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

Anne Presanis[1], Daniela De Angelis[2,1]

[1]MRC Biostatistics Unit, Cambridge

[2]Statistics, Modelling and Bioinformatics Unit, Health Protection Agency Centre for Infections, London

28th August 2009

Greek Stochastics, Lefkada

Outline

MRC Biostatistics Unit

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

MRC Biostatistics Unit

Human Immunodeficiency Virus

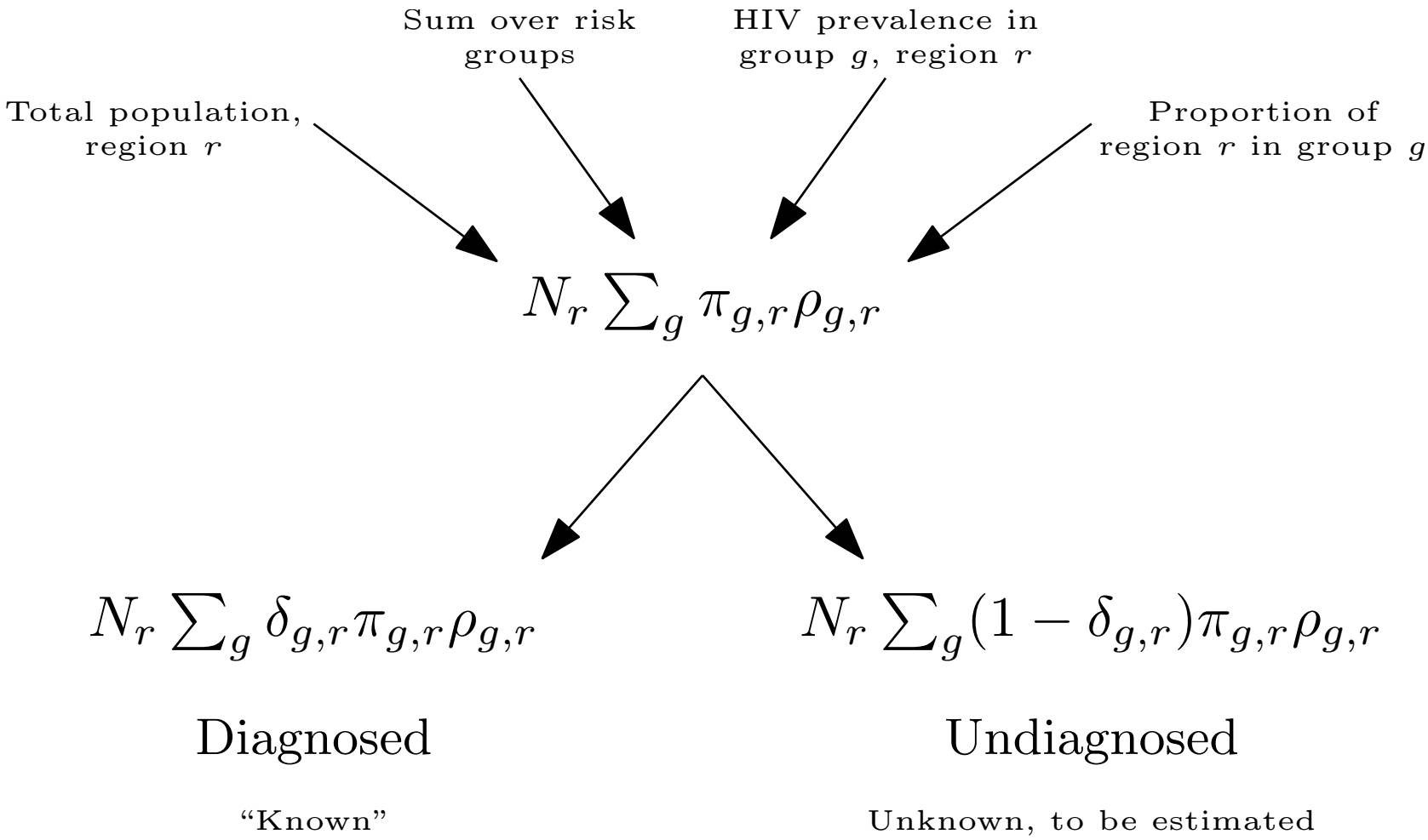
Estimates of HIV *prevalence* and *incidence* are essential for *understanding* and *monitoring* the epidemic, as well as for *assessing* the impact of public health interventions.

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

MRC Biostatistics Unit



Estimate for each g and r:

MRC | Biostatistics Unit

$\rho_{g,r}$ 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.

