Epidemic monitoring using synthesis of information from multiple sources

Daniela De Angelis

MRC Biostatistics Unit, Cambridge
and
Health Protection Agency, London

InFER 2011, Warwick, 29th March 2011

Outline

Motivation

Planning, implementation and evaluation of public health policies in the UK *e.g.*

- Department of Health (DoH) National Strategy for Sexual Health
- DoH Action Plan for Hepatitis C
- Pandemic influenza Preparedness Strategy

  rely on the monitoring of fundamental aspects of the disease of interest, such as
  - prevalence (undiagnosed prevalence)
  - incidence

  - by age groups and locations
  - at regular intervals (real time!)

Problem

- These characteristics are typically not easily directly measurable (if at all) with little *direct* information available on them
- There is plenty of *indirect* information on functions of these quantities from diverse sources (surveillance, ad hoc surveys etc)
- Estimated from the synthesis of both *direct* and *indirect* information
- This has been common problem underlying most of the work I have been recently involved with

  [Goubar *et al*, 2008], [Presanis *et al*, 2008], [Sweeting *et al*, 2008], [De Angelis *et al*, 2009], [Sweeting *et al*, 2009], [Presanis *et al*, 2010, 2011], [Birrell *et al*, 2011].

Evidence synthesis - a long-established idea

Methods for combining evidence are *not* new:

- The *Bayesian* paradigm
  - combining prior knowledge with new evidence
- *Meta-analysis*
  - combining studies of same type
- *Confidence Profile Method* [Eddy *et al* (1992)]
  - combining information of different types/study designs (medical-decision making literature)
- *Multi-parameter evidence synthesis* [Spiegelhalter *et al* (2004), Ades & Sutton (2006)]
  - epidemiology

A few examples

- Hepatitis C Virus (HCV)
- HIV
- A/H1N1
  - type of challenges posed by the data
  - how tackled by the proposed approach
  - methodological/epidemiological considerations and open questions

HCV

Infection with the Hepatitis C virus (HCV): how many are infected?

- Acquired through exchange of blood with infected individual (e.g. injecting drug use)
- Disease with long incubation; progressive fibrosis of the liver to cirrhosis, hepatocellular carcinoma and death
- Antiviral therapy very effective
- Planning for prevention and treatment implementation needs reliable estimates of the number currently infected

Problem: HCV prevalence estimation

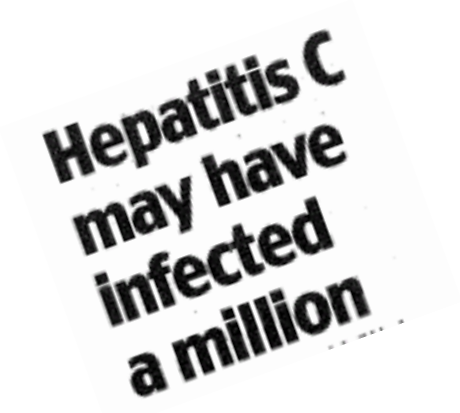
No prevalence study/surveillance representatively covering the general population exists

- Estimates of proportion of infected derived in specific (opportunistic) groups
- Resulting estimates are either not interpretable or heavily biased

Widely different figures have been suggested

Evening standard, 24^{th} May 2006

Daily Mail, 23^{rd} March 2006



Daily Express, 24^{th} May 2006

Proposed Approach

- Information from any available study expressed in terms of 3 main risk groups g: current injecting drug users (CUR); Ex-injecting drug users (EX); Non-IDUs ($NON-IDU$)
- Parameters of interest
  - ρ_{CURrsa}, ρ_{EXrsa}, $\rho_{NON-IDUrsa}$

    prevalence (*i.e.* the proportion) of current, EX, and NON-IDU in the population for region r, gender s, and age-group a.

    - $\rho_{NONrsa} = 1 - \rho_{CURrsa} - \rho_{EXrsa}$
    - $\rho_{IDUrsa} = \rho_{CURrsa} + \rho_{EXrsa}$
  - $\pi_{CURrsa}, \pi_{EXrsa}, \pi_{NON-IDUrsa}$

corresponding prevalence of HCV. Any other quantities can be derived from these *e.g.* π.

Data on ρ_{grsa}

- **Capture re-capture study** in 15-44 years old in London estimate of number of current IDUs
- **Household surveys**:
  - **British Crime Survey** (HO)
  - **Survey of Psychiatric Morbidity** (ONS)
  - **Offending Crime and Justice Survey** (HO)
  - **National Survey of Sexual Attitudes and Lifestyles** (NATSAL)
    - use of non-prescribed IDU drugs - ever - past year

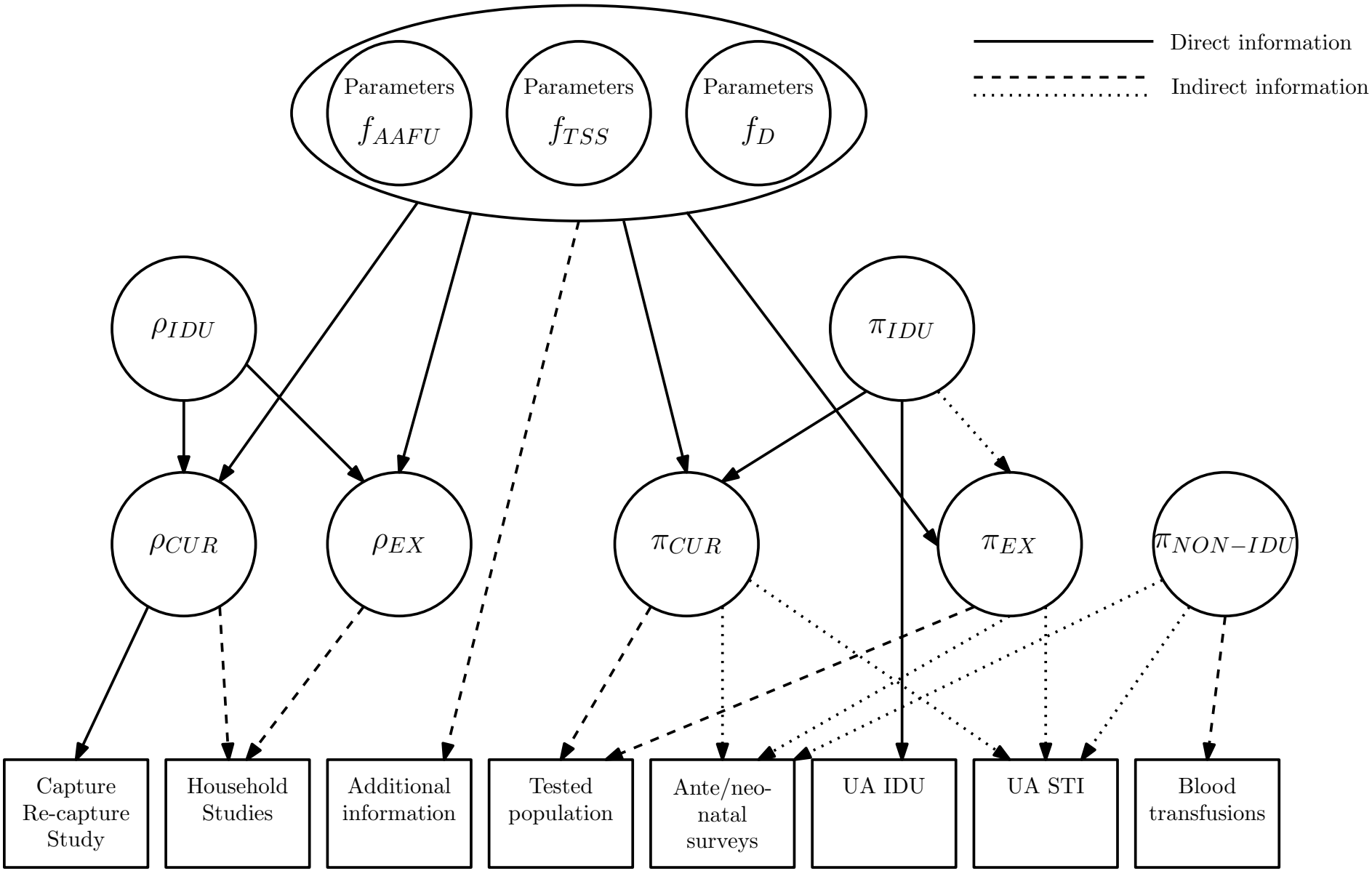
Data on π_{grsa}

- **UA programme in STI clinics**
  - HCV prevalence in ever IDUs and non-IDUs
- **UA programme in current IDUs attending specialist clinics**
- **UA programme in antenatal clinics and neonatal samples**
- **Studies in blood donors**
  - HCV prevalence in low risk population
- **Sentinel laboratory surveillance**
  - HCV prevalence in populations testing for HCV

