Evolution and Diversity of HIV in Southeastern Michigan

Erik M Volz
Department of Epidemiology, University of Michigan– Ann Arbor

April 20, 2011
The HIV epidemic in Southeastern Michigan

- Statewide there are an estimated 19,500 people living with HIV; 12,850 infections in the Detroit MSA
- Rates of infection remain stable at 13.2 per 100k per year (21.4 in Detroit)
- African Americans and Detroit residents have a higher disease burden (11.1×)
- 20% of infected individuals develop AIDS prior to diagnosis with HIV
HAART introduced in 1995
It is difficult to estimate incidence in the early epidemic
PHIs based on HAART

- Two public health interventions have been proposed based on the use of ARVs
- Viral loads become undetectable under ART, transmission is greatly reduced
  - Test and Treat (T&T): Begin ART immediately following diagnosis
    - Ecological studies have shown correlation of community viral loads and incidence 1. Montaner et al., Lancet 376 9740, 2010.
  - Pre-Exposure Prophylaxis (PrEP): Microbicide or pill taken daily
Our study...

- Motivation: Viral diversity and evolution sheds light on epidemiology of HIV that is hard to observe directly
  - Who infects who? When do transmissions occur? Transmission rates by stage of infection?
- It is difficult to estimate these parameters using conventional study designs
- Understanding these hard-to-observe factors is important for designing good PHIs
Our study...

- 2004-2010 Michigan Department of Community Health has collected HIV genetic sequences from newly diagnosed; the rate of sequencing has increased year over year. Collected for surveillance of drug resistance mutations and optimizing ARV treatment regimens.
- Behavioral, demographic, and clinical covariates of the patients from whom the virus was isolated are linked to each sequence.
- Correlation of patient-level covariates with phylogenetic relationships of the virus.
- All data used in this study are secondary use and deidentified; would be easy to reproduce these analyses in other cities and for other at-risk populations.
Methodology: Initial focus on MSM in Detroit MSA

- Approx. half of new infections occur in those with confirmed MSM risk factor
- Approx. two thirds of new infections occur in Detroit and surrounding counties

Total: 3525 → MSM: 1881 → Detroit MSA: 1153
→ Tested for recent/acute infection: 711 → HIV-1 Subtype-B, high-quality: 612
Phylogenetic clusters (Who infects who?)

- Sets of closely related virus from infected hosts
- If virus from two hosts is very closely related, they are likely separated by a short chain of transmissions
- Clusters can be informative about transmission dynamics
  - Clustered hosts often have similar attributes; indicates population structure
A simple definition of a cluster

- Our goal is to characterize clustering patterns as a function of demographic & behavioral variables; statistical support for a cluster is of secondary interest
- Straightforward to illustrate the kind of inference that can be drawn from clusters using a very simple definition:
  - Clades of size 2 ("cherries")
Clustering and race matrix

- Size indicates strength of phylogenetic linkage
- Color indicates deviation from null
- Sexual assortativity by race compounds differential disease burden (Morris et al., *AJPH*, 699, 2009)
Clustering and age

The diagram shows the distribution of clustering across different age groups and years. The color scale indicates the value of \((A-E(A))/E(A)\), which is a measure of deviation from the expected clustering. The colors range from yellow (low deviation) to dark red (high deviation), with intermediate shades indicating varying degrees of deviation. The specific values are not provided in the image but can be inferred from the color gradient. The grid represents the intersection of age groups (0 - 12 yrs, 13 - 19 yrs, 20 - 24 yrs, 25 - 29 yrs, 30 - 39 yrs, 40 - 49 yrs, 50 - 59 yrs, 60 and over) and years (12 yrs, 19 yrs, 24 yrs, 29 yrs, 39 yrs, 49 yrs, 59 yrs, over).
To what extent does clustering patterns reflect sexual network structure?