Risk groups in England & Wales

MRC | Biostatistics Unit

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

MRC Biostatistics Unit

Risk group		N	ρ	π	δ	$\pi(1-\delta)$	$f(\rho, \pi, \delta)$
Men	MSM		NATSAL			UA GUM, GMSHS	SOPHID
	IDUs		HOCR, SEA	UA IDU	UA IDU		SOPHID
	SSA		Census, ONS births				SOPHID
	GUM		NATSAL			UA GUM	
	LR						
	ALL	ONS MYEs					SOPHID
Women	IDUs		HOCR, SEA	UA IDU	UA IDU		SOPHID
	SSA		Census, ONS births	UA PW	NSHPC, UA PW		SOPHID
	GUM		NATSAL			UA GUM	
	LR						
	non-SSA			UA PW	NSHPC, UA PW		
	ALL	ONS MYEs					SOPHID

Multi-parameter evidence synthesis

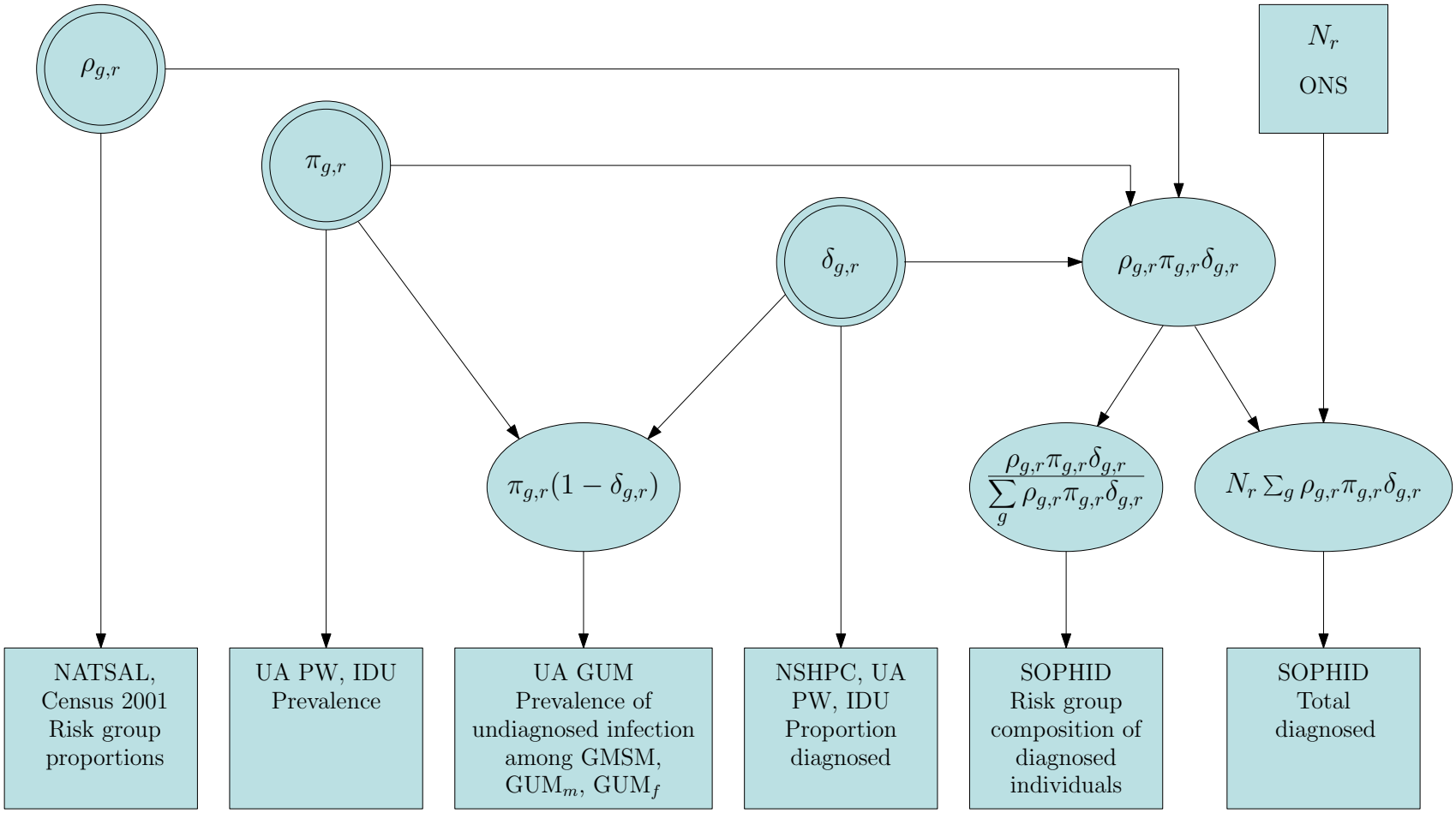
MRC | Biostatistics Unit

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



Inference

MRC Biostatistics Unit

Priors

Unif(0,1) for basic parameters ρ, π, δ

Likelihood

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

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

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.