Challenges

- data structure simple as mostly of the form $\{r_{grsa}, n_{grsa}\}$
- But the observed proportions are typically
  - biased estimates of the true proportions of interest (*e.g.* size of the populations)
  - only interpretable as mixtures of proportions
- lack of direct information on specific proportions of interest (*e.g.* the size of the ex-IDU population)

Graphical model



Challenges: example (1)

- Household studies
  i^{th} study provides information on ρ_{grsa} in the form of $\{r^i_{grsa}, n^i_{grsa}\}$.
  We assume that

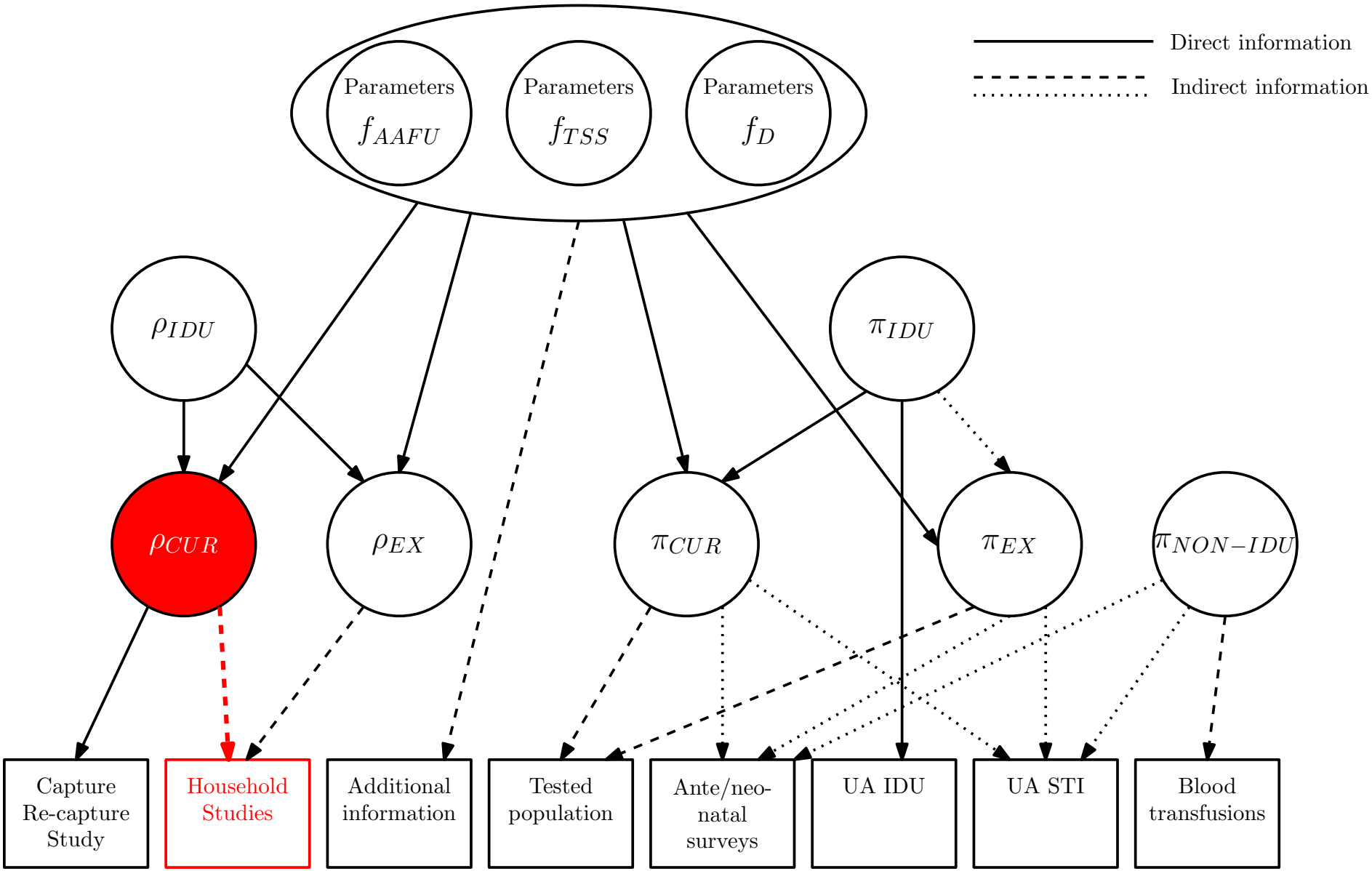
$$r^i_{grsa} \sim \text{Binomial}(n^i_{grsa}, \rho^i_{grsa})$$

and

$$logit(\rho^i_{grsa}) = logit(\rho_{grsa}) + b$$

where b is a bias parameter.

Graphical model



Challenges: example (2)

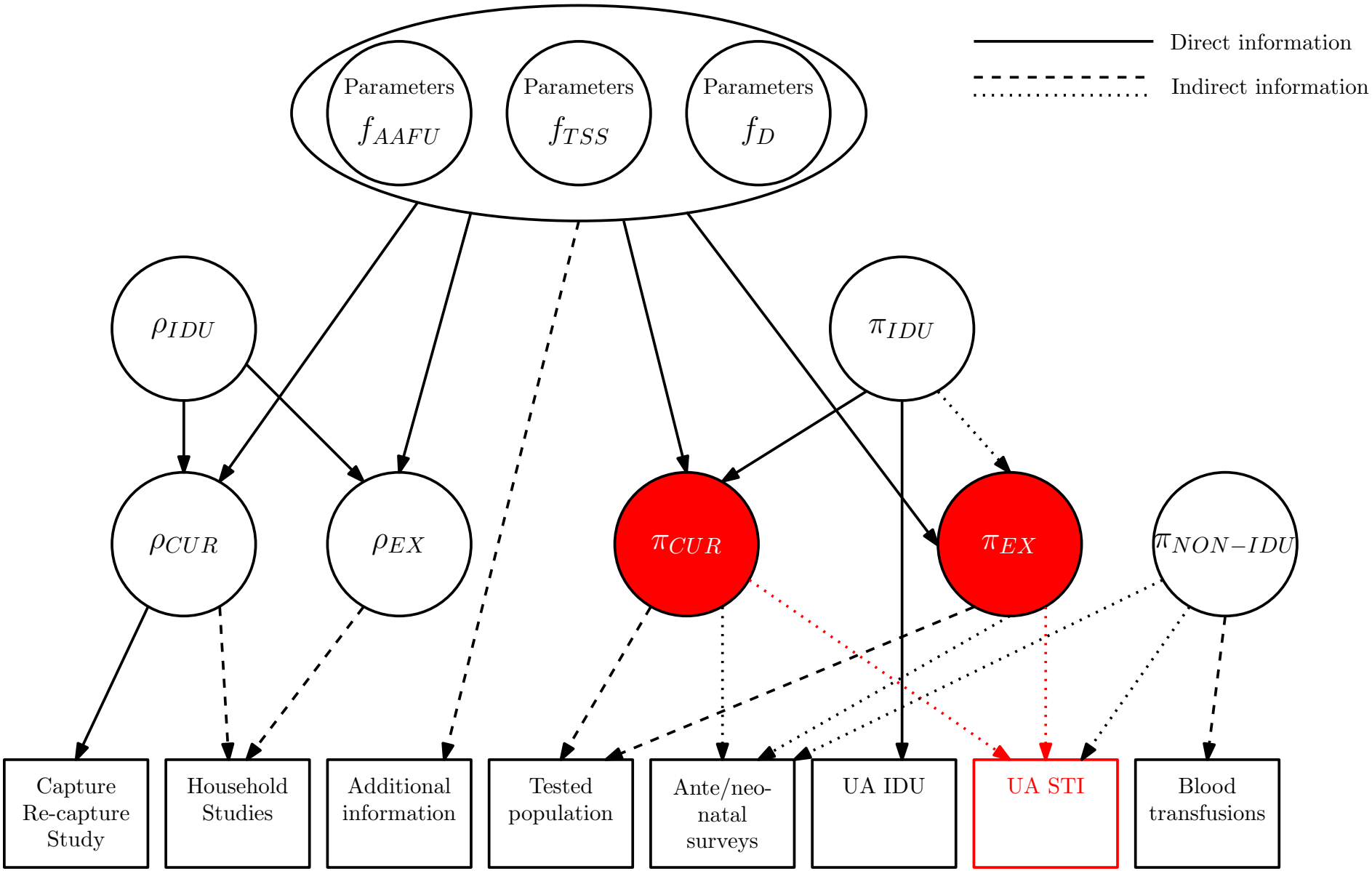
- Mixture of proportions
  the HCV prevalence in STI clinic attendees who have ever injected π^{STI}_{IDUrsa} estimated using $\{r^{STI}_{IDUrsa}, n^{STI}_{IDUrsa}\}$ can only be interpreted as being

$$\pi^{STI}_{IDUrsa} = \psi\pi^{STI}_{CURrsa} + (1-\psi)\pi^{STI}_{EXrsa}$$

  where the mixture coefficient ψ is informed by the NATSAL survey. Thus we assume that

$$r^{STI}_{IDUrsa} \sim \text{Binomial}(n^{STI}_{IDUrsa}, \pi^{STI}_{IDUrsa})$$

Graphical model



HIV

Infection with HIV: what are trends in prevalence of undiagnosed infections?

