The Annals of Applied Statistics
2011, Vol. 5, No. 4, 2359–2384
DOI: 10.1214/11-AOAS488

MODELING OF THE HIV INFECTION EPIDEMIC IN THE NETHERLANDS: A MULTI-PARAMETER EVIDENCE SYNTHESIS APPROACH[1]

BY STEFANO CONTI, ANNE M. PRESANIS, MAAIKE G. VAN VEEN, MARIA XIRIDOU, MARTIN C. DONOGHOE, ANNEMARIE RINDER STENGAARD AND DANIELA DE ANGELIS

Health Protection Agency, Medical Research Council Biostatistics Unit, National Institute for Public Health and the Environment, National Institute for Public Health and the Environment, World Health Organization Regional Office for Europe, World Health Organization Regional Office for Europe, and Health Protection Agency and Medical Research Council Biostatistics Unit

Multi-parameter evidence synthesis (MPES) is receiving growing attention from the epidemiological community as a coherent and flexible analytical framework to accommodate a disparate body of evidence available to inform disease incidence and prevalence estimation. MPES is the statistical methodology adopted by the Health Protection Agency in the UK for its annual national assessment of the HIV epidemic, and is acknowledged by the World Health Organization and UNAIDS as a valuable technique for the estimation of adult HIV prevalence from surveillance data. This paper describes the results of utilizing a Bayesian MPES approach to model HIV prevalence in the Netherlands at the end of 2007, using an array of field data from different study designs on various population risk subgroups and with a varying degree of regional coverage. Auxiliary data and expert opinion were additionally incorporated to resolve issues arising from biased, insufficient or inconsistent evidence. This case study offers a demonstration of the ability of MPES to naturally integrate and critically reconcile disparate and heterogeneous sources of evidence, while producing reliable estimates of HIV prevalence used to support public health decision-making.

1. Introduction. Refining and advancing the current understanding of the dynamics of the HIV epidemic attracts a continued interest from the epidemiological and medical community. Both national and international public health institutes recognize the importance of improving current methods to monitor HIV prevalence, as this constitutes a key input to inform public health-care policies and resource allocation.

Received June 2010; revised May 2011.

[1]Supported by the World Health Organization Regional Office for Europe (WHO/Europe) and Medical Research Council Grants U.1052.00.007 and G0600675.

Key words and phrases. Bayesian inference, bias adjustment, evidence synthesis, hierarchical models, HIV infection.

A number of approaches have been proposed in the statistical literature, starting from the back-calculation method [Brookmeyer and Gail (1988)], initially devised to obtain an estimate of HIV prevalence. The most popular estimation methods (so-called "direct") typically rely on evidence specifically around HIV prevalence [Giesecke et al. (1994); Petruckevitch et al. (1997); Houweling et al. (1998); Karon, Khare and Rosenberg (1998); Ramón et al. (2002); McGarrigle et al. (2006)]. In broad terms, direct methods assume a target population of size $N = \sum_g N_g$ to be divided into mutually exclusive subgroups $g = 1, \ldots, G$ of corresponding size N_g. Each subgroup is characterized by a given degree of risk behavior and consists of $N_g(1-\pi_g)$ uninfected and $N_g\pi_g$ infected individuals, where π_g denotes the unknown subgroup-specific HIV prevalence. Prevalent cases $N_g\pi_g$ can in turn be split into $N_g\pi_g\delta_g$ diagnosed and $N_g\pi_g(1-\delta_g)$ undiagnosed individuals, as determined by the (unknown) proportion δ_g of HIV positive cases diagnosed within each subgroup. Provided enough cross-sectional surveillance- or survey-based information is available to estimate subgroup sizes and parameters, the number of subgroup-specific diagnosed and undiagnosed prevalent cases can be inferred by multiplying corresponding estimates of N_g and π_g with δ_g and $1-\delta_g$, respectively. These in turn can be summed across subgroups to obtain a point estimate of the total number of HIV infections in the population.

While at a first glance appealing, direct methods suffer from both conceptual and practical complications. Data may: (i) be insufficient to inform directly relevant parameters, like prevalence in hard-to-reach subgroups; (ii) relate to individuals matching multiple risk profiles; and/or (iii) be affected by selection and reporting biases. Without the inclusion of supplementary evidence, these problems are normally tackled via unverifiable assumptions, ad-hoc adjustments and/or removal of selected data [Goubar et al. (2008)]. Moreover, the common practice of using only as many items of (highest quality) evidence as the number of parameters of interest is hardly justified under a decision-making perspective. Decisions around research prioritization and service provision are more rationally and robustly taken when driven by a comprehensive, rather than selective, use of available information [Claxton, Sculpher and Drummond (2002)], provided the varying degree of accuracy of the components of the evidence base is correctly recognized and taken into account in the analysis.

Conversely, *multi-parameter evidence synthesis* (henceforth MPES) offers a coherent analytical framework designed to make rational and exhaustive use of the whole body of information available [Ades and Sutton (2006)], thus circumventing the above shortcomings. A disparate pool of evidence is naturally accommodated within a MPES model structure through its formal specification of the relationships between data and parameters, which dictate how (*direct*) evidence on the parameters of interest can be supplemented by (*indirect*) information available on arbitrarily complex functions of those parameters. A MPES approach thus presents a number of advantages over direct methods: first, since it incorporates more data,

a MPES model is expected to produce more accurate parameter estimates. Consequently, the inferences it produces correctly reflect the uncertainty surrounding the whole evidence base. Moreover, where there are more data points than estimands, MPES flags any inconsistency potentially affecting a collection of heterogeneous items of data. These conflicts are important to detect, as they may highlight biases in, or misinterpretations of, the data, which can be then addressed.

As an analytical perspective, MPES has in recent years rapidly gained a foot in medical statistics, health technology assessment and epidemiological modeling of infectious diseases like HIV and hepatitis C [Welton and Ades (2005); Goubar et al. (2008); Presanis et al. (2008); Sweeting et al. (2008); De Angelis et al. (2009)]. Since 2005 the UK Health Protection Agency employs a MPES approach to estimate diagnosed and undiagnosed HIV prevalences in the UK using data from routine surveillance and ad-hoc surveys [HIV & STI Department (2005, 2006, 2007, 2008, 2009)]. These evidence synthesis exercises have typically been carried out from a Bayesian perspective, due to its computational convenience, coherent decision-theoretic foundation and automatic synthesis between empirical and prior/subjective information.

This paper describes the development of a Bayesian MPES model to estimate HIV prevalence in different population subgroups and areas across the Netherlands, through reliance on its national surveillance network and an array of regional registries and surveys. The proposed model produces estimates of prevalence, proportions diagnosed and sizes for a number of pre-defined subgroup profiles at risk of HIV infection within the target population of 15- to 70-year old individuals living in the Netherlands in 2007. The paper is organized as follows: Section 2 formally defines the MPES approach adopted. Section 3 describes the body of evidence compiled by the National Institute for Public Health and the Environment in the Netherlands to enable estimation. Section 4 details the MPES model building process, and results are illustrated in Section 5. Model criticism and concluding remarks are outlined in Section 6.

2. The synthesis of evidence. The practice of synthesising evidence from multiple sources, through the combination of direct and/or indirect information from differently designed studies, dates well before dedicated work emerged under an explicit MPES header. Besides the vast body of literature on meta-analytis [see Sutton et al. (2000)], of which MPES represents an extension, a methodological stepping stone in the subject of collating direct and indirect evidence is widely recognized to be the Confidence Profile Method [Eddy and Hasselblad (1992)]. Instances of complex synthesis include, but are not limited to, indirect and mixed treatment comparisons [e.g., Dominici et al. (1999); Song et al. (2003); Lu and Ades (2004); Caldwell, Ades and Higgins (2005)], cross-design synthesis [Drioycour, Silberman and Chelimsky (1993); Benson and Hartz (2000)], hierarchical models [extensively reviewed in, e.g., Sutton et al. (2000); Ades and Cliffe (2002); Whitehead (2002); Gelman and Hill (2007)], Bayesian melding [Poole and

Raftery (2000); Fuentes and Raftery (2005); Alkema, Raftery and Clark (2007)], bias adjustment [Spiegelhalter and Best (2003); Wolpert and Mengersen (2004); Turner et al. (2009)] and multiple/surrogate endpoint synthesis [Berkey et al. (1998); Nam, Mengersen and Garthwaite (2003); Burzykowski, Molenberghs and Buyse (2004)]. These examples attempt to integrate separate sources of evidence to draw inferences that are not only more efficient than those instead obtained from a selective "best data" approach, but also consistent with *all* available information.

Formally, a MPES setup follows closely the characterization of the Confidence Profile Method: assume interest lies in learning about I *basic* parameters $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_I)$, and that for estimation purposes n data points (i.e., sufficient statistics) $\mathbf{y} = (y_1, \ldots, y_n)$ have been separately collected. Any data point may inform either a basic parameter ϑ_i or some *functional* parameter $\psi_j = \psi_j(\boldsymbol{\vartheta})$, $j = 1, \ldots, J$, which can be expressed as a function of known form of the basic parameters. Data unbiasedly reporting on basic parameters are normally referred to as "direct" evidence; samples informing functional parameters are also included in the evidence base, in that they provide "indirect" evidence about their defining basic parameters. Indicating with $L_r(\boldsymbol{\vartheta}; y_r)$ the likelihood contribution from y_r to (elements of) the basic parameter vector $\boldsymbol{\vartheta}$, from the independence of elements in $\mathbf{y}$ the full likelihood model

$$L(\boldsymbol{\vartheta}; \mathbf{y}) = \prod_{r=1}^{n} L_r(\boldsymbol{\vartheta}; y_r) \tag{1}$$

follows.

