The limit of sexual networks data: Implications for mathematical modelling of STI

Marie-Claude Boily
Department of Infectious Disease Epidemiology

Thanks to Roel Bakker for network picture
We know

Network structure is important

On disease transmission:

• Individual level
  • risk of infection
  • risk of transmission

• Population level
  • establishment
  • persistence
• Mean, Variance, Core group

• Mixing by sexual activity, age, race, etc

• Duration

• Interval between partnerships

• Concurrency
  Koumans et al (2001)

• Local centrality, Global centrality
• Mean, Variance, Core group

Critical level of sexual activity

\[ R_0 = \frac{B_c D}{C_c} > 1 \]

\[ C_c = \frac{1}{(B_c)} \]

Minimum number of partners \( (C_c) \) vs. Duration of infectiousness \( (D) \) (months)

- Infectivity = 10%
- Infectivity = 30%
- Infectivity = 50%
- Infectivity = 80%
• Mean, Variance, Core group

Garnett (2002), STI
• Mean, Variance, Core group

• Mixing by sexual activity, age, race, etc
• Mean, Variance, Core group

• Mixing by sexual activity, age,race, etc
Detailed micro-simulation/network models

• More intuitive
• More realistic?
  – Full network structure, sexual behaviour
  – Dynamics of infection

• Flexibility
  – Easy to introduce complexity
  – Wide range of research questions
    • Epidemiological and methodological
    • Individual and population level
Incomplete understanding of human behaviour

Network model

Individual rules

Behavioural process
- Partnership formation
- Concurrent partners
- Partnership duration
- Preferred partner type
- etc

Simulated Network

Validation

Sampled network

Incomplete network data

Impact of disease on network over time

Thanks to Roel Bakker for network pictures
Outline

• Do not know if network simulated representative of the REAL population network

1. Incomplete understanding of human behaviour

2. Limited data on sexual behaviour

3. Change in the network structure over time
   – due to the spread of disease itself
Defining behavioural process

- Do not know which set of individual rules determine real sexual population networks completely
  - Most obvious rules:
    - Observed behaviour
  - More subtle rules:
    - Intuition

→ Influence on network structure?
Simulation of Large Static Multi-Component Networks

900 simulated networks

Fixed number of links between activity classes:
- Fixed size = 100 000
- Fixed distribution in sexual activity (m^{1-15}=1.5, v=3)
- Fixed Mixing (Proportionate)
Between components

Number of components

300 simulated networks

Size of largest component

Partnership formation (L2)

- low-low
- random
- high-high

Partnership selection rule (L2)

Individual selection rule: Old
Size of largest component

300 simulated networks

![Graph showing the size of largest components](image)

- Individual selection rule: Old

Partnership selection rule (L2):
- low-low
- random
- high-high
Size of largest component
600 simulated networks

- Individual selection rule: Old
- Individual selection rule: New

Partnership selection rule (L2)
- low-low
- random
- high-high
Size of largest component
900 simulated networks

- Individual selection rule: Old
- Individual selection rule: New
- Individual selection rule: random

Proportionate mixing

Partnership selection rule (L2)
Implications

• Choice of rules/algorithm
  – Can biased simulated networks toward a region of networks with specific structural properties
  – Random variation small:
    • Difficult to create a large variety of network structures
    • Too restraint variety of network structures
      → Miss impact of risk factors
  – Random variation large:
    • Can obtain various network structures
    • Too restraint variety of network structures
      → Miss impact of risk factors
Outline

• Do not know if network simulated representative of the REAL population network

  1. Incomplete understanding of human behaviour

  2. Limited data on sexual behaviour

  3. Change in the network structure over time
     – due to the spread of disease itself
Behavourial data

† Empirical studies on sexual behaviour, sexual network

Key words

• Number of sexual partners
• Representative study (ies), survey(s), sexual behaviour, sexual activity level
• Number of sexual partners in general population, homosexual (heterosexual) population, representative sample
• Sexual activity in homosexual, heterosexual population
• Sexual mixing, mixing patterns
• (Empirical) sexual networks, sexual partnership (s), partner (s) network (s)
Individual based data

