A Bayesian synthesis of evidence for a dynamic transmission model: estimating HIV incidence among MSM

Anne Presanis\textsuperscript{1}, Daniela De Angelis\textsuperscript{2,1}

\textsuperscript{1}MRC Biostatistics Unit, Cambridge
\textsuperscript{2}Statistics, Modelling and Bioinformatics Unit, Health Protection Agency Centre for Infections, London

28th August 2009
Greek Stochastics, Lefkada
Outline

1. Introduction
2. Prevalence model
3. Incidence model
4. Two-stage linear model
5. Two-stage transmission model
6. Combined prevalence & incidence model
7. Ongoing work
Motivation

Human Immunodeficiency Virus

Estimates of HIV prevalence and incidence are essential for understanding and monitoring the epidemic, as well as for assessing the impact of public health interventions.

Challenges

- HIV has a long asymptomatic incubation period, so many infections undiagnosed
- Surveillance systems available only for certain risk groups and populations
- Surveillance and other survey/ad-hoc data subject to biases
- Data sometimes tell us only indirectly about the quantities of interest
Estimating HIV prevalence

\[ N_r \sum_g \pi_{g,r} \rho_{g,r} \]

- **Total population, region** \( r \)
- **Sum over risk groups**
- **HIV prevalence in group** \( g \), region \( r \)
- **Proportion of region** \( r \) **in group** \( g \)

\[ N_r \sum_g \delta_{g,r} \pi_{g,r} \rho_{g,r} \]

Diagnosed

“Known”

\[ N_r \sum_g (1 - \delta_{g,r}) \pi_{g,r} \rho_{g,r} \]

Undiagnosed

Unknown, to be estimated
Estimate for each $g$ and $r$: 

$\rho_{g,r}$ proportion of the population of $r$ in $g$;

$\pi_{g,r}$ HIV prevalence;

$\delta_{g,r}$ proportion of infections which are diagnosed.

Then any function of these may be estimated.
Groups are further sub-divided by sex and current/past risk behaviour:

- **MSM**: Men who have sex with men
- **IDUs**: Injecting drug users (non-MSM)
- **SSA-born**: Heterosexual individuals born in Sub-Saharan Africa (non-IDUs)
- **GUM**: Heterosexual individuals (non-SSA), current GUM attendees
- **LR**: Lower risk heterosexual individuals (non-SSA, non-GUM)
## Availability of data

<table>
<thead>
<tr>
<th>Risk group</th>
<th>$N$</th>
<th>$\rho$</th>
<th>$\pi$</th>
<th>$\delta$</th>
<th>$\pi(1 - \delta)$</th>
<th>$f(\rho, \pi, \delta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSM</td>
<td></td>
<td>NATSAL</td>
<td></td>
<td></td>
<td>UA GUM, GMSHS</td>
<td>SOPHID</td>
</tr>
<tr>
<td>IDUs</td>
<td></td>
<td>HOCR, SEA</td>
<td></td>
<td>UA IDU</td>
<td>UA IDU</td>
<td>SOPHID</td>
</tr>
<tr>
<td>SSA</td>
<td></td>
<td>Census, ONS births</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GUM</td>
<td></td>
<td>NATSAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>ALL</td>
<td>ONS MYEs</td>
<td></td>
<td></td>
<td></td>
<td>SOPHID</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDUs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SOPHID</td>
</tr>
<tr>
<td>SSA</td>
<td></td>
<td>HOCR, SEA</td>
<td></td>
<td>UA IDU</td>
<td>UA IDU</td>
<td>SOPHID</td>
</tr>
<tr>
<td>GUM</td>
<td></td>
<td>Census, ONS births</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>non-SSA</td>
<td>NATSAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>ALL</td>
<td>ONS MYEs</td>
<td></td>
<td></td>
<td></td>
<td>SOPHID</td>
</tr>
</tbody>
</table>
Multi-parameter evidence synthesis

Aim

Synthesise evidence from *all* available sources to estimate parameters of interest

- evidence from potentially *different types* of studies
- *direct* evidence on parameters of interest
- *indirect* evidence on complex functions of parameters

*Bayesian* setting:
- Coherent and correct *propagation of uncertainty*
- *Prior* information
- Ease of model formulation to account both for *complex* relationships between data sources and for *biases*
Directed Acyclic Graph

- \( \rho_{g,r} \)
- \( \pi_{g,r} \)
- \( \delta_{g,r} \)
- \( N_r \)
- ONS

NATSAL, Census 2001
Risk group proportions

UA PW, IDU
Prevalence

UA GUM
Prevalence of undiagnosed infection among GMSM, GUMm, GUMf

NSHPC, UA PW, IDU
Proportion diagnosed

SOPHID
Risk group composition of diagnosed individuals

SOPHID
Total diagnosed

\( \rho_{g,r} \pi_{g,r} \delta_{g,r} \)