Within a classical framework, specification of (1) is sufficient to obtain, typically via maximum likelihood, estimates $\hat{\boldsymbol{\vartheta}}$ of the basic parameters and therefore of the J functional parameters $\hat{\boldsymbol{\psi}}_j = \psi_j(\hat{\boldsymbol{\vartheta}})$. Additionally, under a Bayesian perspective, prior (imperfect or even scarce) knowledge around the basic parameters, as expressed through some joint prior distribution $p(\boldsymbol{\vartheta})$, may be updated in the light of the observed data into a posterior distribution $p(\boldsymbol{\vartheta}|\mathbf{y})$ summarizing all information around $\boldsymbol{\vartheta}$ (and thus $\boldsymbol{\psi}$): that is,

$$p(\boldsymbol{\vartheta}|\mathbf{y}) \propto p(\boldsymbol{\vartheta}) L(\boldsymbol{\vartheta}; \mathbf{y}).$$

As in recent MPES modeling work, a Bayesian approach is here proposed since its prior-to-posterior updating mechanism naturally corresponds to the spirit, typical of MPES, of synthesizing multiple items of evidence. Furthermore, the resulting posterior distribution fully reflects both the sampling variability affecting such evidence and the parameter uncertainty surrounding the model.

3. The HIV surveillance network in the Netherlands. In line with Goubar et al. (2008) and Presanis et al. (2008), and compatibly with the socio-demographic coverage and resolution of available data, the population living in the Netherlands

at the end of 2007 was classified by mutually exclusive subgroups and areas of residence. Subgroups are defined as follows:

(1) men who have sex with men (MSM), who have ($\mathrm{MSM}_{\mathrm{STI}}$) or have not ($\mathrm{MSM}_{\overline{\mathrm{STI}}}$) attended a sexually-transmitted infections (STI) clinic in 2007;
(2) intravenous drug users (IDU);
(3) female sex workers (FSW);
(4) heterosexuals attending an STI clinic (STI), further divided into Sub-Saharan Africans ($\mathrm{SSA}_{\mathrm{STI}}$), Caribbeans ($\mathrm{CRB}_{\mathrm{STI}}$) and nonmigrants ($\mathrm{WST}_{\mathrm{STI}}$);
(5) heterosexuals not attending an STI clinic (thus supposedly at low risk of infection), further divided into Sub-Saharan Africans ($\mathrm{SSA}_{\overline{\mathrm{STI}}}$), Caribbeans ($\mathrm{CRB}_{\overline{\mathrm{STI}}}$) and nonmigrants ($\mathrm{WST}_{\overline{\mathrm{STI}}}$).

Let $\mathcal{G}$ denote the set collecting the above subgroups. Broader groups may be defined by merging selected risk categories in $\mathcal{G}$, such as migrants from HIV-endemic areas ($\mathrm{MGR} \doteq \mathrm{MGR}_{\mathrm{STI}} \cup \mathrm{MGR}_{\overline{\mathrm{STI}}}$) either attending ($\mathrm{MGR}_{\mathrm{STI}} \doteq \mathrm{SSA}_{\mathrm{STI}} \cup \mathrm{CRB}_{\mathrm{STI}}$) or not attending ($\mathrm{MGR}_{\overline{\mathrm{STI}}} \doteq \mathrm{SSA}_{\overline{\mathrm{STI}}} \cup \mathrm{CRB}_{\overline{\mathrm{STI}}}$) an STI clinic; likewise, nonmigrant population clusters ($\mathrm{WST} \doteq \mathrm{WST}_{\mathrm{STI}} \cup \mathrm{WST}_{\overline{\mathrm{STI}}}$) may be similarly defined. Here it is assumed that subgroups in $\mathcal{G}$ are ranked by decreasing risk of infection, so that individuals matching multiple risk profiles are allocated into the one highest ranked: for instance, FSW who are at the same time IDU would be classified as IDU.

Group and gender specific estimates of key parameters are derived for three geographic regions: Amsterdam (A), Rotterdam (R) and the rest of the country (O). Let N_r indicate the total population residing in region r, assumed known from census statistics, and $N_{r,g} = \rho_{r,g} N_r$ the unknown (to be estimated) *absolute* size of subgroup $g \in \mathcal{G}$ therein. Basic parameters of interest consist of *relative* subgroup size $\rho_{r,g}$, HIV prevalence $\pi_{r,g}$ and proportion diagnosed with HIV $\delta_{r,g}$ for each combination of 9 subgroups g and 3 regions r. With group-specific estimates being sought by gender (and by STI clinic attendance status for MSM) except for the female-only FSW, the total number of independent basic estimands thus amounts to

$$3 \times (\overbrace{9 \times 2 - 1 - 2}^{\#\{\rho_{r,g}\}} + \overbrace{9 \times 2 - 1}^{\#\{\pi_{r,g}\}} + \overbrace{9 \times 2 - 1}^{\#\{\delta_{r,g}\}}) = 147,$$

given that regional subgroup proportions add up to 1 for each gender: $\sum_g \rho_{r,g} = 1 \; \forall r$.

The HIV surveillance network in place in the Netherlands provides sufficient information to infer basic parameters for most region-subgroup combinations. However, data are partly lacking on proportions diagnosed (notably among migrant subgroups) and more generally outside main urban areas. This lack of information complicates, and in certain cases prevents, estimation of relevant basic parameters, so that a direct approach in the spirit of that outlined in Section 1 would be

inapplicable. On the other hand, an array of registry-based and ad-hoc surveys effectively targeting functional parameters is available to supplement, from a MPES perspective, the available direct data, therefore compensating for the poor evidence on some basic parameters. The overall data set consists of 194 items of data: 65 from Amsterdam, 60 from Rotterdam and 69 from the rest of the Netherlands.

Table 1 details the data collected to directly or indirectly inform HIV epidemic descriptors in the Amsterdam area; the network of surveillance and survey data capturing the HIV epidemic in Rotterdam and the rest of the Netherlands is reported as 6.3. The full array of data shows the extent of coverage of national surveillance and highlights the links between basic and functional parameters. Figure 1 sketches the flow of information within the network of evidence, which is described below.

3.1. *Relative subgroup sizes*. Official figures from Statistics Netherlands (CBS, source o) in Table 1 provide absolute frequencies N_r of regional population sizes. Subgroup sizes for migrants not attending an STI clinic, additionally obtained from CBS, are used to estimate proportions $\rho_{r,g}$ for $g \in \mathrm{MGR}_{\overline{\mathrm{STI}}}$. It should be noted, however, that CBS statistics neither track illegal entries into the country nor distinguish between migrants attending an STI clinic. This implies the following: (i) a downward bias undermining, to an undocumented extent, migrant-related figures; and (ii) the need to decouple STI clinic users from nonusers. Details of how these biases are accommodated within the MPES modeling framework are outlined in Section 4.

Broad MSM subgroup sizes are derived from regional published population studies (Amsterdam and Rotterdam Health Monitors, sources a and r) and random population samples (RNG, source s; Pienter Project, source t). These data, however, enable unbiased estimation only of $\rho_{\mathrm{A,MSM}}$, as they either under- or over-report absolute (and hence relative) sizes of MSM subgroups living outside Amsterdam. These data are then used to inform minimum and maximum values of the underlying subgroup sizes. The municipal registry of opiate and methadone users (source i) presents the same problem when used to inform IDU prevalence in Amsterdam; the size of the IDU population elsewhere is estimated unbiasedly through the Pienter study and a municipal report on addiction and homelessness in Rotterdam (source t).

Direct estimates of subgroup sizes from all STI clinic attending-subgroups are obtained from the national registry of STI clinic consultations (SOAP, source f). Finally, unlinked anonymous (UA) HIV surveys (source q), reporting results from HIV antibody testing of saliva samples from FSW in Amsterdam and Rotterdam, are available to directly inform $\rho_{r,\mathrm{FSW}}$ for $r = \mathrm{A}, \mathrm{R}$. A published study on FSW in The Hague (source u) is utilized to inform a range for the frequency of FSW in the rest of the country.

TABLE 1

Evidence supporting HIV prevalence estimation in Amsterdam ($N_m = 284{,}002$, $N_f = 284{,}067$); letters in brackets link to data sources as detailed in Section 3

		Basic parameters			**Functional parameters**	
Group	**Subgroup**	ρ	π	δ	$\pi\delta$	μ[a]
MSM	STI	$2{,}495/N_m = 0.009$ (f)	$606/2{,}723 = 0.223$[b] (f)	$79/85 = 0.929$ (g)		
	All	$73/776 = 0.094$ (a)		$48/547 = 0.088$ (s)		$2{,}827$[c] (h)
IDU	M	$(720\text{–}1{,}120)/N_m = 0.003\text{–}0.004$ (i)	$45/167 = 0.269$[d] (k)	$31/45 = 0.689$[e] (k)	$37/196 = 0.189$ (p)	99[f] (h)
	F	$(180\text{–}280)/N_f = 6.34E\text{–}04\text{–}9.85E\text{–}04$ (i)	$6/30 = 0.200$[d] (k)	$3/6 = 0.500$[e] (k)	$20/88 = 0.227$ (p)	64[f] (h)
FSW	F	$7{,}440/N_f = 0.026$ (q)	$3/148 = 0.020$ (q)	$0/3 = 0$ (q)		
WST_{STI}	M	$5{,}702/N_m = 0.020$ (f)	$10/5{,}526 = 0.002$[b] (f)	$10/(N_{g,m}\pi_{g,m})$ (f)		
	F	$6{,}586/N_f = 0.023$ (f)	$7/6{,}402 = 0.001$[b] (f)	$7/(N_{g,f}\pi_{g,f})$ (f)		
SSA_{STI}	M	$261/N_m = 0.001$ (f)	$7/237 = 0.030$[b] (f)	$7/(N_{g,m}\pi_{g,m})$ (f)		
	F	$158/N_f = 0.001$ (f)	$10/151 = 0.066$[b] (f)	$10/(N_{g,m}\pi_{g,f})$ (f)		
CRB_{STI}	M	$899/N_m = 0.003$ (f)	$4/855 = 0.005$[b] (f)	$4/(N_{g,m}\pi_{g,m})$ (f)		
	F	$771/N_f = 0.003$ (f)	$6/753 = 0.008$[b] (f)	$6/(N_{g,f}\pi_{g,f})$ (f)		
$SSA_{\overline{STI}}$	M	$9{,}434/N_m = 0.033$[g] (o)	$1/129 = 0.008$ (l)			173[h] (h)
	F	$8{,}233/N_f = 0.029$[g] (o)		$0/50 = 0$ (l)		252[h] (h)
$CRB_{\overline{STI}}$	M	$31{,}200/N_m = 0.110$[g] (o)	$1/215 = 0.005$ (l)			137[h] (h)
	F	$36{,}468/N_f = 0.128$[g] (o)	$1/252 = 0.004$ (l)			111[h] (h)
STI	M		$8/2{,}135 = 0.004$ (g)	$2/8 = 0.250$[e] (g)		
	F		$7/2{,}580 = 0.003$ (g)	$3/7 = 0.429$[e] (g)		

TABLE 1
(*Continued*)

Group	Subgroup	Basic parameters			Functional parameters	
		ρ	π	δ	$\pi\delta$	μ[a]
Mixed	M					145[i] (h)
	F					207[i] (h)
Pregnant women	Nonmigrant		4/13,097 = 1E–04 (m)	3/4 = 0.750 (m)		
	Migrant		27/3,413 = 0.008 (m)	21/27 = 0.778 (m)		

[a]Total Amsterdam residents for each gender are estimated (source: CBS) at 3,522 (M) and 660 (F), also including 141 and 26 cases of unknown exposure respectively (source: SHM).