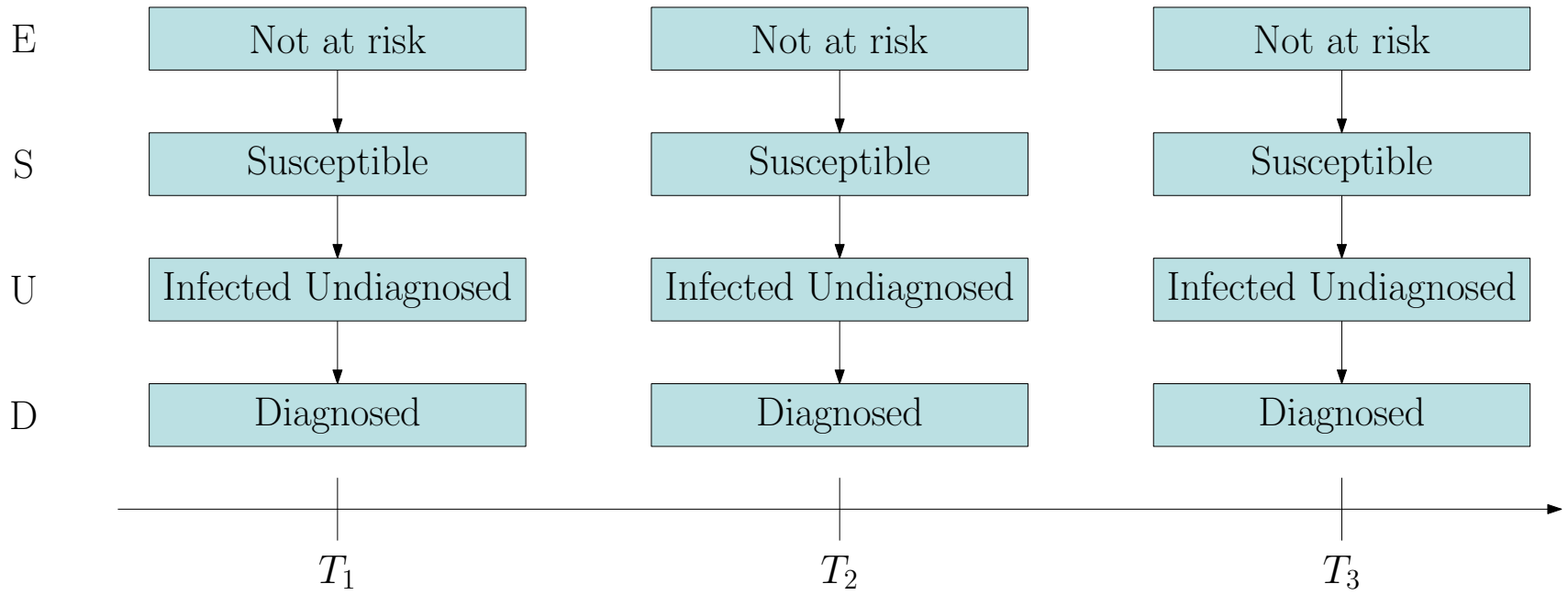
MSM

MRC Biostatistics Unit

Year	2.5%	median	97.5%	Year	2.5%	median	97.5%
	$1-\rho(t)$				$(1-\delta(t))\pi(t)\rho(t)$		
2001	0.9611	0.9667	0.9718	2001	0.0004	0.0005	0.0007
2002	0.9611	0.9667	0.9716	2002	0.0005	0.0007	0.0011
2003	0.9610	0.9667	0.9720	2003	0.0005	0.0007	0.0010
2004	0.9609	0.9667	0.9722	2004	0.0005	0.0006	0.0009
2005	0.9607	0.9664	0.9713	2005	0.0004	0.0005	0.0007
2006	0.9608	0.9667	0.9719	2006	0.0005	0.0006	0.0008
2007	0.9608	0.9665	0.9718	2007	0.0004	0.0005	0.0007
	$(1-\pi(t))\rho(t)$				$\delta(t)\pi(t)\rho(t)$		
2001	0.0268	0.0318	0.0373	2001	0.0009	0.0009	0.0009
2002	0.0268	0.0316	0.0371	2002	0.0009	0.0009	0.0010
2003	0.0263	0.0316	0.0372	2003	0.0010	0.0010	0.0011
2004	0.0262	0.0315	0.0373	2004	0.0011	0.0011	0.0011
2005	0.0271	0.0319	0.0375	2005	0.0011	0.0012	0.0012
2006	0.0263	0.0315	0.0372	2006	0.0012	0.0013	0.0013
2007	0.0264	0.0316	0.0373	2007	0.0013	0.0013	0.0014

Table: Posterior mean and median estimates from stage 1 model of proportion of MSM in each compartment, with 95% credible intervals.

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.

Context

MRC Biostatistics Unit

- 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*
  - e.g. Ades & Medley (1994), Gregson et al (1996), Hallett et al (2008)

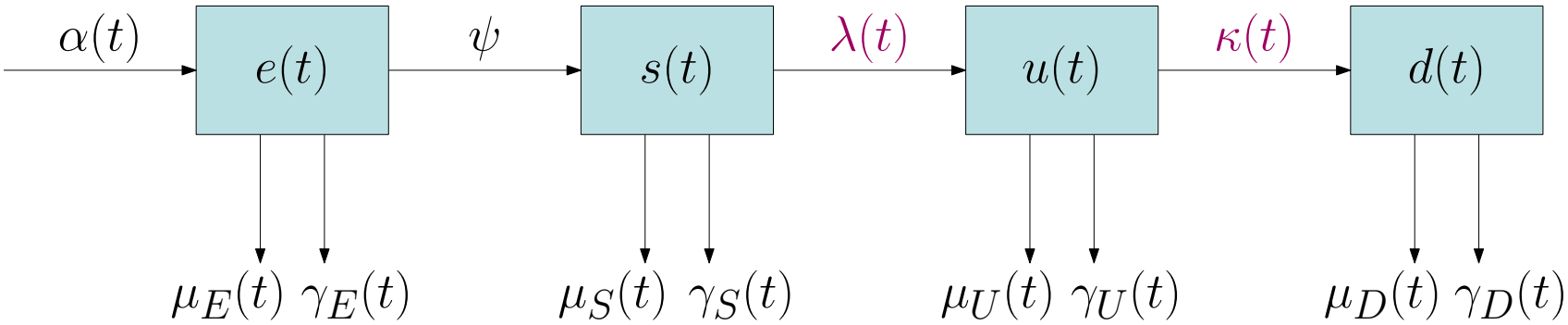
Our work:

- 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

MRC Biostatistics Unit

- $N(t)$ = 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

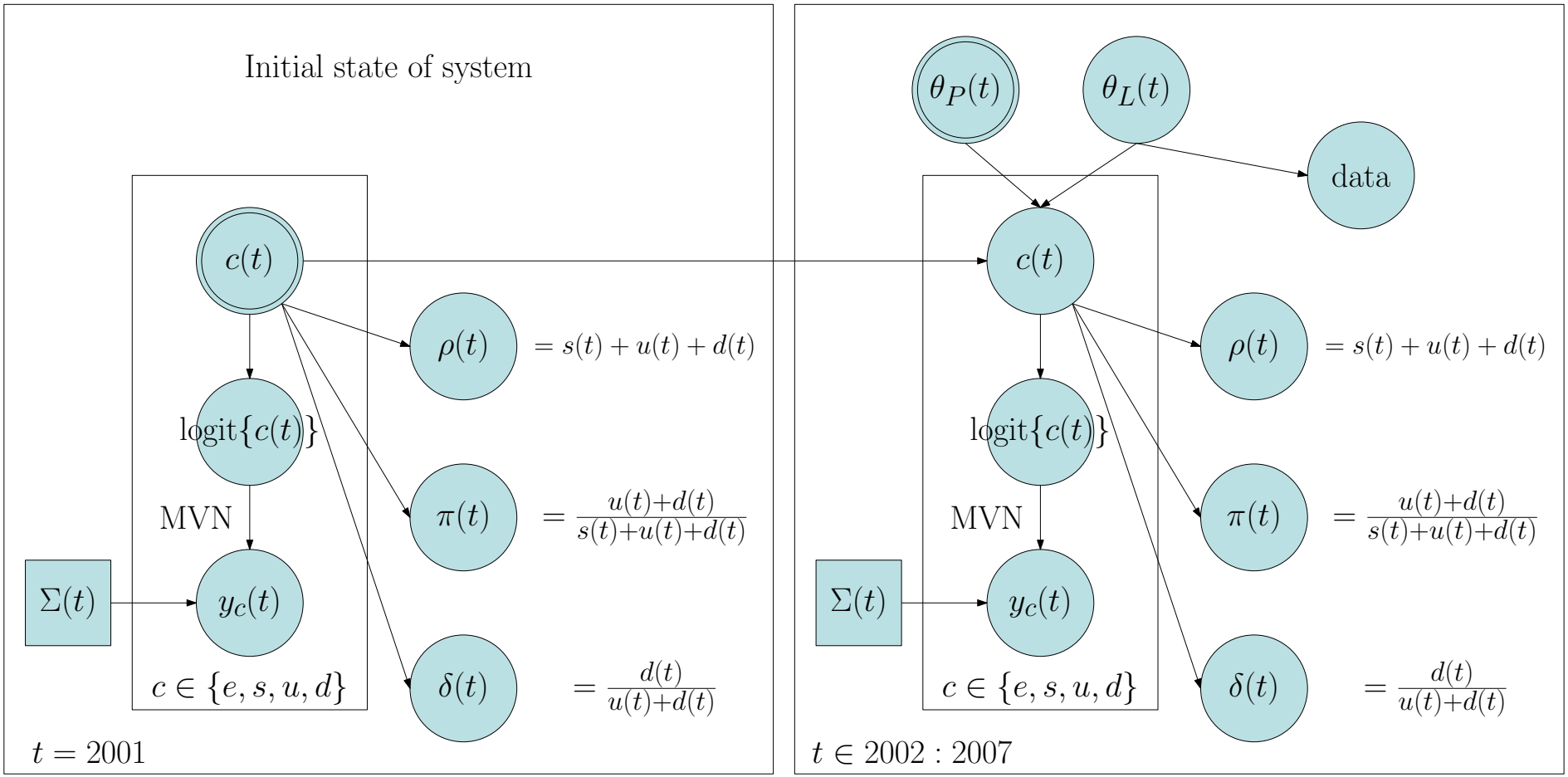


System of equations

MRC Biostatistics Unit

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

Directed acyclic graph



Likelihood

MRC Biostatistics Unit

$$L\Big(\mathbf{c}, \gamma, \mu_{\mathbf{D}}, \mu_{\mathbf{E}}, \kappa \quad | \quad \mathbf{y_c}, \mathbf{\Sigma}, \mathbf{y_{44}}, \mathbf{N}, \mathbf{y}_{\mu_{\mathbf{D}}}, \mathbf{D}, \mathbf{y}_{\text{deaths}}, \mathbf{y}_{\text{diagnoses}}\Big)$$