\( N_r \sum_g \rho_{g,r} \pi_{g,r} \delta_{g,r} \)

\( \pi_{g,r}(1 - \delta_{g,r}) \)
Inference

Priors
Unif(0,1) for basic parameters $\rho, \pi, \delta$

Likelihood
- Poisson counts for total men and women diagnosed
- Remaining data available as proportions $y/n$

$$y \sim \text{Binomial}(n, p)$$
$$p = f(\rho, \pi, \delta)$$

Hierarchy
We *borrow strength* across regions and risk groups in order to estimate parameters for $(g, r)$ combinations where there is a lack of data.
### Table: Posterior mean and median estimates from stage 1 model of proportion of MSM in each compartment, with 95% credible intervals.

<table>
<thead>
<tr>
<th>Year</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>Year</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$1 - \rho(t)$</td>
<td></td>
<td></td>
<td></td>
<td>$(1 - \delta(t))\pi(t)\rho(t)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>0.9611</td>
<td>0.9667</td>
<td>0.9718</td>
<td>2001</td>
<td>0.0004</td>
<td>0.0005</td>
<td>0.0007</td>
</tr>
<tr>
<td>2002</td>
<td>0.9611</td>
<td>0.9667</td>
<td>0.9716</td>
<td>2002</td>
<td>0.0005</td>
<td>0.0007</td>
<td>0.0011</td>
</tr>
<tr>
<td>2003</td>
<td>0.9610</td>
<td>0.9667</td>
<td>0.9720</td>
<td>2003</td>
<td>0.0005</td>
<td>0.0007</td>
<td>0.0010</td>
</tr>
<tr>
<td>2004</td>
<td>0.9609</td>
<td>0.9667</td>
<td>0.9722</td>
<td>2004</td>
<td>0.0005</td>
<td>0.0006</td>
<td>0.0009</td>
</tr>
<tr>
<td>2005</td>
<td>0.9607</td>
<td>0.9664</td>
<td>0.9713</td>
<td>2005</td>
<td>0.0004</td>
<td>0.0005</td>
<td>0.0007</td>
</tr>
<tr>
<td>2006</td>
<td>0.9608</td>
<td>0.9667</td>
<td>0.9719</td>
<td>2006</td>
<td>0.0005</td>
<td>0.0006</td>
<td>0.0008</td>
</tr>
<tr>
<td>2007</td>
<td>0.9608</td>
<td>0.9665</td>
<td>0.9718</td>
<td>2007</td>
<td>0.0004</td>
<td>0.0005</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
<th>Year</th>
<th>2.5%</th>
<th>median</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(1 - \pi(t))\rho(t)$</td>
<td></td>
<td></td>
<td></td>
<td>$\delta(t)\pi(t)\rho(t)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>0.0268</td>
<td>0.0318</td>
<td>0.0373</td>
<td>2001</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0009</td>
</tr>
<tr>
<td>2002</td>
<td>0.0268</td>
<td>0.0316</td>
<td>0.0371</td>
<td>2002</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0010</td>
</tr>
<tr>
<td>2003</td>
<td>0.0263</td>
<td>0.0316</td>
<td>0.0372</td>
<td>2003</td>
<td>0.0010</td>
<td>0.0010</td>
<td>0.0011</td>
</tr>
<tr>
<td>2004</td>
<td>0.0262</td>
<td>0.0315</td>
<td>0.0373</td>
<td>2004</td>
<td>0.0011</td>
<td>0.0011</td>
<td>0.0011</td>
</tr>
<tr>
<td>2005</td>
<td>0.0271</td>
<td>0.0319</td>
<td>0.0375</td>
<td>2005</td>
<td>0.0011</td>
<td>0.0012</td>
<td>0.0012</td>
</tr>
<tr>
<td>2006</td>
<td>0.0263</td>
<td>0.0315</td>
<td>0.0372</td>
<td>2006</td>
<td>0.0012</td>
<td>0.0013</td>
<td>0.0013</td>
</tr>
<tr>
<td>2007</td>
<td>0.0264</td>
<td>0.0316</td>
<td>0.0373</td>
<td>2007</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0014</td>
</tr>
</tbody>
</table>

### Prevalence model

Posterior estimates

### MSM

Presanis & De Angelis (MRC BSU) Bayesian HIV dynamic transmission model 28th August 2009
Incidence from prevalence

- Linear multi-state model, HIV incidence $\lambda(t)$ given flat prior

- Non-linear *dynamic transmission model*, $\lambda(t) = \beta(t)\pi(t)$, coefficient $\beta(t)$ is the *effective contact rate*. $\beta(t)$ may be further parameterised as the overall contact rate multiplied by the transmission rate following an infectious contact.
Literature on dynamic transmission models (see Isham (2005), Hethcote (2000) for reviews, Anderson & May (1991)):
- Deterministic vs Stochastic
- Individual vs population-level
- Scenario-based sensitivity analyses
- Cross-sectional prevalence data
- Discrete/Continuous state space
- Discrete/Continuous time