[b]Data inform minimum prevalences, due to 278/2,495 = 0.111 (STI MSM), 182/5,702 = 0.032 (M WST), 186/6,586 = 0.028 (F WST), 26/261 = 0.100 (M SSA), 9/158 = 0.057 (F SSA), 45/899 = 0.050 (M CRB) and 19/771 = 0.025 (F CRB) STI users opt-out fractions.

[c]Registered cases are underestimated by a 91/98 = 0.929 fraction (source: Schorer Monitor).

[d]Data inform maximum prevalence.

[e]Data inform minimum proportions diagnosed.

[f]Registered cases are underestimated by uninformed gender-specific fractions.

[g]Counts also include STI clinic users, but exclude illegal immigrants.

[h]Recorded infections include both STI clinic attending and nonattending immigrants.

[i]Mixture of respectively male and female registered infections for IDU, FSW (F only), nonmigrant STI clinic users and other.

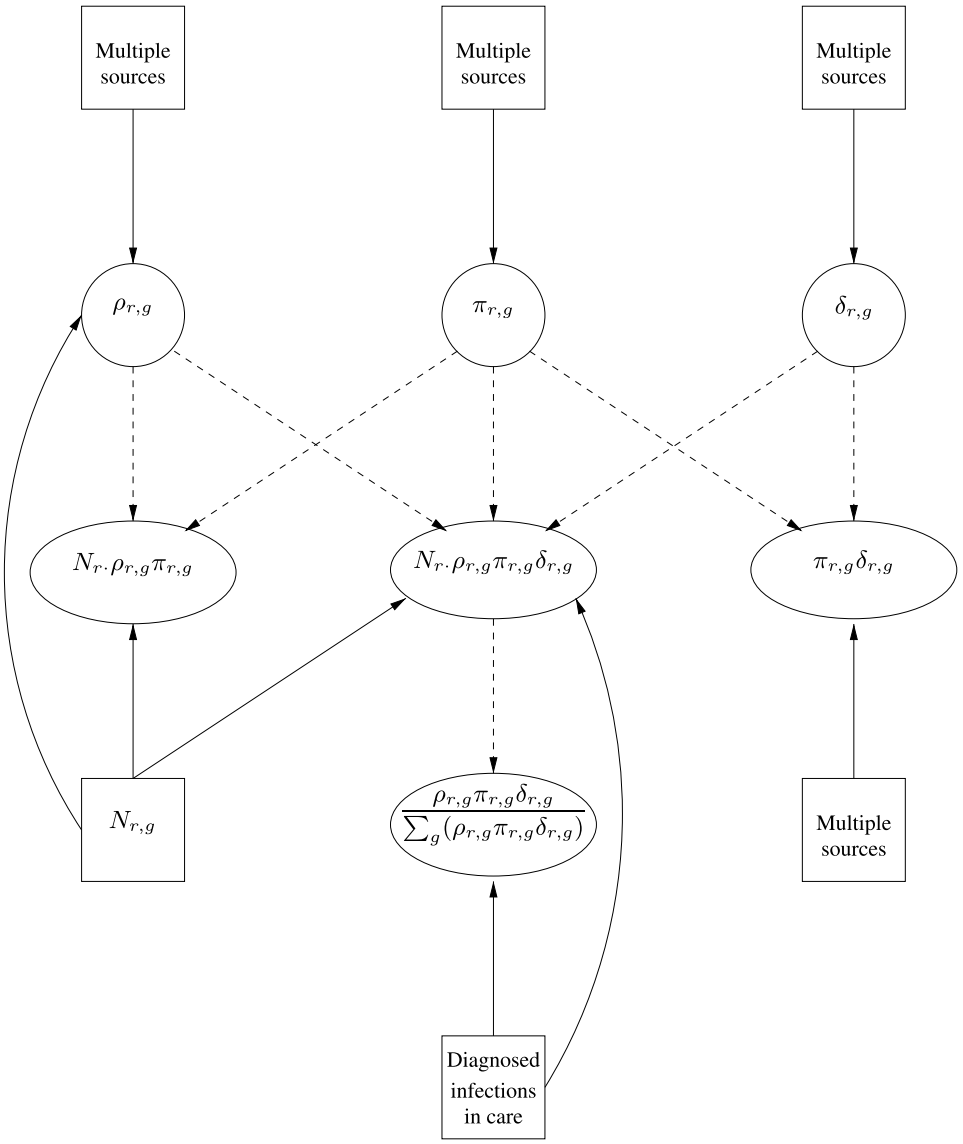

FIG. 1. *Schematic representation of the evidence network informing epidemiological parameters in the MPES HIV model*: *different samples* (*squares*) *provide data-based information* (*solid arrows*) *around key basic* (*circles*) *and functional* (*ellipses*) *estimands*, *where the latter are functionally related* (*dashed arrows*) *to the former* (*i.e.*, $\rho_{r,g}$, $\pi_{r,g}$ *and* $\delta_{r,g}$).

3.2. *HIV prevalences*. Evidence around HIV prevalence is more fragmented than that on subgroup sizes. Information relating to MSM individuals is sparse, with the only direct source of evidence on $\pi_{\mathrm{A,MSM}}$ consisting of the Amsterdam Cohort Study (source s). Information outside urban concentrations is indirectly

derived through data on *diagnosed* prevalence (i.e., $\pi_{O,MSM}\delta_{O,MSM}$) from the Schorer Monitor (source e), that is, the national institute responsible for the coordination of primary HIV/STI prevention policies targeting MSM in the Netherlands, and the Pienter Project. Since these sources are biased downward and upward, respectively, they provide upper and lower limits for the product $\pi_{O,MSM}\delta_{O,MSM}$. Moreover, the Pienter data set also supplies information on diagnosed low-risk prevalence outside main urban areas ($\pi_{O,WST_{\overline{STI}}}\delta_{O,WST_{\overline{STI}}}$).

Separate UA surveys carried out across the country allow direct estimation of HIV prevalences among IDU (source k), FSW (sources k and q) and non-STI clinic attending migrant subgroups (source l). As particularly the UA survey covering the $CRB_{\overline{STI}}$ population is suspected to suffer from underreporting bias, this is specifically utilized to inform a lower bound for the corresponding prevalence parameter.

SOAP records are likely to underestimate HIV prevalence, due to an opt-out policy in place on HIV testing in STI clinics across the Netherlands. Information on HIV prevalence in STI clinic users is limited to those individuals actually submitting to HIV testing, while only information on attendance is retained from the remaining patients. Since reluctance to submit to HIV testing is indicative of a higher risk of HIV infection [Van der Bij et al. (2008)], it is reasonably assumed that STI clinic users opting out of HIV testing are more likely to be HIV positive. Details on how opt-in and opt-out contributions to HIV prevalence parameters are decoupled and modeled are given in Section 4.1.1. UA surveys in Amsterdam (DWAR; source g) and Rotterdam (ROTan; source v) are also included into the network of evidence, as they inform HIV prevalence among all STI clinic users (regardless of ethnicity) in urban areas.

Last, very little information exists on HIV prevalence affecting low-risk subgroups. Two indirect anonymized sources can be identified: the national antenatal screening program (source m), which monitors seroprevalence in pregnant women across the Netherlands in 2007; and the national registry of blood donors (Sanquin Foundation, source w), which keeps records of HIV infections among new and regular donors in the Netherlands in 2007.

Data on blood donors are unlikely to provide unbiased evidence on HIV prevalence in the low-risk group, as blood donors are at especially low risk of contracting HIV. Moreover, information from Sanquin is not categorized by either gender or region, so it captures HIV prevalence at a very coarse subgroup level. Equally problematic, data from the national antenatal screening program, which are broadly classified by ethnicity, provide evidence on HIV prevalence on a population subgroup not explicitly contemplated by the model, but rather resulting from a mixture of female subgroups in $\mathcal{G}$. An assumption of equal representativeness, in terms of risk group composition, of pregnant women with respect to the wider female population is introduced to allow modeling of this indirect (“mixed”) evidence [Presanis et al. (2008)].

3.3. *Proportions diagnosed with HIV.* Many data sources already informing HIV prevalence also provide evidence on the extent of disease diagnosis within the target subgroups. This information, however, is markedly sparse: sample sizes are small and coverage does not extend to all region-subgroup combinations, notably excluding MSM, $\mathrm{MGR}_{\overline{\mathrm{STI}}}$ and $\mathrm{WST}_{\overline{\mathrm{STI}}}$.