- HIV is a long, asymptomatic disease with many infections undiagnosed
- Undiagnosed infections contribute to transmission - lack of access to treatment
- Reliable estimates of the number infected, particularly those still undiagnosed, over time are required for public health policy

Parameters of interest

- ρ_{gtr} prevalence (*i.e.* the proportion in the population) of risk-group g in the population at time t for region r
- π_{gtr} corresponding prevalence of HIV
- δ_{gtr} proportion of infections diagnosed in risk-group g, region r

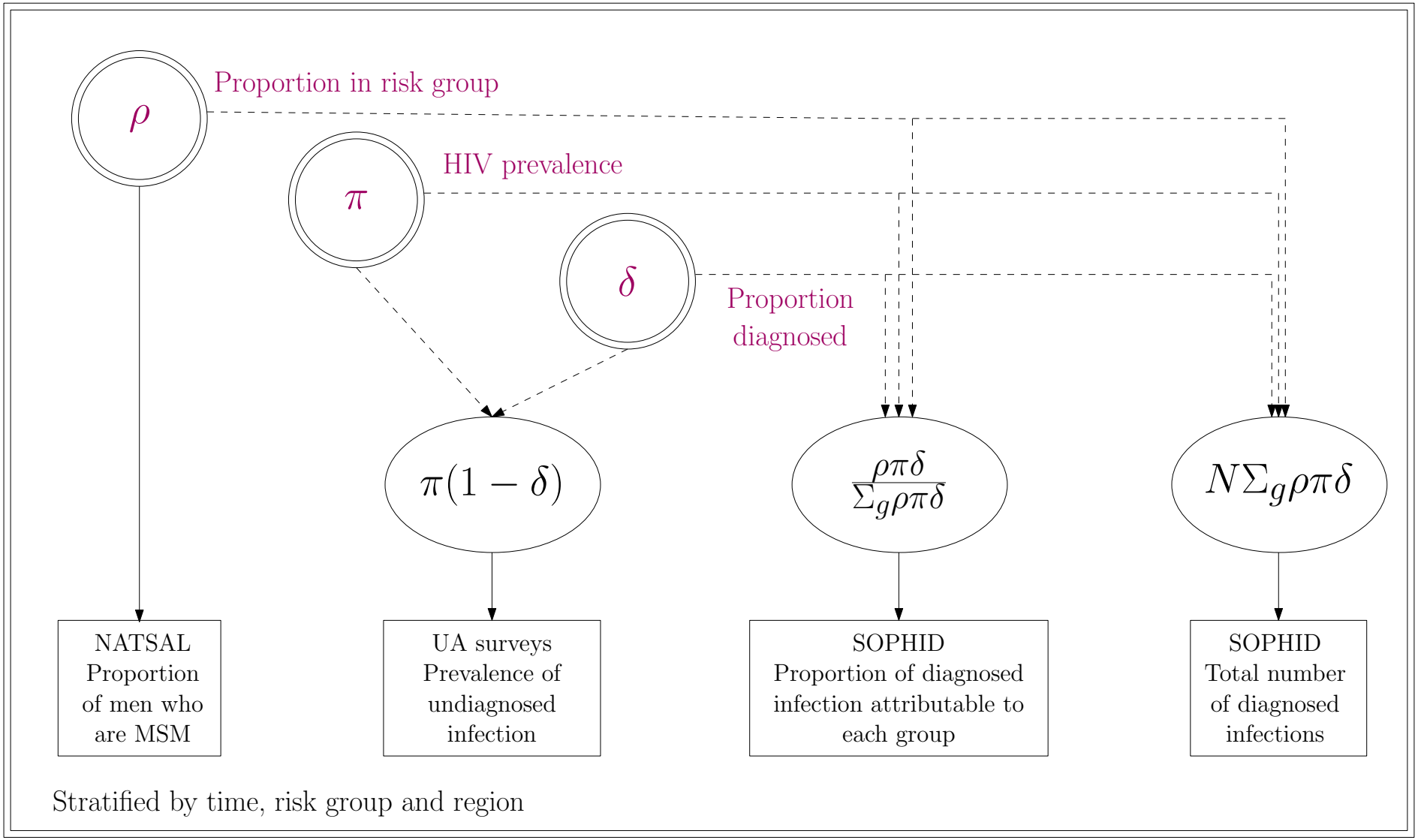
Any other quantities can be derived from these *e.g.* $\pi_{gtr}(1-\delta_{gtr})$.

Availability of data: 13 risk groups, 3 regions, over time

	Risk group	N	ρ	π	δ	$\pi(1-\delta)$	$f(\rho,\pi,\delta)$
Men	MSM		✓			✓	
	IDUs		✓	✓	✓		✓
	Born sub-Saharan Africa		✓				✓
	STI clinic attendees		✓			✓	
	Lower risk						
	ALL	✓					✓
Women	IDUs		✓	✓	✓		✓
	Born sub-Saharan Africa		✓	✓	✓		✓
	STI clinic attendees		✓			✓	
	Lower risk						
	Born UK/elsewhere			✓	✓		
	ALL	✓		✓			✓

Note: multiplicity of data for some parameters

Graphical model: MSM



A/H1N1

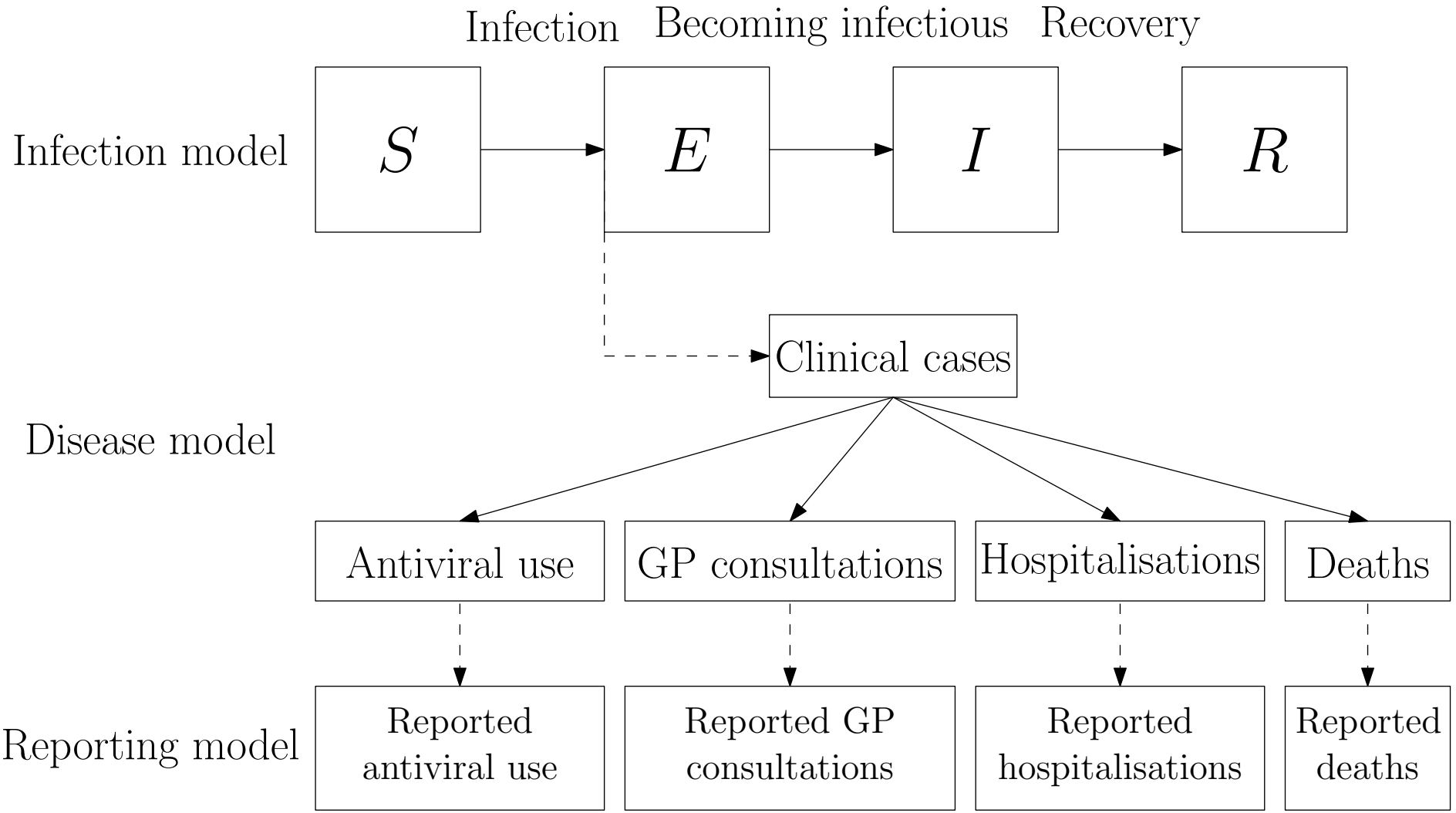
A/H1N1 2009 in England: is it possible to inform policy in the midst of an epidemic?