Estimating incidence from prevalence data

Our work:
- Full probability model
- Estimate both prevalence and incidence
- Bayesian framework allows correct and complete propagation of uncertainty in data sources through to posterior estimates of prevalence and incidence
Multi-state model

- \( N(t) = \) total number of men aged 15-44 in England and Wales at time \( t \)
- \( \rho(t) = s(t) + u(t) + d(t) = \) proportion of \( N(t) \) who are men who have sex with men (MSM): Susceptible, Undiagnosed and Diagnosed respectively
- \( 1 - \rho(t) = e(t) = \) proportion of \( N(t) \) who are not MSM
System of equations

\[
\begin{align*}
\frac{d}{dt} e(t) &= \frac{\alpha(t)}{N(t)} - (\mu_E(t) + \gamma_E(t) + \psi)e(t) - e(t)f(t) \\
\frac{d}{dt} s(t) &= \psi e(t) - (\lambda(t) + \mu_S(t) + \gamma_S(t))s(t) - s(t)f(t) \\
\frac{d}{dt} u(t) &= \lambda(t)s(t) - (\kappa(t) + \mu_U(t) + \gamma_U(t))u(t) - u(t)f(t) \\
\frac{d}{dt} d(t) &= \kappa(t)u(t) - (\mu_D(t) + \gamma_D(t))d(t) - d(t)f(t) \\
\end{align*}
\]

\[
f(t) = \left[ \frac{\alpha(t)}{N(t)} - (\mu_E(t) + \gamma_E(t))e(t) - (\mu_S(t) + \gamma_S(t))s(t) \\
- (\mu_U(t) + \gamma_U(t))u(t) - (\mu_D(t) + \gamma_D(t))d(t) \right]
\]
Two-stage linear model

Inference

Directed acyclic graph

Initial state of system

\[ c(t) \]

\[ \logit\{c(t)\} \]

\[ \text{MVN} \]

\[ \Sigma(t) \]

\[ y_c(t) \]

\[ c \in \{e, s, u, d\} \]

\[ t = 2001 \]

\[ \rho(t) \]

\[ = s(t) + u(t) + d(t) \]

\[ \pi(t) \]

\[ = \frac{u(t)+d(t)}{s(t)+u(t)+d(t)} \]

\[ \delta(t) \]

\[ = \frac{d(t)}{u(t)+d(t)} \]

\[ t \in 2002 \to 2007 \]

\[ \theta_P(t) \]

\[ \theta_L(t) \]

\[ \text{data} \]

\[ \text{MVN} \]

\[ \Sigma(t) \]

\[ y_c(t) \]

\[ c \in \{e, s, u, d\} \]

\[ t = 2001 \]

\[ \rho(t) \]

\[ = s(t) + u(t) + d(t) \]

\[ \pi(t) \]

\[ = \frac{u(t)+d(t)}{s(t)+u(t)+d(t)} \]

\[ \delta(t) \]

\[ = \frac{d(t)}{u(t)+d(t)} \]

\[ t \in 2002 \to 2007 \]
Likelihood

\[ L(c, \gamma, \mu_D, \mu_E, \kappa \mid y_c, \Sigma, y_{44}, N, y_{\mu_D}, D, y_{\text{deaths}}, y_{\text{diagnoses}}) \]

\[ \propto \prod_{t=1}^{7} \left\{ \exp \left( - \frac{1}{2} (y_c(t) - \logit(c(t)))^T \Sigma^{-1}(t) (y_c(t) - \logit(c(t))) \right) \right. \\
\times \left. \left( \gamma(t)^{y_{44}(t)} (1 - \gamma(t))^{T(t) - y_{44}(t)} \right) \right. \\
\times \left. \left( \gamma_D(t)^{y_{44}^D(t)} (1 - \gamma_D(t))^{D(t) - y_{44}(t)} \right) \right. \\
\times \left. \left( \mu_D(t)^{y_{\mu_D}(t)} (1 - \mu_D(t))^{D(t) - y_{\mu_D}(t)} \right) \right. \\
\times \left. \left( \mu(t)^{y_{\text{deaths}}(t)} (1 - \mu(t))^{T(t) - y_{\text{deaths}}(t)} \right) \right. \\
\times \left. \left( (\kappa(t)U(t))^{y_{\text{diagnoses}}} \exp(-\kappa(t)U(t)) \right) \right\} \]
Priors

\[ \psi \sim \text{Normal}(0.0049, 0.0006^2) \]
\[ \lambda(t) \sim \text{Beta}(1, 1) \]
\[ \kappa(t) \sim \text{Beta}(1, 1) \]
\[ c(2001) \sim \text{Dirichlet}(1, 1, 1, 1) \]

\[ \mu_E(t) \sim \text{Beta}(1, 1) \]
\[ \mu_D(t) \sim \text{Beta}(1, 1) \]
\[ \gamma_E(t) \sim \text{Beta}(1, 1) \]
Migration assumptions

Model 1: As described above - includes data on proportions $c(t)$, aging, mortality and diagnosis rates. Net migration assumed 0 in each state.