Biases also undermine parts of the evidence base. For instance, UA evidence is known to underestimate $\delta_{r,\mathrm{IDU}}$, due to the especially hard-to-reach nature of this subgroup. Data are therefore assumed to inform a lower bound for corresponding proportions. Similarly, DWAR records on all HIV infections diagnosed in STI clinics in Amsterdam, due to intrinsic design limitations, provide a downward-biased estimate of $\delta_{\mathrm{A,STI}}$.

3.4. *Diagnosed HIV infections.* The HIV Monitoring Foundation (SHM; source h) compiles and maintains a registry of (almost[2]) all diagnosed HIV cases in specialized care in the Netherlands, classified by socio-demographic factors. Absolute counts from the relevant registry inform the regional risk group composition of (mixtures of) prevalent HIV diagnoses: namely,

$$\mu_{r,g} = N_r \rho_{r,g} \pi_{r,g} \delta_{r,g}, \tag{2}$$

which form a set of functional parameters (see Figure 1) involving all basic parameters of interest. The risk subgroup classification adopted by SHM does not coincide with that in the MPES model, since it includes mixed pregnant women, SSA, CRB, WST and unclassified individuals (none in $\mathcal{G}$) as well as MSM and IDU. Additionally, cross-matching with records from the Schorer Monitor reveals an underreporting bias affecting SHM records on prevalent MSM cases diagnosed across the country. Finally, SHM is also known to underreporting IDU cases in Amsterdam.

4. The MPES model structure. The above array of data on HIV prevalence in the Netherlands is synthesized in a Bayesian statistical model relying upon suitably chosen standard distributions, in the spirit of case studies already documented in the literature [e.g., Ades and Sutton (2006); Goubar et al. (2008); Presanis et al. (2008)].

4.1. *Sampling distributions.* Count data $x_{r,g}$ from a census- or survey-type study of fixed size $n_{r,g}$ on subgroup $g \in \mathcal{G}$ in region r (like, e.g., SOAP records on HIV diagnoses in STI clinics across the Netherlands) and characterized by a generic probability parameter $\lambda_{r,g}$ are naturally modeled via Binomial likelihoods

$$x_{r,g} | \lambda_{r,g} \sim \mathrm{Bin}(n_{r,g}, \lambda_{r,g}).$$

[2]In reality, not every diagnosed HIV case makes timely (if any) contact with national treatment facilities.

The total number of diagnosed HIV cases $m_{r\cdot}$ on SHM record as in care in region r is assumed to follow a Poisson distribution with regional rate $\mu_{r\cdot} = \sum_g \mu_{r,g}$, with $\mu_{r,g}$ defined as in (2). At the same time, the subgroup sizes $m_{r,g}$ within each region are assigned a Multinomial distribution with size parameter $m_{r\cdot}$ and probability vector $\boldsymbol{\xi}_r = (\xi_{r,g}; g \in \mathcal{G})$ with $\xi_{r,g} = \frac{\mu_{r,g}}{\mu_{r\cdot}}$, so that

$$m_{r\cdot}|\mu_{r\cdot} \sim \text{Poi}(\mu_{r\cdot}),$$

$$m_{r,g}|m_{r\cdot}, \boldsymbol{\xi}_r \sim \text{Multin}(m_{r\cdot}, \boldsymbol{\xi}_r).$$

In practice, as explained in Section 2, interest does not always lie in the (often functional) λ or ξ parameters, but rather in the basic parameters they subsume in their definition. The relationship between basic and functional parameters is formally determined by the type of mixed, biased or otherwise indirect evidence available. Examples are illustrated in the following sections.

4.1.1. *Mixed subgroup modeling*. By classifying individuals into risk groups other than those being modeled, registry-type records provide information on proportions of diagnosed cases in each risk category (i.e., ratios of the form $\mu_{r,g}/\sum \mu_{r,g}$, rather than $\delta_{r,g}$), possibly on suitably defined mixtures of subgroups in $\mathcal{G}$.

This is, for instance, the case with SHM which, as outlined in Section 3.4, poses a number of modeling challenges. Unclassified individuals within its records are distributed proportionately across modeled risk groups, in line with Presanis et al. (2008). Additionally, records on mixed migrant subgroups are modeled by STI clinic attendance status via the likelihood term

$$m_{r,g}|m_{r\cdot}, \xi_{r,g} \sim \text{Bin}(m_{r\cdot}, \xi_{r,g})$$

for $g \in \{\text{SSA}, \text{CRB}\}$, where

$$\xi_{r,\text{SSA}} = \frac{\mu_{r,\text{SSA}_{\text{STI}}} + \mu_{r,\text{SSA}_{\overline{\text{STI}}}}}{\mu_{r\cdot}}$$

and

$$\xi_{r,\text{CRB}} = \frac{\mu_{r,\text{CRB}_{\text{STI}}} + \mu_{r,\text{CRB}_{\overline{\text{STI}}}}}{\mu_{r\cdot}},$$

where $\xi_{r,g}$ denotes the fraction of reported regional HIV diagnoses in the $\text{SSA} = \text{SSA}_{\text{STI}} \cup \text{SSA}_{\overline{\text{STI}}}$ and $\text{CRB} = \text{CRB}_{\text{STI}} \cup \text{CRB}_{\overline{\text{STI}}}$ subgroups.

As explained in Section 3.2, estimation of the HIV prevalence $\pi_{r,g}^{\text{out}}$ unobserved in subgroups opting out of HIV testing requires some modeling assumption. Here it is assumed that prevalence among STI clinic users declining an HIV test would be at least that of patients with the same risk profile but not submitting to the test [Van der Bij et al. (2008)]. This is formalized for $g \in \{\text{SSA}_{\text{STI}}, \text{CRB}_{\text{STI}}, \text{WST}_{\text{STI}}\}$

through the decomposition

$$\begin{aligned}\pi_{r,g} &= \frac{\#\text{infections}}{N_r \rho_{r,g}} \\ &= \frac{\#\text{opt-in infections}}{N_r \rho_{r,g}} + \frac{\#\text{opt-out infections}}{N_r \rho_{r,g}} \\ &= \pi_{r,g}^{\text{in}} + \pi_{r,g}^{\text{out}},\end{aligned}$$

where the HIV prevalence $\pi_{r,g}^{\text{in}}$ among those submitting to the diagnostic test is the parameter actually being captured by regional SOAP statistics.

In Section 3.2 it was also mentioned that, in addition to SOAP, DWAR and ROTan provide independent information on HIV prevalence and proportions diagnosed among STI clinic users of ethnicity in Amsterdam and Rotterdam, respectively. These aggregate-level data are retained into the model to estimate corresponding parameters for $r \in \{A, R\}$, that is,

$$\tilde{\pi}_{r,\text{STI}} = \frac{\sum_{g \in \text{STI}} \rho_{r,g} \pi_{r,g}}{\sum_{g \in \text{STI}} \rho_{r,g}},$$

$$\tilde{\delta}_{r,\text{STI}} = \frac{\sum_{g \in \text{STI}} \rho_{r,g} \pi_{r,g} \delta_{r,g}}{\sum_{g \in \text{STI}} \rho_{r,g} \pi_{r,g}}.$$

Data on low-risk women from the national antenatal screening program are dealt with similarly.

4.1.2. *Bias modeling*. In general, any sample estimating some basic parameter ϑ with bias $\Delta \neq 0$ can be regarded as providing indirect evidence, on some suitable scale, on ϑ through the functional parameter $\psi(\vartheta) = \vartheta + \Delta$. An example of biased evidence from the case study at hand is offered by CBS immigration records which, as explained in Section 3.1, do not include illegal entries and are not classified by STI clinic attendance status. Letting $\gamma_{r,g}$ indicate the proportion of *legal* migrants in region r with ethnicity $g \in \{\text{SSA}, \text{CRB}\}$, CBS provides unbiased information on the relative size $\tilde{\rho}_{r,g}$ of immigrant subpopulations legally living in the Netherlands: in functional terms,

$$\tilde{\rho}_{r,\text{SSA}} = \gamma_{r,\text{SSA}} (\rho_{r,\text{SSA}_{\text{STI}}} + \rho_{r,\text{SSA}_{\overline{\text{STI}}}})$$

and

$$\tilde{\rho}_{r,\text{CRB}} = \gamma_{r,\text{CRB}} (\rho_{r,\text{CRB}_{\text{STI}}} + \rho_{r,\text{CRB}_{\overline{\text{STI}}}}).$$

Since no auxiliary data are available to inform the number of illegal immigrants, either overall or by ethnicity, it is assumed that the proportions $1 - \gamma_{r,g}$ of SSA-born (CRB-born) illegal migrants in each region ranged between 10% and 20%[3] (0% and 5%) across the country [van Veen (2009)].

[3]Unlike immigrants from Sub-Saharan African countries, most individuals from the Caribbean are actually entitled lawful entry into the Netherlands.

Furthermore, downward-biased SHM records on MSM and IDU (see Section 3.4) are modeled by dividing parameters $\mu_{r,\mathrm{MSM}}$ and $\mu_{\mathrm{A,IDU}}$ with underreporting proportions, in turn separately estimated from the Schorer Monitor and Amsterdam Cohort Study, respectively.

4.2. *Prior distributions*. Within a Bayesian framework, parameters of a statistical model are given some prior probability distribution reflecting the imperfect knowledge around them. In the present work basic parameters are typically assigned vague prior distributions. At times, lack of information around a significant number of parameters required introducing more structured priors to express either known constraints or expert opinion.