- Detailed representative sexual behaviour still limited for many population
  - Low risk population often under-documented
  - Unbalanced sexual activity between sexes when reported by men or female
    - Sampling bias
    - Misreporting bias
      - Culture specific

<table>
<thead>
<tr>
<th>Survey</th>
<th>12 months Unadjusted</th>
<th>12 months Adjusted</th>
<th>5 years Unadjusted</th>
<th>5 years Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSS, 1988–1991</td>
<td>1.74</td>
<td>1.19</td>
<td>2.44</td>
<td>1.01</td>
</tr>
<tr>
<td>NHSLS, 1992</td>
<td>1.47</td>
<td>0.98</td>
<td>2.12</td>
<td>0.91</td>
</tr>
</tbody>
</table>

– Behaviour crudely reported: categories

### Table 1  Sexual behaviour of the sampled Quebec and British populations

<table>
<thead>
<tr>
<th>No. of lifetime partners</th>
<th>Quebec</th>
<th>Britain*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men number</td>
<td>Women number</td>
</tr>
<tr>
<td></td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>Sexually inactive 0</td>
<td>94 (7.2)</td>
<td>121 (6.7)</td>
</tr>
<tr>
<td>Sexually active†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>212 (16.3)</td>
<td>472 (26.2)</td>
</tr>
<tr>
<td>2 to 4</td>
<td>343 (26.3)</td>
<td>616 (34.2)</td>
</tr>
<tr>
<td>5 to 9</td>
<td>283 (21.7)</td>
<td>329 (18.3)</td>
</tr>
<tr>
<td>10 to 19</td>
<td>196 (15.1)</td>
<td>187 (10.4)</td>
</tr>
<tr>
<td>20+</td>
<td>174 (13.4)</td>
<td>77 (4.3)</td>
</tr>
<tr>
<td>Total</td>
<td>1302 (100.0)</td>
<td>1802 (100.0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Categorical data (midpoint estimates)‡‡</th>
<th>Quebec</th>
<th>Britain*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Variance</td>
<td>Mean</td>
</tr>
<tr>
<td>10.8</td>
<td>162.6</td>
<td>6.2</td>
</tr>
</tbody>
</table>

| Lognormal distribution† | 14.8 | 1611.3 | 6.9 | 211.5 |
| Negative binomial**     | 9.4  | 119.6  | 5.3 | 35.4  |

| Lognormal distribution‡‡ | 17.8 | 6903.3 | 4.1 | 45.0 |
| Negative binomialff      | 7.4  | 70.8   | 2.4 | 7.9   |

NA = not available.

*British National Survey of Sexual Attitudes and Lifestyles (ACS'F Investigators); †Sexually active population = those declaring ≥ 1 lifetime sexual partners. ‡The midpoint mean (variance) estimates are based on the mid-interval value for the first five categories. ‡‡For the (20+) category of Quebec the mid-interval is fixed to the average number of annual partners (≥40 lifetime partners). ††For Britain, the mid-interval value of the (10+) category is fixed to 20. Best fit variables values; for the lognormal (mean = exp (μ+σ²/2), variance = mean x exp (σ²-1)); Quebec-μ₁ = 1.5; sd₁ = 1.5, μ₂ = 0.8, sd₂ = 1.4; Britain-μ₁ = 1.2, sd₁ = 1.8, μ₂ = 0.3, sd₂ = 1.3; for the negative binomial (mean = k(1-p)/p, variance = mean/p); Quebec-p₁ = 0.07, k₁ = 0.63, p₂ = 0.12, k₂ = 0.58; Britain-p₁ = 0.09, k₁ = 0.53; p₂ = 0.18, k₂ = 0.31.
• Sexual activity distribution and mixing does not define the network uniquely

$N_{\text{pop}} = 100,000$

Fixed DSA
$m = 1.5, s^2 = 3$

Good fit does not mean we have simulated the right population network
## Network data