Monitoring the epidemic through the (real time) estimation and prediction of

- the number of infections
- the number of symptomatic cases
- the number of severe cases (hospitalisations, ICU admissions, deaths)

by age group and region

Ideally

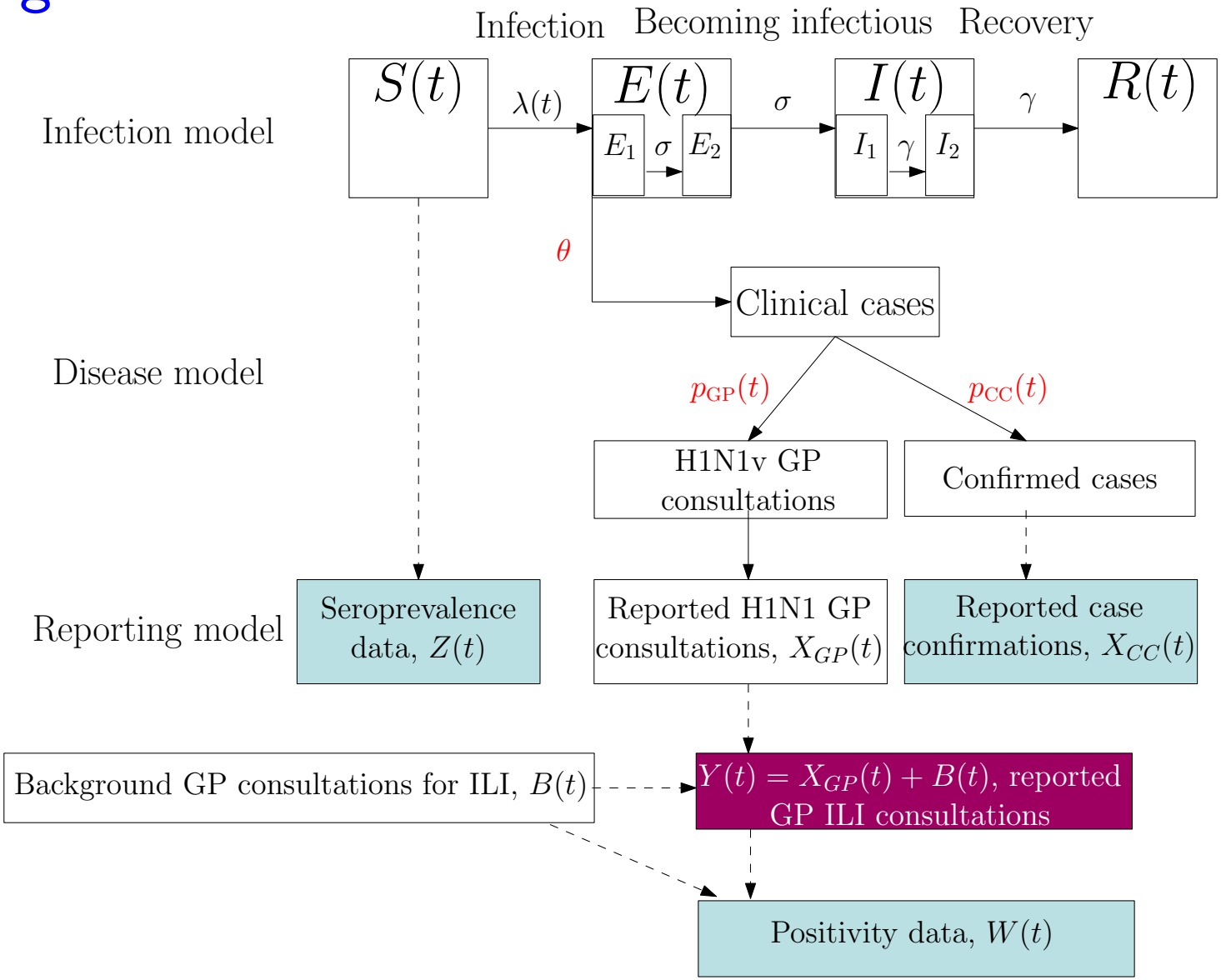

Infection
Becoming infectious
Recovery
Infection model
S
E
I
R
Clinical cases
Disease model
Antiviral use
GP consultations
Hospitalisations
Deaths
Reporting model
Reported antiviral use
Reported GP consultations
Reported hospitalisations
Reported deaths

In reality

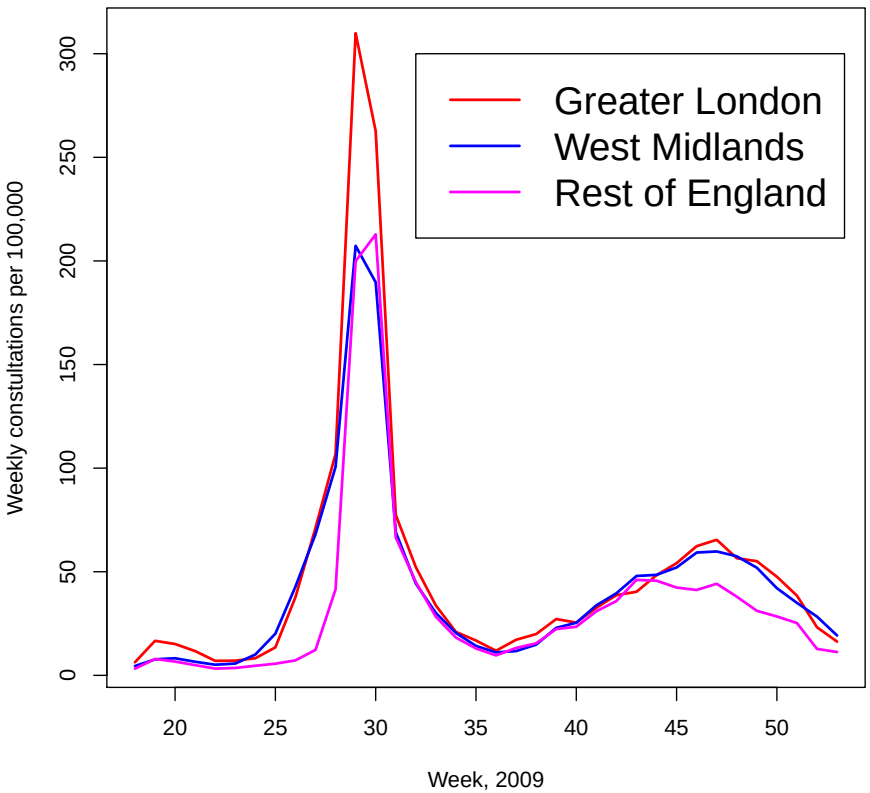
- Q-Surveillance Daily counts of individuals reporting symptoms of influenza-like illness (ILI) at general practices (GPs).
- RMN + RCGP Sentinel GP surveillance schemes, swabbing ILI patients. Provides data on swab positivity for A/H1N1.
- HPA returns service Cross-sectional seroepidemiological surveys.
- Virologically confirmed cases.

All data stratified by age class (< 1, $1-4$, $5-14$, $15-24$, $25-44$, $45-64$, and $65+$ years) from the 1^{st} May until 31^{st} December.

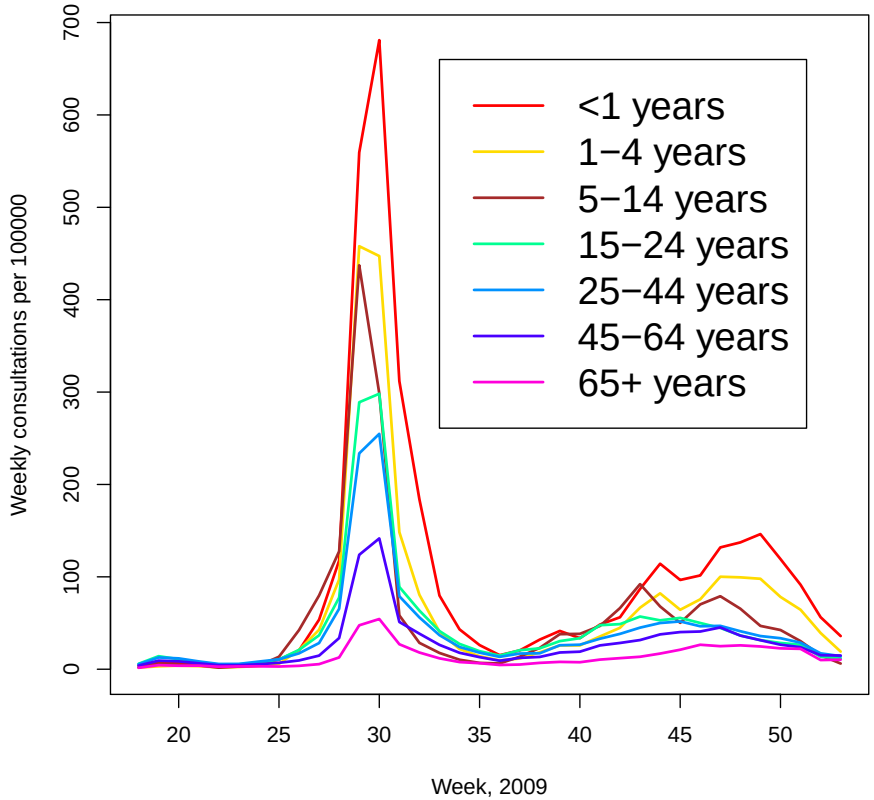
Integrated model for transmission and reporting