4.2.1. *Parameter constraints*. A relatively simple example of the need for a constraining prior distribution was offered by data from the Sanquin Foundation. As pointed out in Section 3.2, HIV prevalence among blood donors is expected to be significantly lower than among the wider $\mathrm{WST}_{\overline{\mathrm{STI}}}$ subgroup. This is accommodated within the model by assuming that information from blood donors provides a *lower bound* $\pi^{\mathrm{L}}_{\mathrm{WST}_{\overline{\mathrm{STI}}}}$ for $\mathrm{WST}_{\overline{\mathrm{STI}}}$ HIV prevalence at a national level, where

$$\pi^{\mathrm{L}}_{\mathrm{WST}_{\overline{\mathrm{STI}}}} \leq \pi_{r,\mathrm{WST}_{\overline{\mathrm{STI}}}} \qquad \forall r.$$

These assumptions are introduced to model any sample that is only known to be biased, but without any additional information as to the extent of the bias. These data points are annotated in detail in Table 1. In all cases the modeling structure is naturally completed by Uniform priors defined over appropriate bounds, as was done in Section 4.1.1 with the parameters $\pi^{\mathrm{out}}_{r,g}$. It is then assumed that $\pi_{r,\mathrm{WST}_{\overline{\mathrm{STI}}}} \leq \min_{g\in\mathcal{G}} \pi_{r,g}$, implying that $\mathrm{WST}_{\overline{\mathrm{STI}}}$ prevalences should not exceed that exhibited by any other subgroup in the same region. Last, diagnosed HIV prevalences in any STI clinic-attending subgroup are conservatively assumed to be at least 20%, to prevent unrealistically low parameter estimates.

4.2.2. *Expert opinion*. Sometimes subjective prior distributions were elicited from collaborating epidemiologists. This was the case with parameters characterizing low-risk individuals and, more broadly, outside Amsterdam and Rotterdam (see Section 3.2). Similarly to Goubar et al. (2008) and Presanis et al. (2008), let $\pi^{s}_{r,g}$ denote HIV prevalence among male and female ($s = \mathrm{m}, \mathrm{f}$) individuals with risk profile $g \neq \mathrm{MSM}, \mathrm{FSW}$ in region r; the male-to-female prevalence log-odds ratio is then defined as

$$\eta_{r,g} = \operatorname{logit} \pi^{\mathrm{m}}_{r,g} - \operatorname{logit} \pi^{\mathrm{f}}_{r,g}.$$

A two-stage hierarchical model is formulated for prevalence log-odds ratios: in the first level these are pooled across subgroups to produce shrunk estimates $\bar{\eta}_r$; the

second then pools regional log-odds ratios $\bar{\eta}_r$ across the Netherlands to derive an overall estimate $\bar{\bar{\eta}}$. The complete model specification is thus given by

$$\begin{aligned} \eta_{r,g}|\bar{\eta}_r, \sigma_r &\sim \mathcal{N}(\bar{\eta}_r, \sigma_r^2), \\ \bar{\eta}_r|\bar{\bar{\eta}}, \tau, \sigma_r &\sim \mathcal{N}(\bar{\bar{\eta}}, \tau^2). \end{aligned} \tag{3}$$

Vague hyperpriors on the national log-odds ratio ($\bar{\bar{\eta}}$) and on the regional (σ_r) as well as national (τ) standard deviations, respectively, measuring the degree of between-subgroup and between-region heterogeneity among prevalence log-odds ratios, complete the hierarchical model structure.

While absolute HIV prevalences should not be reasonably expected to be distributed homogeneously across subgroups within each region, corresponding male-to-female log-odds ratios can instead be more plausibly thought of as arising from a common region-specific distribution, as implied by (3). Shrinkage toward a regional mean allows information available around some subgroups to supplement that around others poorly informed; see, for example, Gelman and Hill (2007) and annotated bibliography for a comprehensive review of the concept of "borrowing strength."

Finally, expert opinion helps: to categorize individuals from HIV-endemic countries by legal entry status when modeling respective subgroup sizes (as described in Section 4.1.2); and to infer HIV prevalences $\pi_{r,g}^{\text{out}}$ among STI clinic-attending subgroups declining HIV testing (as seen in Section 4.1.1). Additional assumptions relating to migrants and STI clinic users are motivated by the expectation of a higher proportion of HIV diagnoses among STI clinic users, relative to nonusers with the same ethnicity and sexual orientation. In more formal terms,

$$\delta_{r,\text{SSA}_{\text{STI}}} \geq \delta_{r,\text{SSA}_{\overline{\text{STI}}}},$$

$$\delta_{r,\text{CRB}_{\text{STI}}} \geq \delta_{r,\text{CRB}_{\overline{\text{STI}}}}$$

and

$$\delta_{r,\text{MSM}_{\text{STI}}} \geq \delta_{r,\text{MSM}_{\overline{\text{STI}}}}.$$

In a similar fashion, the MPES model also includes the constraints

$$\delta_{r,\text{CRB}_{\text{STI}}} \geq \delta_{r,\text{SSA}_{\text{STI}}}$$

and

$$\delta_{r,\text{CRB}_{\overline{\text{STI}}}} \geq \delta_{r,\text{SSA}_{\overline{\text{STI}}}}.$$

The above are motivated by a better integration in the Netherlands of Caribbean migrants compared to Sub-Saharan Africans, who tend to be less familiar with HIV treatment facilities and access routes to health-care services [van Veen et al. (2005)].

4.3. *Model appraisal*. Recalling notation introduced in Section 2, the standardized deviance of a particular model is defined as

$$D(\mathbf{y}, \boldsymbol{\vartheta}) = -2 \ln \frac{L(\boldsymbol{\vartheta}; \mathbf{y})}{L(\hat{\boldsymbol{\vartheta}}; \mathbf{y})},$$

where $L(\hat{\boldsymbol{\vartheta}}; \mathbf{y})$ is the likelihood of the saturated model (where the number of parameters equals the number of observations), evaluated at the maximum-likelihood estimate of $\boldsymbol{\vartheta}$.

The use of the posterior mean deviance

$$\bar{D}(\mathbf{y}) = \mathbb{E}[D(\mathbf{y}, \boldsymbol{\vartheta})|\mathbf{y}] \tag{4}$$

has been suggested by Spiegelhalter et al. (2002) as a measure of goodness of fit: under standard regularity conditions, if the model is true, $\mathbb{E}[\bar{D}(\mathbf{y})] \approx n$, so that, in particular, $\mathbb{E}[\bar{D}(\mathbf{y})] \gg n$ would be suggestive of lack of model fit. As the sampling distribution of $\bar{D}(\mathbf{y})$ is not well understood [Seaman, De Angelis and Presanis (2011)], this idea is here used informally to identify conflicting information on specific parameters (see Section 6.1), through the decomposition $\bar{D}(\mathbf{y}) = \sum_{i=1}^{n} \bar{D}(\mathbf{y}_i)$ of the deviance (4) into the individual contributions $\bar{D}(\mathbf{y}_i)$ made by each data point $\mathbf{y}_i, i = 1, \ldots, n$. The fact that, for a true model, $\mathbb{E}[\bar{D}(\mathbf{y}_i)] \approx 1$, can be used to identify specific data points responsible for a potential lack of fit and to investigate the likely inconsistency in the information they provide.

5. Results. The MPES model was fitted to the collection of surveillance and survey data via McMC simulation using the `WinBUGS` statistical package [Lunn et al. (2000)]. The code and data required to produce model estimates are provided as 6.3. The sampling algorithm was started at three independent initial states, with convergence ascertained by both visual and formal diagnostic means [Gelman and Rubin (1992)] after 30,000 iterations. After thinning, a 30,000-sized sample from the full posterior distribution was subsequently retained for drawing inferences.

5.1. *Model inferences*. Point and interval estimates around basic population and HIV-related parameters are presented by risk group for the Amsterdam area in Table 3, together with predicted number of HIV infections classified by diagnosis status; inferences for the remaining georgraphic areas across the Netherlands are presented as 6.3.

Most posterior distributions tend to concentrate around parameter values regarded as plausible by the collaborating epidemiologists, usually with a reasonable level of accuracy, given the uncertainty affecting the underlying data. In particular, predicted numbers of prevalent ($N_r \rho_{r,g} \pi_{r,g}$) and undiagnosed ($N_r \rho_{r,g} \pi_{r,g} (1 - \delta_{r,g})$) cases—the key inputs to health-care decision-making—appear in line with expectations and concur with results from alternative analytical frameworks [van Veen et al. (2011)].

Estimates in general reflect the varying accuracy of regional collection networks, as well as local patterns of subgroup composition: the precision of inferences can be seen to broadly decrease when moving from urban concentrations (like Amsterdam, Table 3) into smaller subgroups/areas across the Netherlands (see 6.3). This is, for instance, the case with estimated proportions of prevalent MSM cases diagnosed outside Amsterdam and Rotterdam ($\hat{\delta}_{\mathrm{O,MSM}}$, see 6.3), which are also significantly lower than those from the two main cities (see, e.g., $\hat{\delta}_{\mathrm{A,MSM}}$ from Table 3). Due to the lack of studies targeting $\delta_{\mathrm{O,MSM}}$, this can only be inferred indirectly from diagnosed HIV cases ($\mu_{\mathrm{O,MSM}}$) and diagnosed prevalence ($\pi_{\mathrm{O,MSM}}\delta_{\mathrm{O,MSM}}$), the latter in turn being informed by two biased studies. The uncertainty around resulting MPES estimates is just a consequence of the synthesis between such scarce evidence and the mild ranking assumptions on $\delta_{r,g}$ detailed in Section 4.2.2.