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Objective</th>
<th>Sample</th>
<th>Location</th>
<th>Network characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alden and Klovelhal et al, 1992</td>
<td>1985</td>
<td>Study a network of HIV infected persons</td>
<td>44 homosexuals</td>
<td></td>
<td>egocentric, sociocentric</td>
</tr>
<tr>
<td>Auerbach, et al, 1984</td>
<td></td>
<td>Study cluster of HIV infection</td>
<td>44 homosexuals</td>
<td></td>
<td>egocentric sociocentric</td>
</tr>
<tr>
<td>Potterat et al, 1999</td>
<td>1988-1992</td>
<td>Exploration of change in structural change in 2 different population networks</td>
<td>•99 adolescents&lt;br&gt;•595 heterosexual at risk of HIV</td>
<td>•Atlanta, Georgia, US&lt;br&gt;•Colorado Spring</td>
<td>egocentric sociocentric</td>
</tr>
<tr>
<td>Friedman et al, 2000</td>
<td>1991-1993</td>
<td>To study reasons of HIV prevalence saturation in IDU</td>
<td>767 IDU over 30 days, 2 years</td>
<td>New-York, US</td>
<td>egocentric but mainly sociocentric</td>
</tr>
<tr>
<td>Day et al, 1998</td>
<td>1994-1996</td>
<td>Describe sexual networks of Gc patients from 2 UK departments of genitourinary medicine</td>
<td>•Heterosexual and homosexuals&lt;br&gt;•510 Gc cases+ 1228 contacts&lt;br&gt;•235 Gc cases+ 335 contacts&lt;br&gt;•over 3 months</td>
<td>•London, UK&lt;br&gt;•Sheffield, UK</td>
<td>egocentric sociocentric</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Objective</td>
<td>Sample</td>
<td>Location</td>
<td>Network characteristics</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Potterat et al, 1999</td>
<td>1996-1997</td>
<td>Identify individual and population-level determinants of Ct transmission</td>
<td>1309 Gc cases + 2409 contacts over 6 months</td>
<td>Colorado Springs, US</td>
<td>egocentric sociocentric</td>
</tr>
<tr>
<td>Wylie &amp; Jolly, 2001</td>
<td>1997-1998</td>
<td>To study patterns of Gc and Ct</td>
<td>4544 Gc or Ct male and female cases + contacts</td>
<td>Manitoba, Canada</td>
<td>sociocentric</td>
</tr>
<tr>
<td>Day et al, 1998</td>
<td>?</td>
<td>To “test” the use of genetic sero-typing data for the study of sexual networks</td>
<td>17 individuals over 6 months</td>
<td>Sheffield, Derby, Southampton, UK</td>
<td>limited sociocentric</td>
</tr>
<tr>
<td>Pickering et al, 1997</td>
<td>?</td>
<td></td>
<td>48 CSW + 38 clients over 12 weeks</td>
<td>Fishing villages, Uganda</td>
<td>egocentric</td>
</tr>
<tr>
<td>Johnson et al, 2003</td>
<td>1996-1997</td>
<td>Study relation between HIV prevalence and network size</td>
<td>Females in pre or post maternal care: 75 HIV+ 41partners &amp; 137 HIV- + 70 partners over 1 year, 5 years, lifetime</td>
<td>Lima, Peru</td>
<td>egocentric</td>
</tr>
<tr>
<td>Rothenberg et al, 2000</td>
<td></td>
<td>To study the network structure+ behaviour among persons at risk of HIV</td>
<td>over 6 months, 2 years</td>
<td>Atlanta, Georgia, US</td>
<td>egocentric sociocentric</td>
</tr>
</tbody>
</table>
Contact tracing data

- Sample of infected individuals + contacts
  - Completeness of data:
    - Diagnosis, sensitivity, number of contacts
  - Not necessarily representative of population networks
    - Initial sample: high-risk population
    - Following waves: not representative of populations of links
Sexual behaviour of gonorrhoea patients compared with ‘general’ US population

Garnett et al, 1999
Laumann et al, 1994
Changes in sample composition over time

- Contact tracing sample depends on the stage of the epidemic

True Mean = 1.5 partner
Example:
Estimation of mixing pattern based on snowball sampling