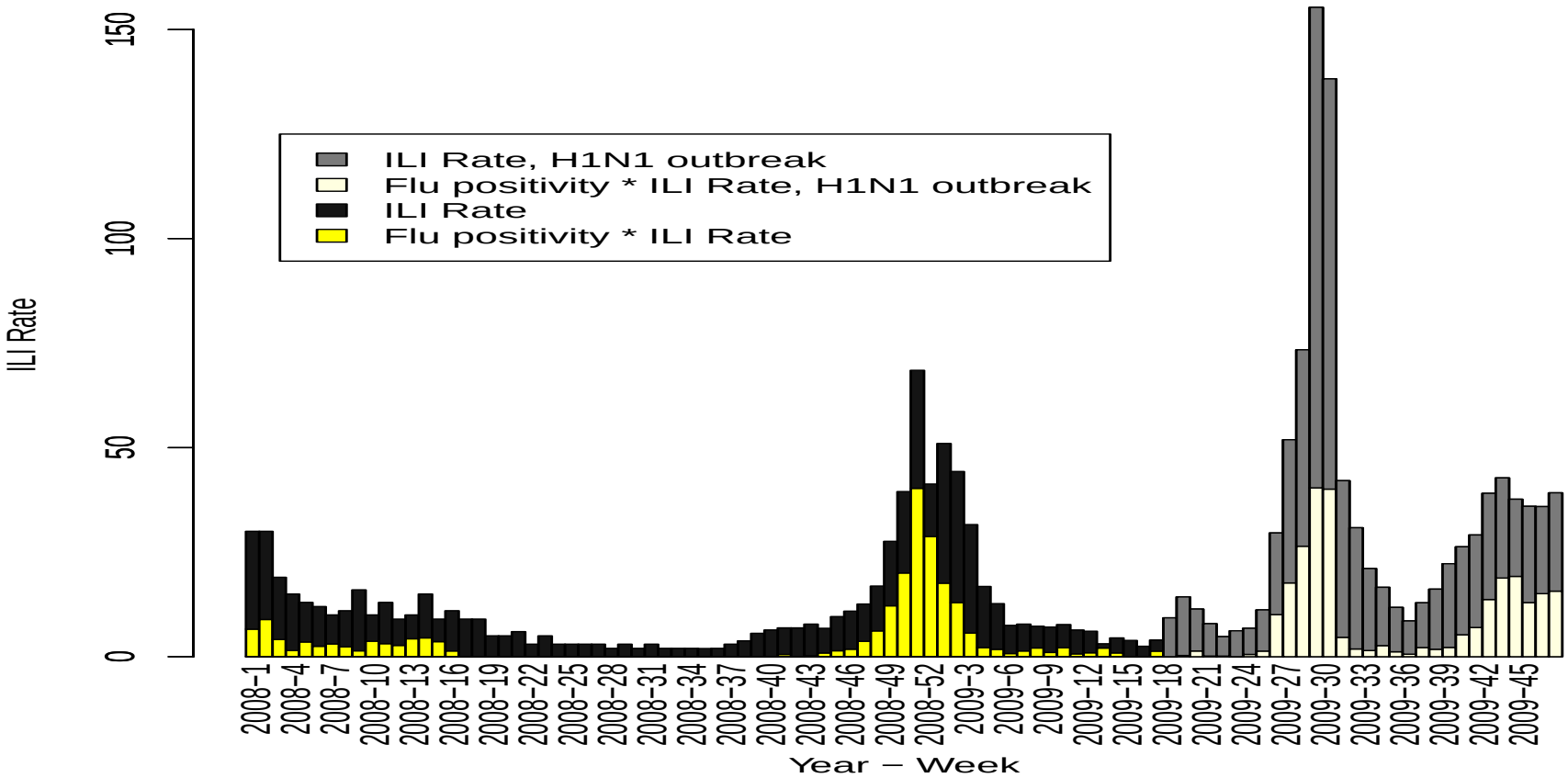
(a)



(b)



(a) Weekly consultations by region (Greater London, West Midlands, Rest of England) and (b) by age group



GP consultations for ILI and virological positivity for the presence of ANY flu, by week from end-2005 to week 50, 2009

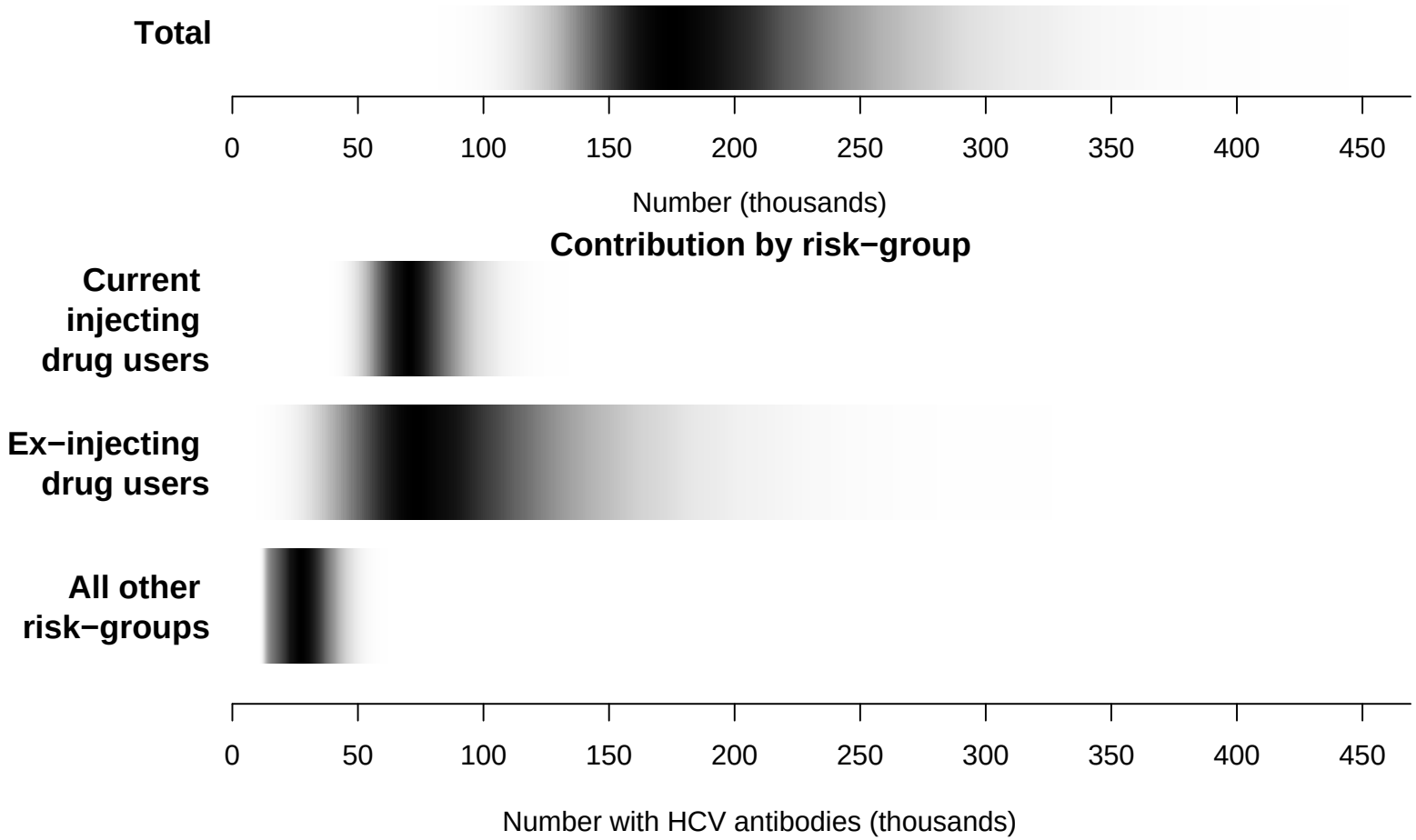
Common statistical formulation

- Interest: estimation of $\boldsymbol{\theta} = (\theta_1, \theta_2 \ldots, \theta_k)$ on the basis of a collection of data $\mathbf{y} = (y_1, y_2 \ldots, y_n)$
- Each y_i provides information on
  - a *single* component of $\boldsymbol{\theta}$, or
  - a *function* of one or more components, *i.e.* on a quantity $\psi_i = f(\boldsymbol{\theta})$