Estimated proportions of prevalent IDU cases diagnosed in Amsterdam ($\delta_{\mathrm{A,IDU}}$ from Table 3) are higher than elsewhere in the Netherlands ($\delta_{r,\mathrm{IDU}}$ for $r \neq A$ from 6.3). Critical appraisal of $\hat{\delta}_{\mathrm{A,IDU}}$ is complicated by the large number of data sources involved. Direct data-based estimates (Table 1) can be seen to be much lower than those produced by the MPES model (Table 3). At the same time, however, the number $\mu_{\mathrm{A,IDU}}$ of diagnosed HIV infections, which suffers from underreporting (see Section 3.4), disagrees with direct information separately available on its building blocks $\pi_{\mathrm{A,IDU}}$ and $\delta_{\mathrm{A,IDU}}$ (see Section 3.3). On the other hand, evidence listed in Table 1 on $\pi_{\mathrm{A,IDU}}$ and $\mu_{\mathrm{A,IDU}}$, while biased, is overall firmer than that around $\delta_{\mathrm{A,IDU}}$ and therefore weighs more in the synthesis process. In broad terms, records on diagnosed infections can be seen as an "anchor" to the balance between $\pi_{\mathrm{A,IDU}}$ and $\delta_{\mathrm{A,IDU}}$: by definition, the same number of diagnosed HIV infections $\mu_{r,g}$ can be obtained with different combinations of prevalent cases $N_r\rho_{r,g}\pi_{r,g}$ and fractions diagnosed $\delta_{r,g}$. Since available information allows for more accurate estimation of $\pi_{\mathrm{A,IDU}}$ compared to $\delta_{\mathrm{A,IDU}}$, the MPES model reconciles conflicting evidence around $\mu_{\mathrm{A,IDU}}$ by favoring larger estimates of the more uncertain $\delta_{\mathrm{A,IDU}}$ over correspondingly lower values of $\pi_{\mathrm{A,IDU}}$.

6. Discussion. Recent applied work has consolidated the role of MPES as a modeling framework for the estimation of epidemiological indicators of infectious diseases [Ades (2003); Welton and Ades (2005); Ades et al. (2006, 2008); Goubar et al. (2008); Presanis et al. (2008); De Angelis et al. (2009)]. This has paved the way for governmental institutions (e.g., the Medical Research Council, the National Institute for Health and Clinical Excellence and the Health Protection Agency in the UK) and international bodies (e.g., the World Health Organization and UNAIDS) to increasingly rely on formal evidence synthesis as an analytic tool to advance epidemiological understanding and support medical decision-making. The present work represents an additional step toward expansion of the range of

applicability of the MPES approach, as it illustrates the experience of HIV prevalence estimation in the Netherlands, a western European country with a concentrated HIV epidemic and reasonably consolidated and accessible HIV specialist care.

While relatively extensive in terms of geographic and behavioral coverage, the array of surveillance and survey data made available by the Dutch National Institute for Public Health and the Environment required a comprehensive reappraisal at an evidence synthesis stage. Problems in the evidence body were identified through an informal use of deviance statistics in terms of conflicting, biased or insufficient data on certain region-subgroup combinations. Inconsistencies thus detected were mostly resolved by using additional data and/or expert beliefs provided by collaborating epidemiologists. Nevertheless, some evidence of conflict remained, as indicated by the overall mean posterior deviance of 258.139, compared to a total of 186 observations.[4] This conflict is mainly around evidence informing HIV prevalence among migrant women in the rest of the Netherlands, for which collection of further information was consequently recommended. Ultimately, model estimates broadly met the expectations of the pool of epidemiologists involved in the case study.

6.1. *The role within MPES of direct and indirect evidence*. In general, the process of amalgamating all knowledge available is *expected* to produce more accurate inferences than those resulting from a partial or no synthesis. However, as anticipated in Section 1, the availability of multiple evidence sources on given parameters can lead to the utilization of discrepant, if not conflicting, items of information. These discrepancies typically originate from an incorrect interpretation of what the data represent (e.g., unrecognized biases), which consequently are inadequately modeled. If these inconsistencies remain unresolved, MPES inferences may be less accurate than those obtained from using direct information alone. This is because MPES estimates arise as a compromise between estimates separately informed by direct and indirect evidence only, the more precise of the two weighting more in the balance. The MPES approach allows resolution of inconsistencies by explicitly modeling the conflicting items of evidence [e.g., by accounting for biases in the data, as in Ades and Cliffe (2002); Presanis et al. (2008)]. In practice, this is achieved via an interactive reappraisal process, involving the statisticians and collaborating epidemiologists, of the data sources flagged by the MPES model as conflicting. Ultimately, any unresolved conflict of evidence on some parameter would be symptomatic of the need for additional information—either in the form of field data or of expert opinion—to be collected.

[4]This differs from the nominal total of 194 (see Section 3) in that it excludes overly sparse samples—like those from SOAP leading to 0 or 1 maximum likelihood estimates of some $\delta_{r,g}$ parameters—not meeting the regularity conditions mentioned in Section 4.3.

TABLE 2

Posterior mean deviances (computed from direct items of evidence only) and posterior medians with 95% credibility intervals for selected basic parameters, obtained from modeling available direct evidence only, indirect evidence only and all evidence

		Inferences (%)		
Parameter	**Deviance**	**Direct**	**Indirect MPES**	**Full MPES**
$\pi^{\mathrm{m}}_{\mathrm{R,CRB_{STI}}}$	1.664	1.194 (0.416, 2.598)[a]	1.393 (1.000, 2.194)	1.492 (1.014, 2.451)
$\pi^{\mathrm{f}}_{\mathrm{R,IDU}}$	1.887	14.489 (8.389, 22.488)	7.461 (4.198, 13.250)	10.060 (6.612, 15.030)
$\rho_{\mathrm{A,MSM}}$	2.102	9.372 (7.456, 11.557)	13.750 (10.240, 21.240)	10.750 (9.173, 12.560)
$\pi^{\mathrm{f}}_{\mathrm{O,CRB}_{\overline{\mathrm{STI}}}}$	8.463	1.515 (0.529, 3.292)	0.191 (0.164, 0.222)	0.194 (0.166, 0.226)
$\pi^{\mathrm{m}}_{\mathrm{O,SSA}_{\overline{\mathrm{STI}}}}$	9.346	1.116 (0.162, 3.658)[a]	3.838 (2.228, 6.214)	3.926 (2.227, 6.345)

[a]Direct evidence is known to be up- or down-ward biased.

As explained in Section 4.3, examination of the contribution $\bar{D}(\mathbf{y}_i)$ provided by each data point $\mathbf{y}_i$ to the posterior mean deviance (4) allows identification of conflicts between direct and indirect evidence: for a given item of direct evidence, the farther from 1 its contribution to (4), the more marked the discrepancy of the information it provides on a particular parameter with the remaining available evidence. This is illustrated in Table 2, which for selected basic parameters reports the posterior mean deviance contributions to (4), based on their respective direct-only evidence, alongside corresponding inferences obtained from separately utilizing direct, indirect and full information. The benefit of full evidence synthesis in the presence of broadly agreeing sources of information is made obvious by estimates of $\pi^{\mathrm{m}}_{\mathrm{R,CRB_{STI}}}$ and $\pi^{\mathrm{f}}_{\mathrm{R,IDU}}$: respective MPES inferences combine direct and indirect evidence, whose consistency is highlighted by deviance statistics close to 1, to produce narrower credibility intervals than those arising from a direct approach. Moderately discrepant information is instead reconciled within the MPES model through a compromise between direct- and indirect-only inferences: this, as in the case of $\rho_{\mathrm{A,MSM}}$, yields a credibility interval not significantly narrower than its direct counterpart, since it conveys not only the uncertainty within, but also the variability between, the items of evidence it involves.

The same rationale applies to MPES inferences on parameters informed by conflicting evidence: similarly to $\rho_{\mathrm{A,MSM}}$, MPES credibility intervals around $\pi^{\mathrm{f}}_{\mathrm{O,CRB}_{\overline{\mathrm{STI}}}}$ and $\pi^{\mathrm{m}}_{\mathrm{O,SSA}_{\overline{\mathrm{STI}}}}$ offer a compromise between the direct and indirect information separately contributing to their estimation. The synthesized inferences, however, are clearly less accurate than their respective direct versions, due to the extent of the inconsistency undermining the information on $\pi^{\mathrm{f}}_{\mathrm{O,CRB}_{\overline{\mathrm{STI}}}}$ and $\pi^{\mathrm{m}}_{\mathrm{O,SSA}_{\overline{\mathrm{STI}}}}$, as also indicated by the correspondingly large deviance statistics. In this case, while seemingly offering no immediate advantage over a direct method, the deviant MPES estimates point to those parts of the evidence body which remain

TABLE 3

Posterior medians with 95% *credibility intervals of epidemiological parameters and total* ($N\rho\pi$) *and undiagnosed* ($N\rho\pi(1-\delta) = N\rho\pi - \mu$) *infections from the MPES model of HIV prevalence in Amsterdam*