$$\propto \prod_{t=1}^{7} \Bigg\{ \exp\Big(-\frac{1}{2}(\mathbf{y_c(t)} - \text{logit}(\mathbf{c(t)}))^{\mathrm{T}} \Sigma^{-1}(t)(\mathbf{y_c(t)} - \text{logit}(\mathbf{c(t)}))\Big)$$

$$\times \quad \Big(\gamma(t)^{y_{44}(t)}(1-\gamma(t))^{T(t)-y_{44}(t)}\Big)$$

$$\times \quad \Big(\gamma_D(t)^{y^D_{44}(t)}(1-\gamma_D(t))^{D(t)-y^D_{44}(t)}\Big)$$

$$\times \quad \Big(\mu_D(t)^{y_{\mu_D}(t)}(1-\mu_D(t))^{D(t)-y_{\mu_D}(t)}\Big)$$

$$\times \quad \Big(\mu(t)^{y_{\text{deaths}}(t)}(1-\mu(t))^{T(t)-y_{\text{deaths}}(t)}\Big)$$

$$\times \quad \Big((\kappa(t)U(t))^{y_{\text{diagnoses}}}\exp(-\kappa(t)U(t))\Big) \Bigg\}$$

Priors

MRC Biostatistics Unit

$$
\begin{aligned}
\psi &\sim \text{Normal}(0.0049, 0.0006^2)I(0,) \\
\lambda(t) &\sim \text{Beta}(1,1) \\
\kappa(t) &\sim \text{Beta}(1,1) \\
\mathbf{c(2001)} &\sim \text{Dirichlet}(1,1,1,1)
\end{aligned}
$$

$$
\begin{aligned}
\mu_E(t) &\sim \text{Beta}(1,1) \\
\mu_D(t) &\sim \text{Beta}(1,1) \\
\gamma_E(t) &\sim \text{Beta}(1,1)
\end{aligned}
$$

Migration assumptions

MRC Biostatistics Unit

Model 1: As described above - includes data on proportions $\mathbf{c(t)}$, 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).

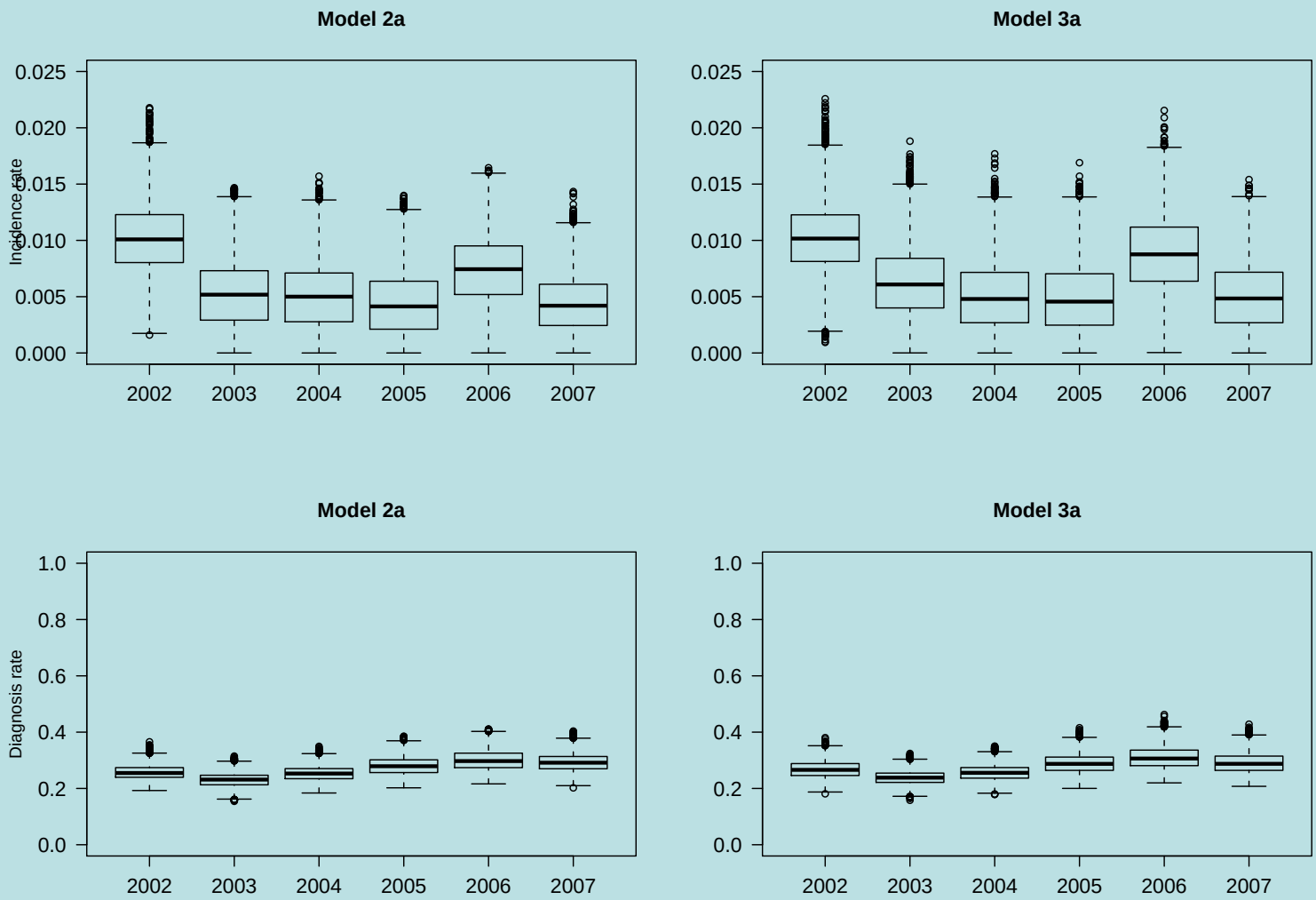
Model 3: As in 2, but assume inward migration of men into E, outward migration from all 4 states (data from ONS).