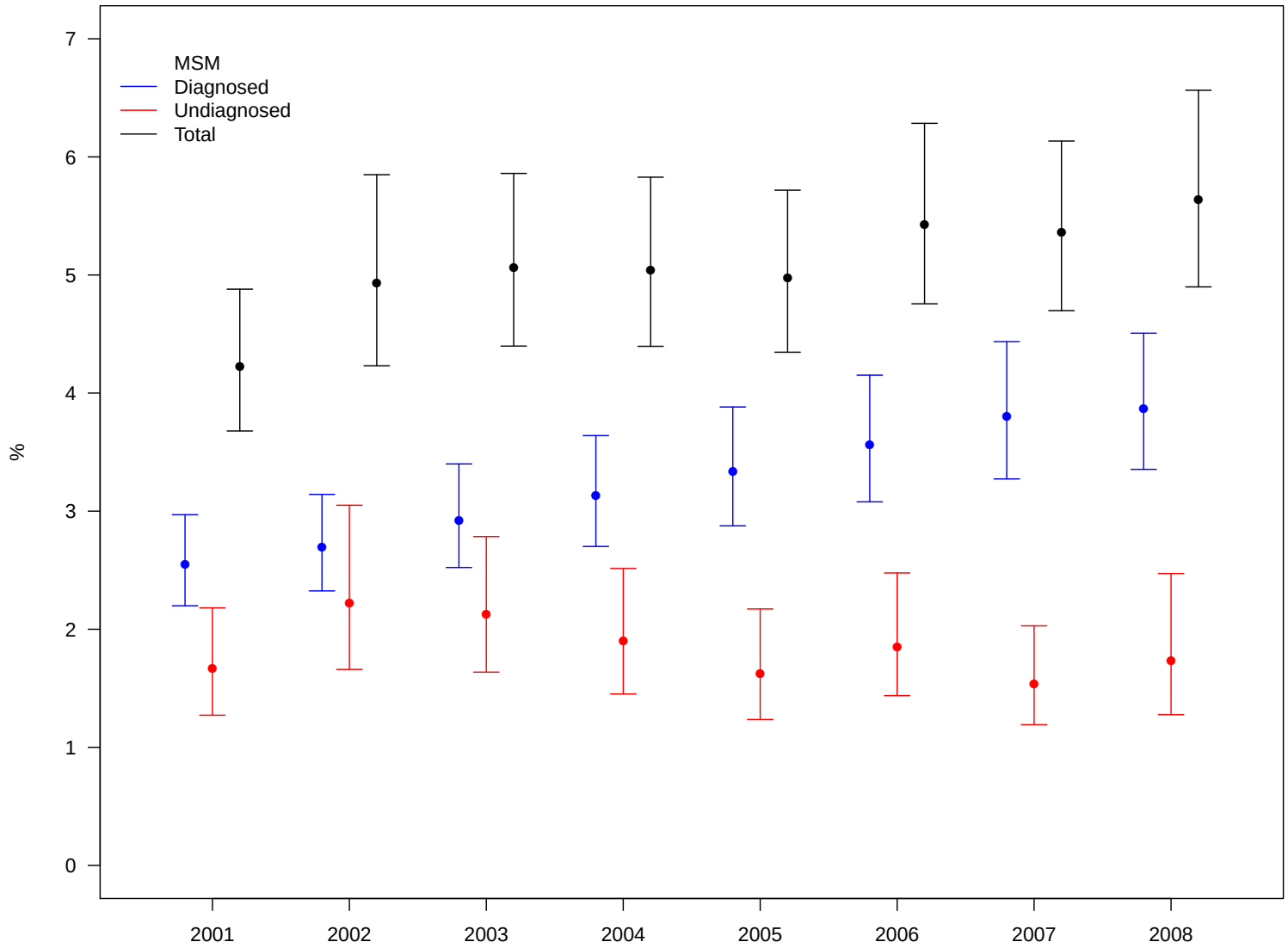
Thus inference is conducted on the basis of both **direct** and **indirect** information.

- Maximum likelihood: $L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(y_i; \boldsymbol{\theta})$
- Bayesian: $p(\boldsymbol{\theta} \mid \mathbf{y}) \propto p(\boldsymbol{\theta}) \times L(\boldsymbol{\theta})$

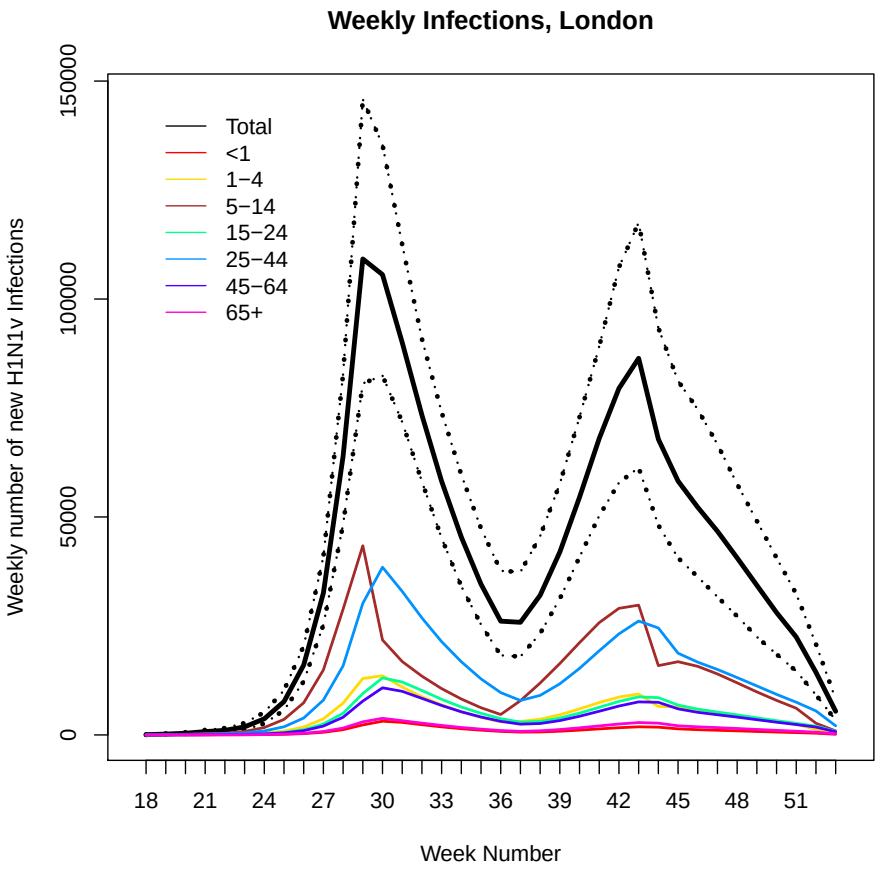
Posterior distributions for the number of individuals with anti-HCV antibodies, E&W 2003



Posterior distributions for the proportion of undiagnosed HIV infections in MSM



London: Infections, Cases, Attack Rate



End August

Infections	672000
	(544000, 826000)
Symptomatic	217000
cases	(133000, 324000)
IAR	9%
	(7%, 11%)

End December

Infections	1441000
	(1229000, 1681000)
Symptomatic	469000
Cases	(298000, 676000)
IAR	19%
	(16%, 22%)

Methodological issues

Powerful approach that allows use of all available information inevitably leading to complex probabilistic models

- How do we assess these complex models?
  - appropriateness compared to alternative models
  - detection of conflicts between data items
  - influence of particular data items on the resulting inference

Model assessment: instruments

Deviance

$$D(\boldsymbol{\theta}) = -2[\log\{\mathrm{p}(\mathbf{y} \mid \boldsymbol{\theta})\} - \log\{\mathrm{p}(\mathbf{y} \mid \hat{\boldsymbol{\theta}})\}]$$

- Posterior mean deviance ($\bar{D}$)

$$\bar{D} = E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})]$$

- Deviance Information Criteria (DIC)

$$DIC = \bar{D} + p_D$$

[Spiegelhalter *et al*, 2002]

