, Miami & Houston (ECOP underway)
National HIV Surveillance System (NHSS)

Collect and analyze data on all persons living with HIV/AIDS (PLWHA) in the U.S.
- Monitor resources nationally and locally
- Improve program implementation in PLWHA

Features
- Data available in aggregate form for HPTN 065
- Unified reporting systems
- Dynamic data extraction: quarterly data uploads
Analysis objectives

Monitor performance of the HIV care cascade

Use surveillance data to estimate

- Proportion of individuals \textit{linked to HIV care} (LC)
- Proportion of individuals \textit{virally suppressed} (VLS)
Monitoring measured VLS in Philadelphia, PA.

Percentage of HIV+ People with Viral Suppression, by Data Upload Quarter: PA

Data Period
- 2010Q1
- 2011Q1
- 2012Q1
- 2013Q1

Data Upload Month (Uploaded Quarterly)
Monitoring measured VLS in Washington, DC.
Outline

1. HPTN 065 Study and National HIV Surveillance
2. Missing Data in HIV Surveillance
3. Conclusion and Future work
Missingness in surveillance data

Adherence
- Inferred VL suppression status

Data quality and surveillance coverage
- Reporting lag
- Administrative missingness (common issue in EMR data)
- Lost specimens and/or records

1. **Identify presence** of bias
2. **Correct** biases in estimates when present
Change point estimation (Tapsoba et al, 2018)

Identify data stabilization using observed aggregated data

- Dependence between observation at different time points
- Develop methods for inference

Predict value of estimand after the change-point, e.g. the “stabilized” value
Bias-corrected value using change points

<table>
<thead>
<tr>
<th>City</th>
<th>CP(q)</th>
<th>LC_{2011} (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td>3.00</td>
<td>82.13</td>
</tr>
<tr>
<td>Miami</td>
<td>3.20</td>
<td>53.54</td>
</tr>
<tr>
<td>Chicago</td>
<td>7.93</td>
<td>81.44</td>
</tr>
<tr>
<td>NY</td>
<td>3.74</td>
<td>89.68</td>
</tr>
</tbody>
</table>

Washington DC

Quarters since first data report

Linkage to care in 3 months (%)

Miami

Quarters since first data report

Linkage to care in 3 months (%)

Chicago

Quarters since first data report

Linkage to care in 3 months (%)

New York

Quarters since first data report

Linkage to care in 3 months (%)
Missingness due to lack of full coverage

Address ‘other’ sources of missingness

For each person, observe a sequence of lab measurements collected over time

\[ \hat{p}_{VLS,q} = \frac{\text{# of people virally suppressed during quarter } q}{\text{# of people in care during quarter } q} \]

Person \( i \) is in care during quarter \( q \)

- Lab results in at least two of the past five quarters
## Individual-level data

<table>
<thead>
<tr>
<th>q1</th>
<th>q2</th>
<th>q3</th>
<th>q4</th>
<th>q5</th>
<th>q6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### True Pop.

True Pop.

- $y_{i1}$
- $y_{i2}$
- $y_{i3}$
- $y_{i4}$
- $y_{i5}$
- $y_{i6}$
- $y_{i7}$

### Update 1

Update 1

- $y_{i11}$
- $y_{i21}$

### Update 2

Update 2

- $y_{i12}$
- $y_{i22}$
- $y_{i31}$

### Update 3

Update 3

- $y_{i13}$
- $y_{i23}$
- $y_{i32}$
- $y_{i33}$
- $y_{i53}$

**Unsuppressed**

**Suppressed**
Missing data in measurements

For individuals in care, HPTN 065 defined VL status as:

\[
VL_{iq} = \begin{cases} 
1, & \text{if } VL_{iq} \text{ observed and } VL_{iq} < 400 \\
1, & \text{if } VL_{iq} \text{ unobserved but } VL_{i(q-1)} < 400 \\
0, & \text{if } VL_{iq} \text{ observed and } VL_{iq} > 400 \\
0, & \text{if } VL_{iq} \text{ unobserved and } VL_{i(q-1)} > 400 \\
0, & \text{otherwise}
\end{cases}
\]  

Assume all missingness to be driven by non-adherence.
Missing data in measurements

For individuals in care, we propose

$$\text{VL status}_{iq} = \begin{cases} 
1, & \text{if } VL_{iq} \text{ observed and } VL_{iq} < 400 \\
1, & \text{if } VL_{iq} \text{ unobserved but } VL_{i(q-1)} < 400 \\
0, & \text{if } VL_{iq} \text{ observed and } VL_{iq} > 400 \\
0, & \text{if } VL_{iq} \text{ unobserved and } VL_{i(q-1)} > 400 \\
0, & \text{otherwise}
\end{cases}$$

(2)

Could be 0 or 1 with probability depending on coverage of lab measurements

Missingness driven by non-adherence or non-coverage
Simulation study

If we knew the true coverage of lab measurements in surveillance

Use auxiliary data to estimate coverage of lab measurements in surveillance.
Estimating surveillance coverage

Use sample survey data from the Medical Monitoring Project (MMP)

- Better coverage of lab measurements
- Restricted to 400 individuals in each jurisdiction

Need individual level data

1. Two-dimensional matching
2. Dual-system estimation to estimate coverage of labs

Use computer-simulated data, and request analysis using real data
Outline

1. HPTN 065 Study and National HIV Surveillance
2. Missing Data in HIV Surveillance
3. Conclusion and Future work
Summary

Data quality raises challenges, beyond what was anticipated during design stage.

Many issues leading to missingness.

Addressed two ways of mitigating bias from missing measurements.

Second project is ongoing:
- Developing software for estimating surveillance coverage.
- Developing method to quantify the uncertainty.
Acknowledgements

Fred Hutch
- Deborah Donnell
- Ying Chen
- Jean de Dieu Tapsoba
- CY Wang

DOH Philadelphia
- Kathleen Brady
- Tanner Nassau
- Antonios Mashas

DOH New York
- Qiang Xia
- Lucia Torian

This work was partially supported by the National Institutes of Health grant MH105857. The HIV Prevention Trials Network is funded by the National Institute of Allergy and Infectious Diseases (UM1AI068619, UM1AI068613, UM1AI1068617), with co-funding from the National Institute of Mental Health, and the National Institute on Drug Abuse, all components of the U.S. National Institutes of Health.