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

MRC Biostatistics Unit



Dynamic transmission modelling

MRC Biostatistics Unit

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 τ are transmission rates given an infectious contact, also assumed to differ between diagnosed/undiagnosed contacts. Homogeneous mixing assumed within groups.

Further stratification

MRC Biostatistics Unit

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 ϕ are matrices of probabilities of an individual in group g choosing a contact from group i, ranging from *fully assortative* to *homogeneous* mixing. The χ still represent the rates of contact.

Priors

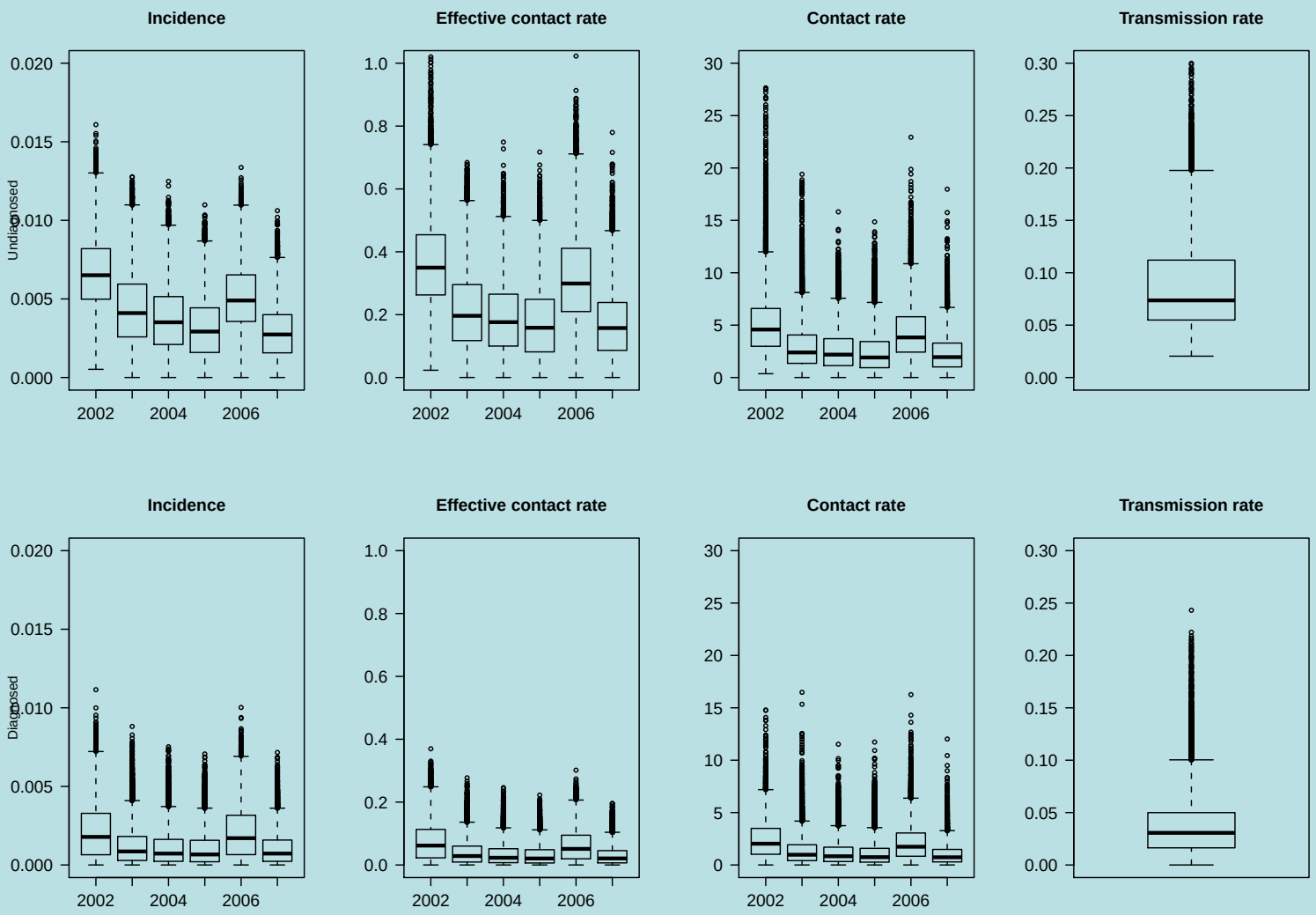
MRC Biostatistics Unit

Different contact groups

$$
\begin{aligned}
\lambda(t) &= \chi_U(t)\tau_U(1-\delta(t))\pi(t) + \chi_D(t)\tau_D\delta(t)\pi(t) \\
\tau_U &\sim \text{Unif}(0, 0.3) \\
\tau_D &\sim \text{Unif}(0, \tau_U) \\
\chi_U(t) &\sim \text{Gamma}(1, 4) \\
\chi_D(t) &\sim \text{Unif}(0, \chi_D(t))
\end{aligned}
$$