Model 2: As in 1, but allow for net outward migration of Diagnosed MSM (data from SOPHID - indistinguishable from loss to follow up).

Model 3: As in 2, but assume inward migration of men into $E$, outward migration from all 4 states (data from ONS).

Model 4: As in 3, but assume inward & outward migration occurs at equal rates in $E$, $S$ and $U$; still assume only net outward migration/LTFU in $D$.

Model 5: As in 4, but assume MSM in $S$ and $U$ have higher migration rates than men in $E$ (data from NATSAL).
Posterior incidence & diagnosis rates

Model 2a

Model 3a

Presanis & De Angelis (MRC BSU) Bayesian HIV dynamic transmission model 28th August 2009
Dynamic transmission modelling

Homogeneous mixing ("mass action")

\[ \lambda(t) = \beta(t)\pi(t) \]

The effective contact rate, \( \beta(t) \) is the average number of contacts per unit time a susceptible individual makes which would be sufficient for transmission if the contact is infectious. The assumption of homogeneous mixing is not particularly realistic for HIV and other STIs.

Different contact groups

\[ \lambda(t) = \chi_U(t)\tau_U(1 - \delta(t))\pi(t) + \chi_D(t)\tau_D\delta(t)\pi(t) \]

\( \chi_U(t) \) is the per-susceptible contact rate with undiagnosed individuals, \( \chi_D(t) \) with diagnosed individuals. The \( \tau \) are transmission rates given an infectious contact, also assumed to differ between diagnosed/undiagnosed contacts. Homogeneous mixing assumed within groups.
Further stratification

Homogeneous mixing within risk groups, no contact across groups

\[ \lambda_g(t) = \chi_g^U(t) \tau_U (1 - \delta_g(t)) \pi_g(t) + \chi_g^D(t) \tau_D \delta_g(t) \pi_g(t) \]

Mixing matrix describing contact across groups

\[ \lambda_g(t) = \sum_i \phi_{gi}^U(t) \chi_g^U(t) \tau_U (1 - \delta_i(t)) \pi_i(t) + \sum_i \phi_{gi}^D(t) \chi_g^D(t) \tau_D \delta_i(t) \pi_i(t) \]

The \( \phi \) are matrices of probabilities of an individual in group \( g \) choosing a contact from group \( i \), ranging from fully assortative to homogeneous mixing. The \( \chi \) still represent the rates of contact.
Priors

Different contact groups

\[ \lambda(t) = \chi_U(t)\tau_U(1 - \delta(t))\pi(t) + \chi_D(t)\tau_D\delta(t)\pi(t) \]

\[ \tau_U \sim \text{Unif}(0, 0.3) \]

\[ \tau_D \sim \text{Unif}(0, \tau_U) \]

\[ \chi_U(t) \sim \text{Gamma}(1, 4) \]

\[ \chi_D(t) \sim \text{Unif}(0, \chi_D(t)) \]
Posterior rates

Incidence

Effective contact rate

Contact rate

Transmission rate

Incidence

Effective contact rate

Contact rate

Transmission rate
Combined prevalence & incidence model

Model structure

Directed acyclic graph

Initial state of system

\[ \rho(t) = s(t) + u(t) + d(t) \]
\[ \pi(t) = \frac{u(t)+d(t)}{s(t)+u(t)+d(t)} \]
\[ \delta(t) = \frac{d(t)}{u(t)+d(t)} \]

\( c \in \{e, s, u, d\} \)

\( t = 2001 \)

\( t \in 2002 : 2007 \)

\( \theta_P(t) \)

\( \theta_L(t) \)

\( c(t) \)

\( \text{data} \)

\( \text{data} \)
Posterior estimates
Ongoing work

- Combined prevalence, incidence & transmission model.
- Incorporate data on resistance to inform transmission from Diagnosed individuals.
- Split MSM into further risk groups by current/past practice, STI clinic attendance and age, incorporate contact/transmission between the groups.
- Expand to other risk groups.
Acknowledgements

- Tony Ades (Bristol)
- Aicha Goubar (INVS, Paris)
- David Spiegelhalter (MRC Biostatistics Unit & University of Cambridge)
- Graham Medley (Warwick)
- Tim Chadborn, Valerie Delpech, Caterina Hill, Viv Hope, Louise Logan, Brian Rice and Noel Gill (Health Protection Agency HIV Department)
References