We can compare the clustering patterns with behavioral surveillance data. National Health and Behavioral Surveillance (NHBS); $n = 519, 158$ partnerships reported, MSM, 2008

- Sample was only collected from Detroit
- Different age composition, and most of the sample was uninfected
Clustering and age: behavioral surveillance

![Heatmap showing clustering and age relations](image-url)
Clustering and age: behavioral surveillance

The diagram shows the clustering of individuals across different age groups. Each cell represents a comparison between two age groups, with the color indicating the difference in clustering. The color scale on the right ranges from -2.0 to 0.8, with darker colors representing a greater difference.
Comparison to survey data

Newman’s coefficient of assortativity:

\[ r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i} \]

\[ r_{race} = 38.2\% \]
\[ r_{county} = 6.9\% \]
\[ r_{age} = 16.1\% \]

NHBS \( r_{age} = 6.5\% \)

- Doherty et al., *Sexually Transmitted Diseases*, 38 12, 2011
- \( n = 2055 \) individuals infected with syphilis
- By race: 38.2%
- By age: 11.9%
□ Muth et al. have conducted a meta-analysis of many social network surveys
□ Assortativity is almost always greater for race than age
□ We found relative high assortativity by both race and age

<table>
<thead>
<tr>
<th>Study</th>
<th>Degree</th>
<th>Age</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antiviral</td>
<td>--</td>
<td>0.08</td>
<td>0.48</td>
</tr>
<tr>
<td>Baltimore</td>
<td>0.08</td>
<td>0.11</td>
<td>0.34</td>
</tr>
<tr>
<td>Bushwick</td>
<td>0.08</td>
<td>0.11</td>
<td>0.62</td>
</tr>
<tr>
<td>Chlamydia</td>
<td>0.13</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td>Flagstaff</td>
<td>-0.03</td>
<td>0.18</td>
<td>0.35</td>
</tr>
<tr>
<td>GC1981</td>
<td>0.03</td>
<td>0.14</td>
<td>0.38</td>
</tr>
<tr>
<td>HIV</td>
<td>0.11</td>
<td>0.14</td>
<td>0.27</td>
</tr>
<tr>
<td>Houston</td>
<td>0.15</td>
<td>0.10</td>
<td>0.73</td>
</tr>
<tr>
<td>Manitoba</td>
<td>0.18</td>
<td>0.20</td>
<td>--</td>
</tr>
<tr>
<td>PPNG</td>
<td>0.16</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td>Project90</td>
<td>0.03</td>
<td>0.10</td>
<td>0.18</td>
</tr>
<tr>
<td>Rockdale</td>
<td>-0.11</td>
<td>-0.08</td>
<td>-0.43</td>
</tr>
<tr>
<td>Syph318</td>
<td>0.03</td>
<td>-0.08</td>
<td>--</td>
</tr>
<tr>
<td>Urban</td>
<td>0.05</td>
<td>0.07</td>
<td>0.02</td>
</tr>
<tr>
<td>Urban2</td>
<td>-0.08</td>
<td>0.07</td>
<td>0.10</td>
</tr>
<tr>
<td>mean</td>
<td>0.06</td>
<td>0.10</td>
<td>0.25 (0.0986) (0.25432)</td>
</tr>
</tbody>
</table>
Clustering: future work

- We have mostly examined demographic variables; behavioral and clinical variables are also important!
- We also observe strong clustering by sexual risk behavior (MSM are strongly clustered)
- And we observe clustering by markers of early/acute infections (CD4, BED, viral diversity within host)
Estimating transmission rates
Building a model for HIV in SE Michigan

\[ S \rightarrow \text{early/acute infection} \rightarrow \text{chronic infection} \rightarrow \text{AIDS/death} \]
\[ \text{chronic infection} \rightarrow \text{diagnosed/ART} \]

- High transmissibility during early/acute infection
- Heterogeneity of risk behavior between individuals and over time
- Population growth and decline; immigration
- Variable duration of risk behavior
- Introduction HAART in mid-90s
- Estimation used a combination of genetic data and reported incidence
  - MDCH data indicate prevalence in 2010 of 4655 with confirmed MSM risk factor
  - An additional 2030 men are infected with unknown or “presumed heterosexual” risk factors
Estimating transmission rates: results