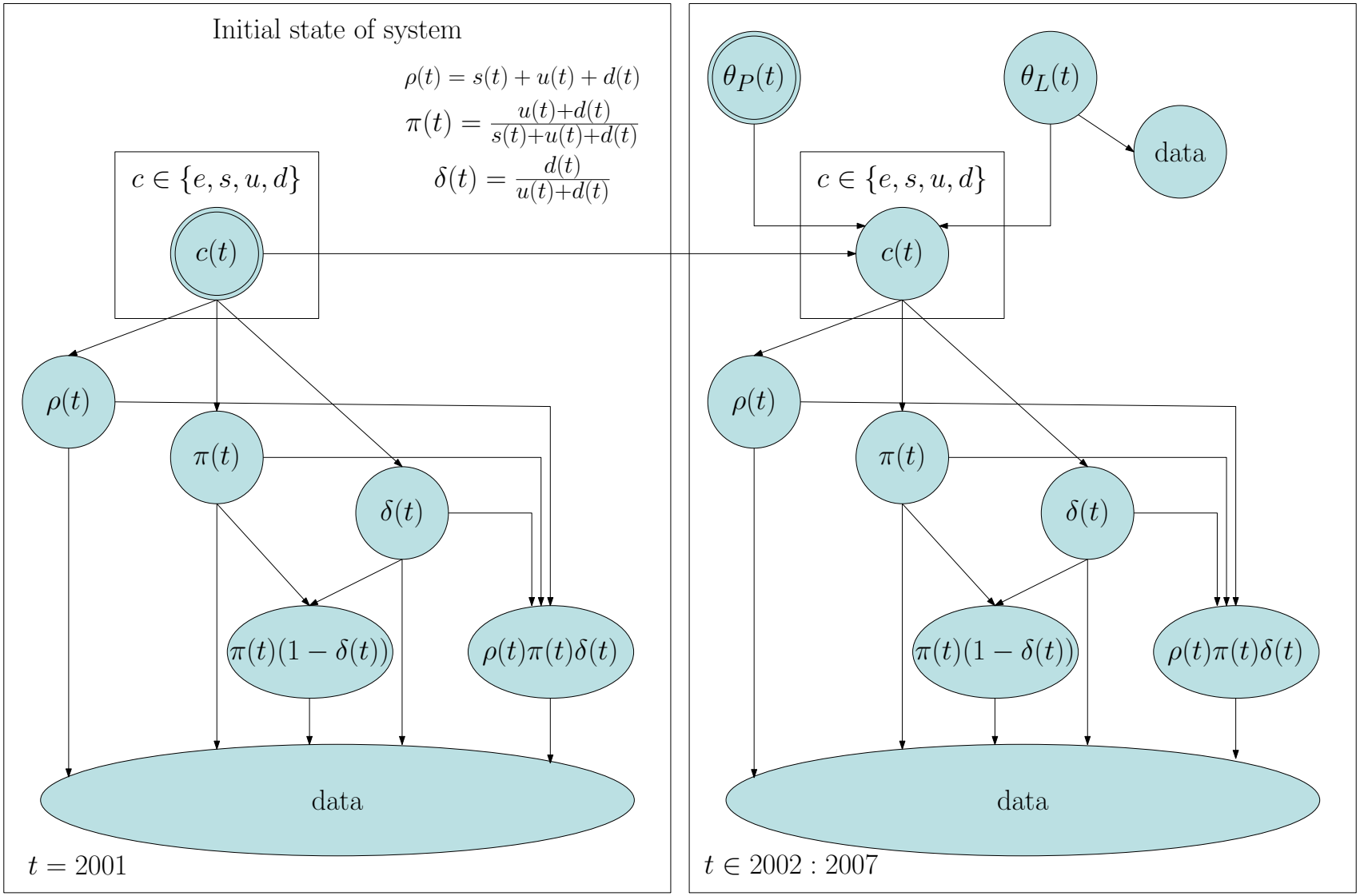
Posterior rates

MRC Biostatistics Unit

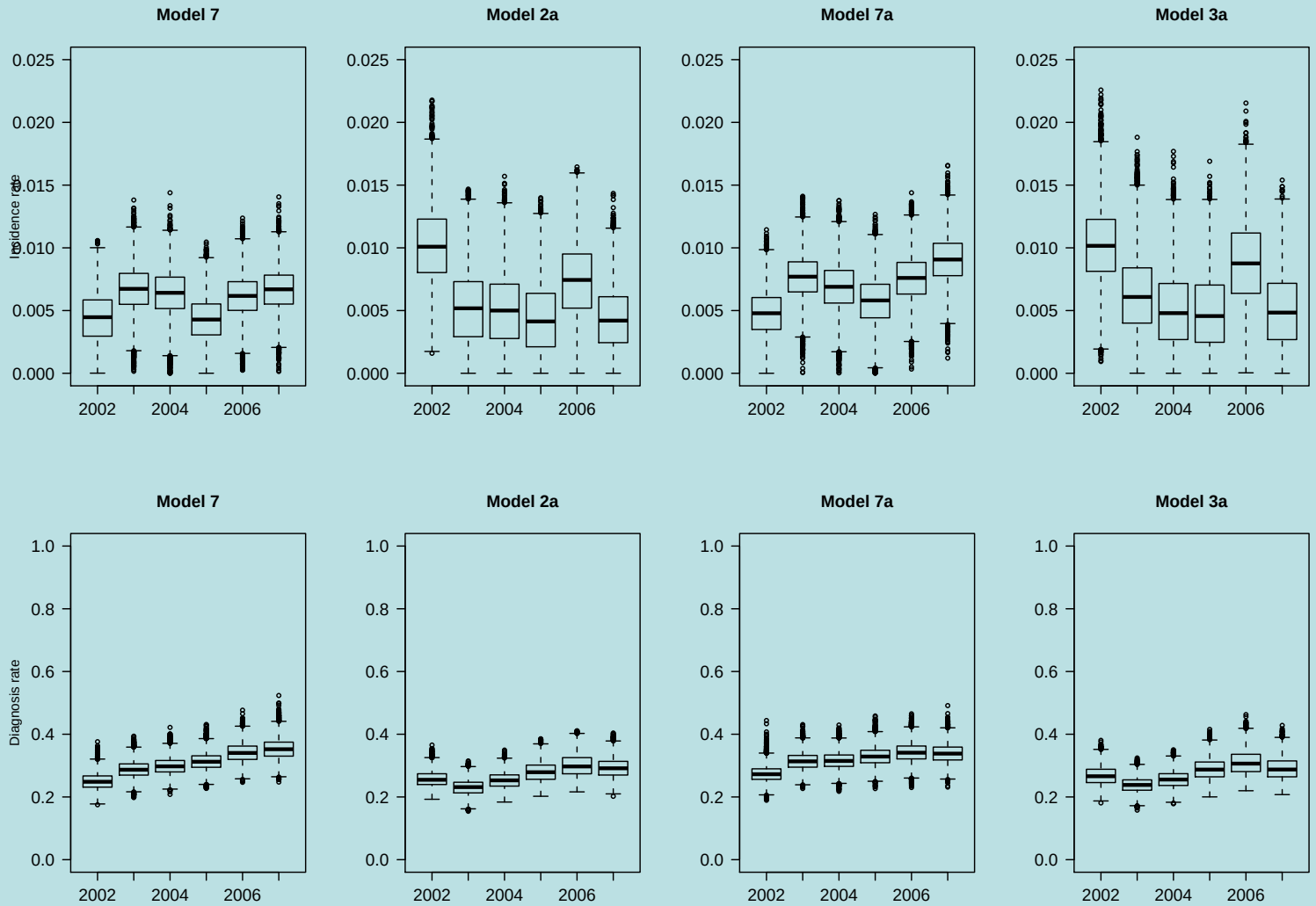


Directed acyclic graph

MRC Biostatistics Unit



Posterior estimates



Ongoing work

MRC Biostatistics Unit

- 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

MRC Biostatistics Unit

- 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

MRC Biostatistics Unit

- Ades and Medley (1994) "Estimates of disease incidence in women based on antenatal or neonatal seroprevalence data: HIV in New York city" *Statist. Med.* **13**(18) pp 1881–1894
- Anderson and May (1991) "*Infectious Diseases of Humans: Dynamics and Control*" Oxford: Clarendon Press
- Goubar et al (2008) "Estimates of HIV prevalence and proportion diagnosed based on Bayesian multi-parameter synthesis of surveillance data" *JRSS(A)* **171**(3) pp 541–580
- Gregson et al (1996) "Demographic approaches to the estimation of incidence of HIV-1 infection among adults from age-specific prevalence data in stable endemic conditions" *AIDS* **10** pp 1689–1697
- Hallett et al (2008) "Estimating incidence from prevalence in generalised HIV epidemics: methods and validation" *PLoS Med.* **5**(4):e80
- Hethcote (2000) "The mathematics of infectious diseases" *SIAM Review* **42** pp 599–653
- Isham (2005) "Stochastic models for epidemics" Chapter 1 in *Celebrating Statistics*, eds A.C. Davison,Y. Dodge and N. Wermuth. OUP, pp 27–54
- Presanis et al (2008) "Conflicting evidence in a Bayesian synthesis of surveillance data to estimate HIV prevalence" *JRSS(A)* **171**(4) pp 915–937