<table>
<thead>
<tr>
<th>Mixing Matrice elements</th>
<th>Female activity classes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>i= 1  2  3  4  5</td>
</tr>
<tr>
<td>Male activity classes</td>
<td>j= 1  2  3  4  5</td>
</tr>
<tr>
<td>P11 P12 P13 P14 P15</td>
<td>P21 P22 P23 P24 P25</td>
</tr>
<tr>
<td>P31 P32 P33 P34 P35</td>
<td>P41 P42 P43 P44 P45</td>
</tr>
<tr>
<td>P51 P52 P53 P54 P55</td>
<td>Q</td>
</tr>
</tbody>
</table>

Q=0 proportionate
Q>0 assortative
Q<0 disassortative
Bias on assortativity indice = (Estimated Q – True Q = 37)
Fig. 3. Errors in measures of network position in sampled networks. High values of p1-p2 indicate good estimates, whereas values close to or less than zero indicate poor estimates; error in measures of global centrality by snowball sampling continuously.

Implications

• Structural properties of network ill represented:
  – For unbiased estimates use sample “strategically”

  • Distribution in sexual activity: index cases
  • Mixing: 1 cycle, random selection of links
  • Measures of centrality (closeness, information centrality, etc): long chains (Ghani et al (2000))
Outline

• Do not know if network simulated representative of the REAL population network

1. Incomplete understanding of human behaviour

2. Limited data on sexual behaviour

3. Change in the network structure over time
Changes over time

• Available behavioural or Network data
  – One point in time (Cross-sectional studies)

• Changes in behaviour and network structure over time
  – Even in absence of prevention
  – For lethal infectious diseases
    • Differential mortality (e.g. AIDS)
• Blower et al (1993)

Closed cohort of MSM in Amsterdam over 7 years
- 21% reduction in mean number of partners (m)
- 33% reduction in variance (v)
- 29% reduction in effective mean rate (c = m + v/m)

Variance in sexual activity

Initial sample
Expected if no individual change
Differential Mortality = -33%
Individual Change
Total Change
Final sample
Example:
Impact of wide-scale use of antiretroviral therapy on STI

Sexual behaviour and STI

Effect of ART at the individual level

– Prolongs survival of AIDS patients + improve quality of life
– Slows disease progression
– Prolong incubation period
– Potentially reduce infectivity of HIV+ treated

→ Differential replenishment

Boily et al., STD (in press)
STI vs sexual behaviour
Over 10 years after ART
HIV vs STI
Over 10 years after ART
STD clinic: gonorrhoea 1981 - 2001
homosexual men in Amsterdam
Unprotected anal sex (%) young (<35 years) homosexual men in cohort

Time periods:
- October ‘84 - April ‘87
- May ‘87 - February ‘92
- March ’92 - June ’96
- July ’96 - December ’99

Reported at cohort visits
Implications (2)

- Interpret current trends
- Prevention
  - Differential AIDS mortality:
    - Overestimate impact of prevention at the beginning of the HIV epidemic
  - Differential ART replenishment:
    - Overestimate relapse to risky sex after treatment availability
  - Developing countries largely afflicted by HIV/AIDS where ART as yet to become widely available
    - Increase in STI and sexual behaviour expected
- Important considerations
  - Impact of population-level changes on individual behaviour
  - Risk of trying altering the network as preventive measures
Impact of AIDS mortality on individual behaviour and network structure

Goes back to uncertainty in behavioural rules
Discussion
Model limitations - Causes for concerns?

- Do not know if network simulated representative of the population network
- Does it really matter?
  - If presuppose network are important → yes
  - More complex model ≠ Necessarily better → False sense of security
- Assessment of the impact of network characteristics on disease transmission depends on the characteristic of network simulated
  - Too restraints variety of network structures
    → Miss impact of risk factors
  - Too large variety of network structures
    → Find non relevant risk factors
- Unfortunately with most complex network models:
  - Behavioural process not always thoroughly described
    → Non Reproducible
  - A description of the structure of network not always presented
    → Non Comparable
• Special thanks to Martina Morris and Claudia Neuhauser

• Robert Poulin (ABB)
• Roel Bakker (Erasmus University)
• Azra Ghani (Imperial College)
• Geoff Garnett (Imperial College)
• Anu Gupta (MHRC)
End