Group	Subgroup	$\hat{\rho}$ (%)	$\hat{\pi}$ (%)	$\hat{\delta}$ (%)	$\widehat{N\rho\pi}$	$\widehat{N\rho\pi} - \hat{\mu}$
	STI	0.879 (0.844, 0.913)	29.100 (24.910, 33.800)	93.510 (87.810, 97.210)	726 (617, 846)	46 (19, 91)
MSM	Non-STI	9.871 (8.298, 11.680)	9.641 (8.074, 11.590)	88.730 (74.900, 95.370)	2,682 (2,380, 3,208)	301 (113, 800)
	All	10.750 (9.173, 12.560)	11.250 (9.645, 13.180)	89.640 (78.130, 95.570)	3,404 (3,138, 3,931)	351 (140, 854)
	M	0.332 (0.256, 0.399)	21.365 (15.690, 28.100)	90.700 (69.660, 99.630)	198 (132, 286)	18 (0, 76)
IDU	F	0.093 (0.074, 0.108)	27.260 (21.180, 35.520)	93.280 (73.751, 99.700)	71 (54, 95)	4 (0, 22)
	MF	0.212 (0.173, 0.248)	22.650 (17.840, 28.590)	90.790 (72.441, 99.610)	270 (198, 362)	24 (0, 90)
FSW	F	2.620 (2.561, 2.679)	3.133 (1.192, 6.252)	33.975 (4.950, 69.159)	233 (88, 467)	148 (43, 367)
	M	2.008 (1.958, 2.061)	0.297 (0.184, 0.577)	57.630 (26.940, 91.490)	16 (10, 32)	7 (0, 23)
WST_{STI}	F	2.318 (2.265, 2.376)	0.168 (0.112, 0.317)	64.910 (32.590, 93.410)	11 (7, 20)	3 (0, 12)
	MF	2.164 (2.126, 2.202)	0.234 (0.153, 0.397)	59.070 (32.140, 90.579)	28 (18, 48)	11 (1, 31)
	M	0.091 (0.080, 0.102)	3.732 (2.686, 6.541)	79.315 (47.120, 97.290)	9 (7, 16)	1 (0, 8)
SSA_{STI}	F	0.056 (0.048, 0.065)	7.348 (5.947, 9.797)	83.700 (68.730, 97.040)	11 (10, 15)	1 (0, 4)
	MF	0.073 (0.067, 0.081)	5.179 (4.160, 7.164)	80.580 (59.801, 96.700)	21 (17, 29)	4 (0, 11)
	M	0.315 (0.295, 0.337)	0.600 (0.452, 1.119)	84.500 (54.521, 97.860)	5 (4, 10)	0 (0, 4)
CRB_{STI}	F	0.271 (0.252, 0.290)	0.917 (0.767, 1.340)	87.630 (73.252, 97.890)	7 (6, 10)	0 (0, 2)
	MF	0.293 (0.279, 0.307)	0.763 (0.624, 1.108)	85.360 (66.271, 97.380)	12 (10, 18)	1 (0, 5)
	M	3.825 (3.590, 4.071)	1.899 (1.186, 3.278)	64.050 (35.030, 88.019)	206 (129, 356)	73 (17, 222)
$SSA_{\overline{STI}}$	F	3.394 (3.161, 3.587)	3.503 (2.877, 4.371)	72.300 (59.431, 83.320)	336 (279, 415)	93 (48, 164)
	MF	3.607 (3.430, 3.781)	2.671 (2.142, 3.531)	68.670 (51.800, 81.030)	546 (441, 719)	170 (88, 342)
	M	10.970 (10.650, 11.280)	0.442 (0.292, 0.693)	75.720 (46.760, 91.440)	137 (90, 214)	32 (9, 106)
$CRB_{\overline{STI}}$	F	12.910 (12.550, 13.270)	0.375 (0.294, 0.479)	80.270 (66.860, 92.130)	137 (107, 175)	26 (9, 55)
	MF	11.940 (11.680, 12.200)	0.407 (0.319, 0.539)	77.410 (59.160, 89.000)	276 (216, 365)	61 (26, 144)
	M	71.710 (69.840, 73.360)	0.065 (0.040, 0.129)	80.750 (39.700, 98.710)	132 (81, 264)	24 (1, 152)
$WST_{\overline{STI}}$	F	78.350 (77.890, 78.800)	0.066 (0.033, 0.109)	85.020 (54.701, 98.800)	147 (72, 242)	21 (1, 81)
	MF	75.030 (74.070, 75.880)	0.067 (0.042, 0.107)	82.180 (49.150, 98.530)	284 (180, 455)	48 (3, 216)
	M	100 (100, 100)	1.459 (1.335, 1.679)	86.655 (75.320, 93.860)	4,143 (3,791, 4,768)	553 (234, 1,170)
Total	F	100 (100, 100)	0.340 (0.286, 0.422)	67.075 (53.860, 78.700)	965 (812, 1,200)	316 (174, 550)
	MF	100 (100, 100)	0.901 (0.831, 1.017)	82.690 (73.560, 89.200)	5,120 (4,720, 5,777)	885 (512, 1,521)

inconsistent, hence suggesting what type of supplementary information would be most useful to resolve the discrepancy.

Furthermore, it can be seen that the leverage each evidence source applies to the final MPES estimate is determined by its corresponding sample size, not by it being graded as "direct" or "indirect": this is shown in Table 2 with $\pi^{\mathrm{f}}_{\mathrm{O},\mathrm{CRB}_{\overline{\mathrm{STI}}}}$, whose estimate is more driven by the stronger indirect evidence. Finally, it is worth noting that the availability, in a MPES setup, of different sets of inferences, each resulting from a different level of evidence synthesis, further stresses the robustness and efficiency of a full MPES approach over any arbitrary selection of items from the complete evidence base. Eventually the MPES approach contributed to a better understanding of the nature of those evidence conflicts which, pending the availability of additional data (the provision of which will be discussed for future updates of national estimates), remain unresolved.

6.2. *Prior information in MPES.* As detailed in Section 4.2, the presented MPES model relies on a number of prior assumptions. This is in line with the MPES spirit of informing the analysis with *all* available evidence, not just "hard" data. Often reliance on expert opinion is regarded as inappropriate, in that if misused it could steer the analysis toward partly subjective outcomes. On the other hand, as notably pointed out in Robert [(2007), Chapter 1], knowledge does not exclusively derive from field data, but actually builds on it. Substantive prior information informing the illustrated case study was typically introduced pragmatically: earlier versions of the MPES model featuring fewer/milder prior assumptions than those listed in Section 4.2 produced overly inaccurate (i.e., with unduly wide credibility bounds) estimates for some poorly informed subgroup-region combinations.[5] An MPES model can help in identifying those parameters whose estimation would benefit the most from the collection of larger/additional samples. To this end, while MPES modeling falls short of indicating which design strategy would yield largest efficiency gains, insights in this respect are naturally offered by more formal decision-theoretic tools, such as those based on the concept of value of information [Parmigiani and Inoue (2009), Chapter 13]. While these are receiving increasing attention by the environmental and health sciences community, they remain the subject of ongoing investigation and fall outside the remit of this paper.

6.3. *Current HIV prevalence estimation platforms.* The MPES approach lends itself as a valuable framework for national HIV prevalence estimation. Alternative options have been freely made available in recent years by the UNAIDS Reference Group on Estimates, Modeling and Projections: that is, the Estimation and

[5]Results not shown.

Projection Package [EPP, Ghys et al. (2004); Brown et al. (2006)] and the Workbook Method [Walker et al. (2004); Lyerla et al. (2006)], each implemented in a bespoken software package (unlike MPES).

EPP assumes the national population is subdivided into nonoverlapping risk subgroups, for which historical records of size and HIV prevalence are available. EPP then fits a simple transmission model to the prevalence data via Sampling Importance Resampling from a Bayesian Melding perspective [Poole and Raftery (2000)], generating a cluster of epidemic curves for each urban/rural and subgroup-specific sub-epidemic [Alkema, Raftery and Clark (2007); Raftery and Bao (2010)]. Resulting national HIV prevalence and incidence projections can then be fed into the stand-alone Spectrum module [Stover (2004)] to predict over time the number of individuals living with HIV or AIDS, new HIV infections, etc.

The Workbook Method estimates and projects HIV prevalence in countries lacking an HIV surveillance network consistently monitoring local prevalence patterns over time. Similarly to MPES and EPP, albeit to a coarser degree, Workbook estimates rely on a classification of the target population by risk profiles for which values of the maximum and minimum size and of HIV prevalence are available. The various combinations of lower-upper bounds are then cross-multiplied and averaged to obtain informal "plausibility" ranges for national HIV prevalence. This in turn can be imported into EPP/Spectrum to obtain a wider array of ancillary HIV epidemic descriptors.

A comparative discussion of the advantages and shortfalls of the three approaches (MPES, EPP/Spectrum, Workbook) is presented elsewhere [van Veen et al. (2011)]. Extensive criticism of Workbook estimates has led to a marked shift toward utilization and development of EPP among epidemiologists and practitioners in the field. While MPES has been only recently extended to HIV prevalence estimation, its flexibility shows promise for application to increasingly varied and complex data structures. Successful implementations have been carried out to incorporate time-series data for the estimation of HIV prevalence and incidence trends and patterns among MSM in England and Wales [Presanis et al. (2011)]. Additional case studies to be conducted via MPES modeling are currently being sought among eastern European countries, since this should facilitate the continuing development necessary for the methodology to reach higher levels of dissemination and maturity.

Acknowledgments. The authors gratefully acknowledge all institutions, data managers and researchers who supplied data as well as valuable discussion and follow-up. The authors are also grateful to three anonymous referees for their helpful advice and comments.

SUPPLEMENTARY MATERIAL

Supplement A: HIV prevalence data in Rotterdam and the rest of the Netherlands (DOI: 10.1214/11-AOAS488SUPPA; .ps). Surveillance- and survey-

type data supporting HIV prevalence estimation Rotterdam and the Rest of the Netherlands.

Supplement B: MPES model and data files (DOI: 10.1214/11-AOAS488SUPPB; .zip). `WinBUGS` code of the MPES model and data inputs enabling HIV prevalence estimation in the Netherlands.

Supplement C: HIV prevalence estimates in Rotterdam and the Netherlands (including and excluding Amsterdam) (DOI: 10.1214/11-AOAS488SUPPC; .ps). Posterior inferences on HIV prevalence descriptors by risk subgroup in Rotterdam and the Netherlands (separately including and excluding Amsterdam and Rotterdam).

REFERENCES

ADES, A. E. (2003). A chain of evidence with mixed comparisons: Models for multi-parameter synthesis and consistency of evidence. *Stat. Med.* **22** 2995–3016.

ADES, A. E. and CLIFFE, S. (2002). Markov Chain Monte Carlo estimation of a multiparameter decision model: Consistency of evidence and the accurate assessment of uncertainty. *Medical Decision Making* **22** 359–371.

ADES, A. E. and SUTTON, A. J. (2006). Multiparameter evidence synthesis in epidemiology and medical decision-making: Current approaches. *J. R. Stat. Soc. Ser. A Stat. Soc.* **169** 5–35. MR2222010

ADES, A. E., SCULPHER, M., SUTTON, A., ABRAMS, K., COOPER, N., WELTON, N. and LU, G. (2006). Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics* **24** 1–19.

ADES, A. E., WELTON, N. J., CALDWELL, D., PRICE, M., GOUBAR, A. and LU, G. (2008). Multiparameter evidence synthesis in epidemiology and medical decision-making. *Journal of Health Services Research & Policy* **13** 12–22.