- We estimate prevalence and incidence over the entire course of the epidemic.
- Estimates based only on genetic data and MDCH estimates for 2010.
- What this model gets right: Rapid rise during the 80s, plateau during the 90s, and gradual rise with HAART.
- What it gets wrong: peak incidence is 81-82; mostly likely occurs 83-84; probably over estimates prevalence ~ 1985.
# Parameter estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median</th>
<th>5pc</th>
<th>95pc</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1^{-1}$ (days)</td>
<td>32.05</td>
<td>30.25</td>
<td>36.50</td>
</tr>
<tr>
<td>$\beta_2^{-1}$ (days)</td>
<td>1373.61</td>
<td>1000.42</td>
<td>1896.85</td>
</tr>
<tr>
<td>Fraction early/acute transmissions</td>
<td>0.56</td>
<td>0.47</td>
<td>0.65</td>
</tr>
<tr>
<td>Start of epidemic (years)</td>
<td>31.70</td>
<td>29.70</td>
<td>32.43</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>7.36</td>
<td>4.39</td>
<td>11.77</td>
</tr>
<tr>
<td>$R_0$</td>
<td>3.31</td>
<td>2.74</td>
<td>3.99</td>
</tr>
</tbody>
</table>

**Acute-stage transmission rate is** $\approx 43 \times$ **chronic-stage transmission**
Using genetic data increases precision

- Inference was based on both genetic data and independent estimate of prevalence in 2010.
- What is the relative contribution of the genetic data and feature matching to the precision of parameter estimates?

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median (FM)</th>
<th>Median (FM+G)</th>
<th>Design Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1^{-1}$ (days)</td>
<td>38.04</td>
<td>32.05</td>
<td>0.47</td>
</tr>
<tr>
<td>$\beta_2^{-1}$ (days)</td>
<td>1726.64</td>
<td>1373.61</td>
<td>0.75</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>8.24</td>
<td>7.36</td>
<td>0.92</td>
</tr>
<tr>
<td>$R_0$</td>
<td>2.76</td>
<td>3.31</td>
<td>0.83</td>
</tr>
</tbody>
</table>

- Increase in precision is greatest for transmission rates.
Predicted prevalence and incidence

- Fitting nonlinear dynamical models makes it easy to extrapolate prevalence over next decade
- \( \approx 2800 \) new infections in MSM over the next 10 years; incidence stable
Fitting nonlinear dynamical models also makes it easy to explore potential interventions.

Test and Treat, optimistic scenario: 80% diagnosed within 2 years, immediately placed on ART, zero transmission while on ART.

≈ 1500 infections in MSM over next 10 years

1300 infections averted
Predicted prevalence and PrEP

- PrEP, optimistic scenario: transmission rates reduced by 40%
- Only around 300 new infections over 10 years, 2500 averted
- Still not enough to get prevalence to fall
Conclusions

- The HIV epidemic remains a major threat to public health in the Detroit MSA.
- Measurements of the genetic diversity of virus are helpful addition to epidemiological surveillance.
- Phylogenetic clustering of the Detroit MSA sequences indicates strong assortativity by race, age and location. Future modeling work may estimate the rates of transmission within and between these populations.
- Initial modeling work suggests that T&T and PrEP can greatly reduce incidence.
- HIV sequences are now widely available. These analyses are easy to replicate for other at-risk populations.
Thanks

- James S. Koopman, University of Michigan
- Ethan Romero Severson, University of Michigan
- Eve Mokotoff, MDCH
- Mary-Grace Brandt, MDCH
- Emily Higgins, MDCH
- Rieza Soelaeman, MDCH
- Simon Frost, University of Cambridge
- Andrew Leigh Brown, University of Edinburgh
- Melissa J Ward, University of Edinburgh