Model assessment: model choice. HCV

Model	DIC	ρ_{CUR} (%)	ρ_{EX} (%)	π_{CUR} (%)	π_{EX} (%)	π_{NON} (%)	π (%)
No bias $b = 0$	1022	0.26 (0.22, 0.30)	0.73 (0.65, 0.81)	33.7 (30.3, 37.3)	19.9 (17.2, 22.8)	0.094 (0.048, 0.152)	0.32 (0.27, 0.39)
Common bias, surveys & risk-groups $b = b$	**976**	0.67 (0.49, 0.93)	**2.69** (1.83, 4.04)	32.7 (29.2, 36.5)	18.9 (16.3, 21.7)	0.098 (0.048, 0.157)	**0.82** (0.60, 1.16)
Risk-group specific bias $b = b_g$	**978**	0.68 (0.49, 0.96)	**1.41** (0.58, 3.19)	33.0 (29.3, 37.2)	19.7 (16.8, 22.7)	0.091 (0.046, 0.150)	**0.60** (0.39, 0.97)
Survey specific bias $b = b^i$	981	0.70 (0.50, 1.01)	2.79 (1.84, 4.27)	32.6 (29.2, 36.5)	18.8 (16.2, 21.6)	0.098 (0.049, 0.158)	0.85 (0.60, 1.21)
Survey & risk-group specific bias $b = b_g^i$	986	0.69 (0.49, 0.96)	1.45 (0.61, 3.31)	33.2 (29.4, 37.3)	19.6 (16.7, 22.7)	0.091 (0.046, 0.151)	0.61 (0.39, 0.99)

Model assessment: conflict between data sources

$$\bar{D} = \sum_{i}^{n} \bar{D}_i$$

assuming independence between the n data sources becomes the sum of the item specific deviance contributions

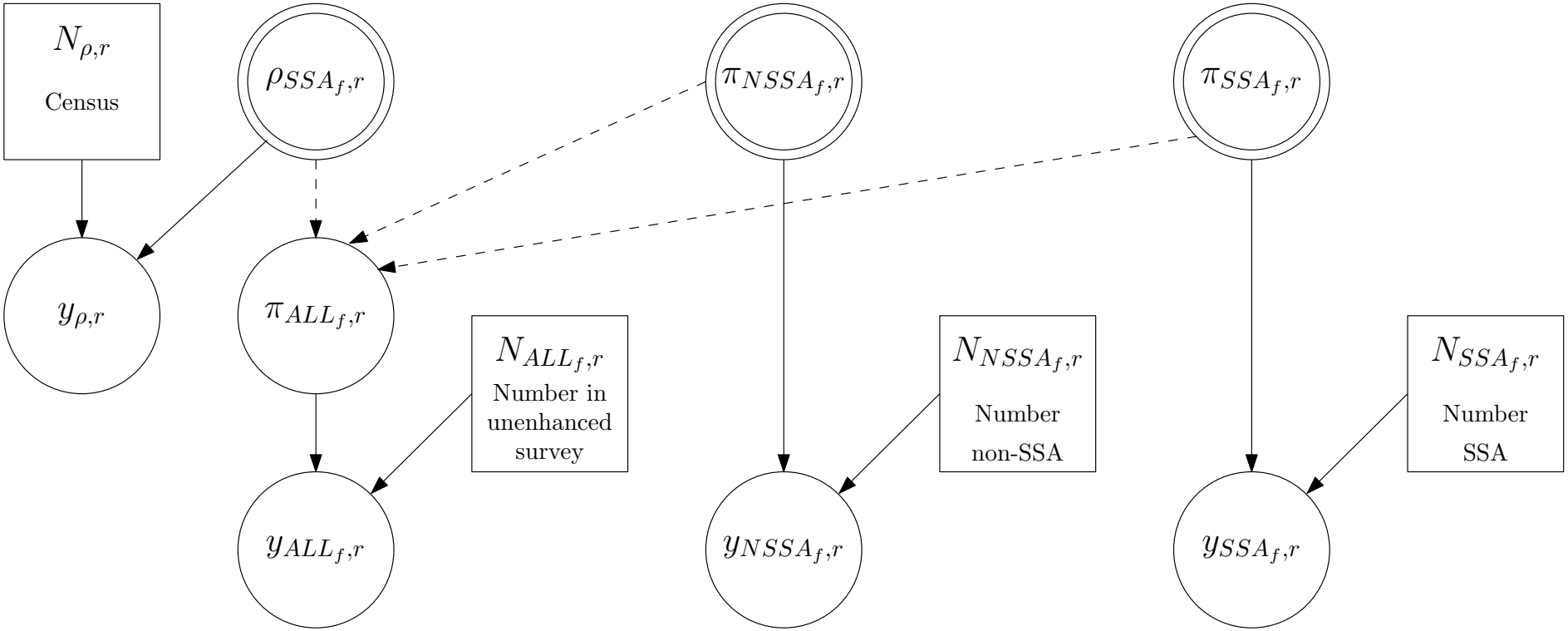
- can be usefully employed to identify conflict between data sources

Availability of data: 13 risk groups, 3 regions, over time

	Risk group	N	ρ	π	δ	$\pi(1-\delta)$	$f(\rho,\pi,\delta)$
Men	MSM		✓			✓	
	IDUs		✓	✓	✓		✓
	Born sub-Saharan Africa		✓				✓
	STI clinic attendees		✓			✓	
	Lower risk						
	ALL	✓					✓
Women	IDUs		✓	✓	✓		✓
	Born sub-Saharan Africa		✓	✓	✓		✓
	STI clinic attendees		✓			✓	
	Lower risk						
	Born UK/elsewhere			✓	✓		
	ALL	✓		✓			✓

Note: multiplicity of data for some parameters

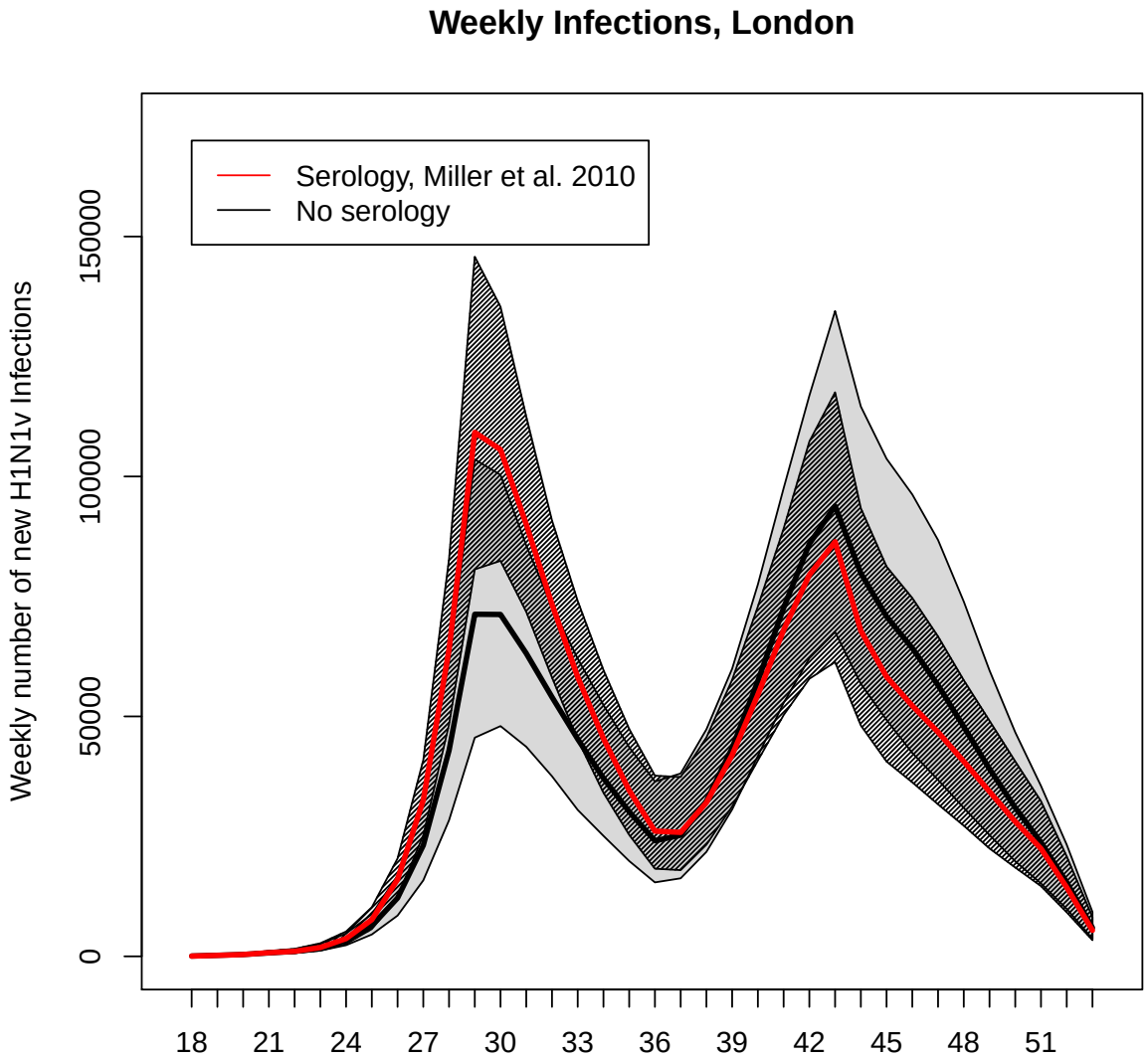
Model assessment: conflict. HIV



$$\pi_{ALL_f,r} = \pi_{SSA_f,r}\rho_{SSA_f,r} + \pi_{NSSA_f,r}(1 - \rho_{SSA_f,r})$$

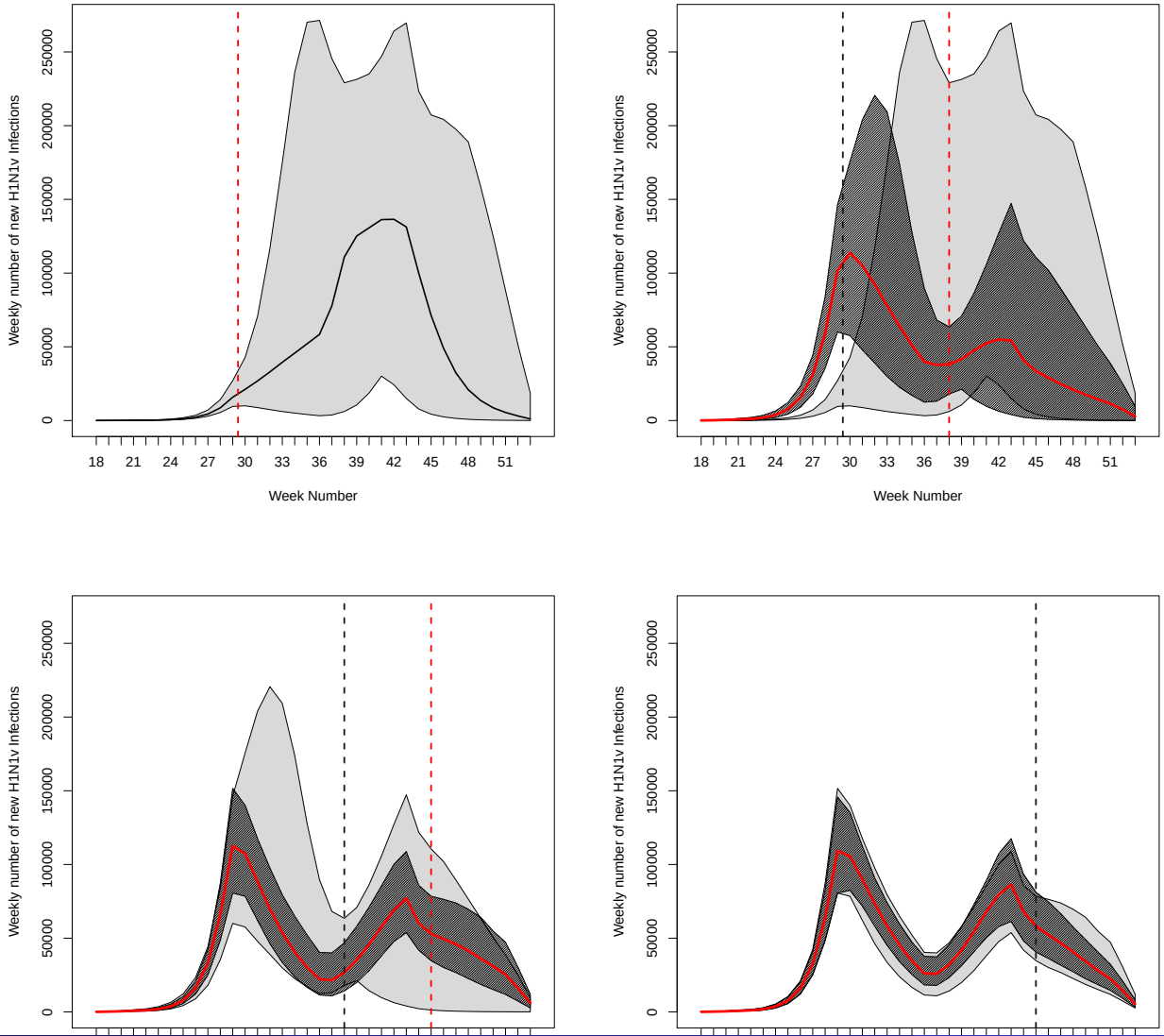
Model assessment: influence. A/H1N1

What is the role of serological studies?



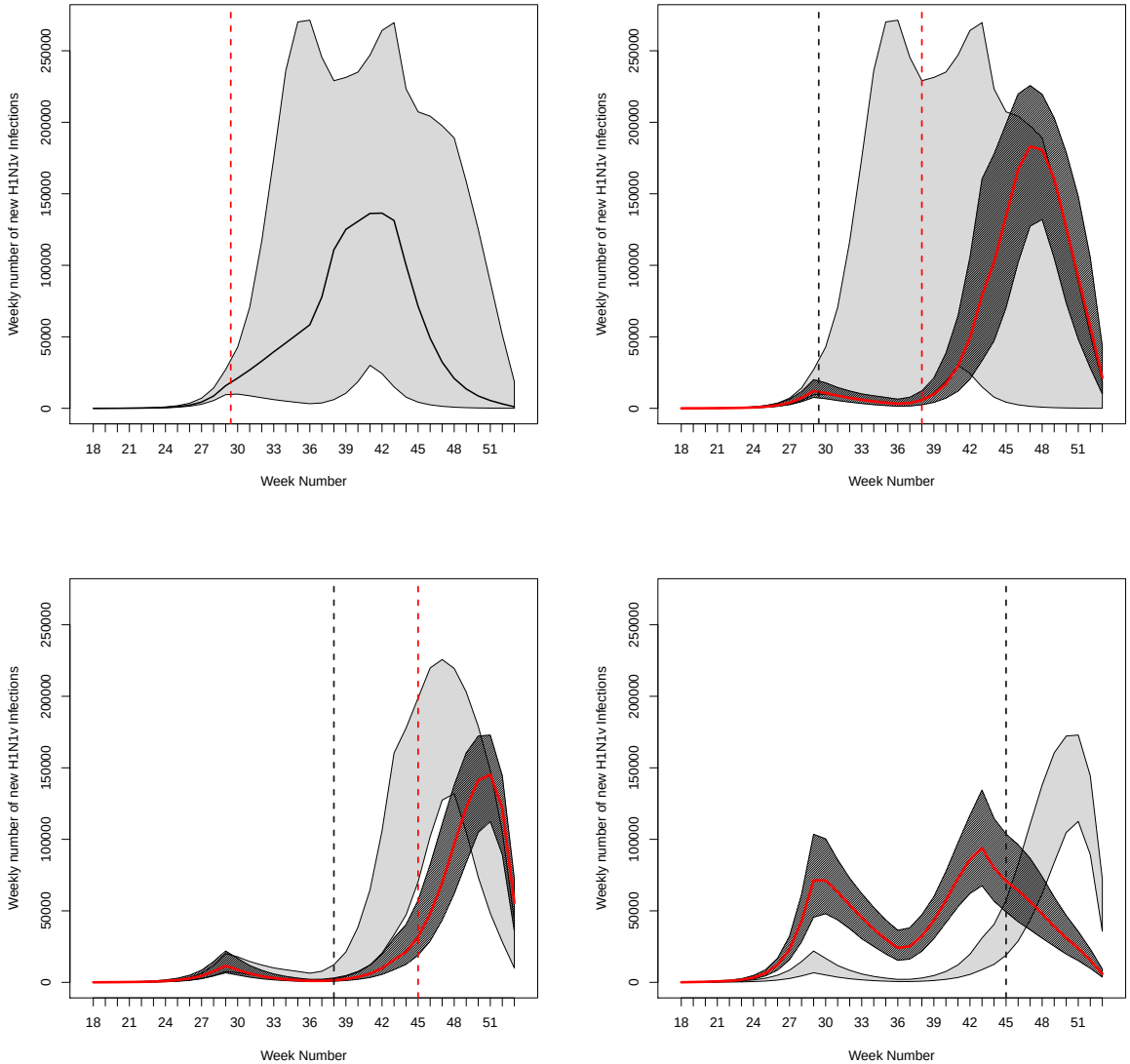
Model assessment: influence. A/H1N1

Sequential analyses with serology based on 83, 143, 192 and 245 days of epidemic surveillance data



In a world without serology...

Other data sources unable to overcome prior information



Discussion

- Example of estimation relevant to problems of public health
- Use of multiple sources of evidence leads to complex probabilistic models
- Increasingly expert at formulating and estimating models
- The establishment of a well-defined iterative process of model criticism lags behind this expertise
- Model criticism becomes more crucial but harder as the number of data sources increases and the model becomes more complex
- Work on approaches to conflict detection, model choice and assessment needed

Co-authors

- Statisticians
  Paul Birrell, Anne Presanis and Michael Sweeting, MRC-BSU
  Tony Ades, Bristol University
- Epidemiologists
  Matthew Hickman, Bristol University
  Vivian Hope, HPA and LSHTM,
  Noel Gill, Richard Pebody, Mary Ramsay, HPA