ALKEMA, L., RAFTERY, A. E. and CLARK, S. J. (2007). Probabilistic projections of HIV prevalence using Bayesian melding. *Ann. Appl. Stat.* **1** 229–248. MR2393849

BENSON, K. and HARTZ, A. (2000). A comparison of observational studies and randomized controlled trials. *New England Journal of Medicine* **342** 1878–1886.

BERKEY, C. S., HOAGLIN, D. C., ANTCZAK-BOUCKOMS, A., MOSTELLER, F. and COLDITZ, G. A. (1998). Meta-analysis of multiple outcomes by regression with random effects. *Stat. Med.* **17** 2537–2550.

BROOKMEYER, R. and GAIL, M. H. (1988). A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *J. Amer. Statist. Assoc.* **83** 301–308.

BROWN, T., GRASSLY, N. C., GARNETT, G. and STANECKI, K. (2006). Improving projections at the country level: The UNAIDS estimation and projection package 2005. *Sexually Transmitted Infections* **82** iii34–iii40.

BURZYKOWSKI, T., MOLENBERGHS, G. and BUYSE, M. (2004). The validation of surrogate end points by using data from randomized clinical trials: A case-study in advanced colorectal cancer. *J. Roy. Statist. Soc. Ser. A* **167** 103–124. MR2043435

CALDWELL, D. M., ADES, A. E. and HIGGINS, J. P. T. (2005). Simultaneous comparison of multiple treatments: Combining direct and indirect evidence. *British Medical Journal* **331** 897–900.

CLAXTON, K. P., SCULPHER, M. J. and DRUMMOND, M. (2002). A rational framework for decision making by the national institute for clinical excellence (NICE). *The Lancet* **360** 711–715.

DE ANGELIS, D., SWEETING, M., ADES, A. E., HICKMAN, M., HOPE, V. and RAMSAY, M. (2009). An evidence synthesis approach to estimating hepatitis C prevalence in England and Wales. *Stat. Methods Med. Res.* **18** 361–379. MR2750101

DOMINICI, F., PARMIGIANI, G., WOLPERT, R. L. and HASSELBLAD, V. (1999). Meta-analysis of migraine headache treatments: Combining information from heterogeneous designs. *J. Amer. Statist. Assoc.* **94** 16–28.

DRIOYCOUR, F., SILBERMAN, G. and CHELIMSKY, E. (1993). Cross-design synthesis: A new form of meta-analysis for combining the results from randomised clinical trials and medical-practice databases. *International Journal of Technology Assessment in Health Care* **9** 440–449.

EDDY, D. M. and HASSELBLAD, V. (1992). *Meta-analysis by the Confidence Profile Method*. Academic Press, Boston, MA.

FUENTES, M. and RAFTERY, A. E. (2005). Model evaluation and spatial interpolation by Bayesian combination of observations with outputs from numerical models. *Biometrics* **61** 36–45. MR2129199

GELMAN, A. and HILL, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. *Analytical Methods for Social Research*. Cambridge Univ. Press, Cambridge.

GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.

GHYS, P. D., BROWN, T., GRASSLY, N. C., GARNETT, G., STANECKI, K. A., STOVER, J. and WALKER, N. (2004). The UNAIDS Estimation and Projection Package: A Software Package to Estimate and Project National HIV Epidemics. *British Medical Journal* **80** i5–i9.

GIESECKE, J., JOHNSON, A., HAWKINS, A., NOONE, A. and NICOLL, A. (1994). An estimate of the prevalence of human immunodeficiency virus infection in England and Wales by using a direct method. *J. Roy. Statist. Soc. Ser. A* **157** 89–103.

GOUBAR, A., ADES, A. E., DE ANGELIS, D., MCGARRIGLE, C. A., MERCER, C. H., TOOKEY, P. A., FENTON, K. and GILL, O. N. (2008). Estimates of human immunodeficiency virus prevalence and proportion diagnosed based on Bayesian multi-parameter synthesis of surveillance data. *J. Roy. Statist. Soc. Ser. A* **171** 541–580. MR2432503

HIV & STI DEPARTMENT (2005). Mapping the issues. HIV and other sexually transmitted infections in the United Kingdom 2005. Technical report, Health Protection Agency.

HIV & STI DEPARTMENT (2006). A complex picture. HIV and other sexually transmitted infections in the United Kingdom 2006. Technical report, Health Protection Agency.

HIV & STI DEPARTMENT (2007). Testing times. HIV and other sexually transmitted infections in the United Kingdom 2007. Technical report, Health Protection Agency.

HIV & STI DEPARTMENT (2008). HIV in the United Kingdom: 2008 report. Technical report, Health Protection Agency.

HIV & STI DEPARTMENT (2009). HIV in the United Kingdom: 2009 report. Technical report, Health Protection Agency.

HOUWELING, H., HEISTERKAMP, S. H., WIESSING, L. G., COUTINHO, R. A., VAN WIJNGAARDEN, J. K. and JAGER, H. J. C. (1998). Methods for estimating HIV prevalence: A comparison of extrapolation from surveys on infection rate and risk behaviour with back-calculation for the Netherlands. *European Journal of Epidemiology* **14** 645–652.

KARON, J. M., KHARE, M. and ROSENBERG, P. S. (1998). The current status of methods for estimating the prevalence of human immunodeficiency virus in the United States of America. *Stat. Med.* **17** 127–142.

LU, G. and ADES, A. E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Stat. Med.* **23** 3105–3124.

LUNN, D. J., THOMAS, A., BEST, N. and SPIEGELHALTER, D. (2000). WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat. Comput.* **10** 325–337.

LYERLA, R., GOUWS, E., GARCIA-CALLEJA, J. M. and ZANIEWSKI, E. (2006). The 2005 workbook: An improved tool for estimating HIV prevalence in countries with low level and concentrated epidemics. *Sexually Transmitted Infections* **82** iii41–iii44.

MCGARRIGLE, C. A., CLIFFE, S., COPAS, A. J., MERCER, C. H., DE ANGELIS, D., FENTON, K. A., EVANS, B. G., JOHNSON, A. M. and GILL, O. N. (2006). Estimating adult HIV prevalence in the UK in 2003: The direct method of estimation. *British Medical Journal* **82** 78–86.

NAM, I. S., MENGERSEN, K. and GARTHWAITE, P. (2003). Multivariate meta-analysis. *Stat. Med.* **22** 2309–2333.

PARMIGIANI, G. and INOUE, L. (2009). *Decision Theory*: *Principles and Approaches*. Wiley, Chichester. MR2604978

PETRUCKEVITCH, A., NICOLL, A., JOHNSON, A. and BENNETT, D. (1997). Direct estimates of prevalent HIV infection in adults in England and Wales for 1991 and 1993: An improved method. *Genitourinary Medicine* **73** 348–354.

POOLE, D. and RAFTERY, A. E. (2000). Inference for deterministic simulation models: The Bayesian melding approach. *J. Amer. Statist. Assoc.* **95** 1244–1255. MR1804247

PRESANIS, A. M., DE ANGELIS, D., SPIEGELHALTER, D. J., SEAMAN, S., GOUBAR, A. and ADES, A. E. (2008). Conflicting evidence in a Bayesian synthesis of surveillance data to estimate HIV prevalence. *J. Roy. Statist. Soc. Ser. A* **171** 915–937. MR2530293

PRESANIS, A. M., DE ANGELIS, D., GOUBAR, A., GILL, O. N. and ADES, A. E. (2011). Bayesian evidence synthesis for a transmission dynamic model for HIV among men who have sex with men. *Biostatistics* **0** 1–16.

RAFTERY, A. E. and BAO, L. (2010). Estimating and projecting trends in HIV/AIDS generalized epidemics using incremental mixture importance sampling. *Biometrics*.

RAMÓN, J. S., ALVARENGA, M., WALKER, N., GARCIA-CALLEJA, J. M. and ZACARIAS, F. (2002). Estimating HIV/AIDS prevalence in countries with low-level and concentrated epidemics: The example of Honduras. *AIDS* **16** S18–S22.

ROBERT, C. P. (2007). *The Bayesian Choice*: *From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. Springer, New York. MR2723361

SEAMAN, S., DE ANGELIS, D. and PRESANIS, A. M. (2011). Using posterior mean deviance to assess Bayesian model fit. Personal communication.

SONG, F., ALTMAN, D. G., GLENNY, A. M. and DEEKS, J. J. (2003). Validity of indirect comparison for estimating efficacy of competing interventions: Empirical evidence from published meta-analyses. *British Medical Journal* **326** 472–476.

SPIEGELHALTER, D. J. and BEST, N. G. (2003). Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Stat. Med.* **22** 3687–3709.

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 583–639. MR1979380

STOVER, J. (2004). Projecting the demographic consequences of adult HIV prevalence trends: The spectrum projection package. *Sexually Transmitted Infections* **80** i14–i18.

SUTTON, A. J., ABRAMS, K. R., JONES, D. R., SHELDON, T. A. and SONG, F. (2000). *Methods for Meta-Analysis in Medical Research*. Wiley, Chichester.

SWEETING, M. J., DE ANGELIS, D., HICKMAN, M. and ADES, A. E. (2008). Estimating hepatitis C prevalence in england and Wales by synthesizing evidence from multiple data sources. Assessing data conflict and model fit. *Biostatistics* **9** 715–734.

TURNER, R. M., SPIEGELHALTER, D. J., SMITH, G. C. S. and THOMPSON, S. G. (2009). Bias modelling in evidence synthesis. *J. Roy. Statist. Soc. Ser. A* **172** 21–47. MR2655603

VAN DER BIJ, A. K., DUKERS, N. H. T. M., COUTINHO, R. A. and FENNEMA, H. S. A. (2008). Low HIV-testing rates and awareness of HIV infection among high-risk heterosexual STI clinic attendees in The Netherlands. *The European Journal of Public Health* **18** 376–379.

VAN VEEN, M. G. (2009). Interim meeting. Personal communication.

VAN VEEN, M. G., WAGEMANS, M. A. J., OP DE COUL, E. L. M., FENNEMA, J. S. A., VAN DER HELM, T. C. M., WALTER, J., PRINS, M. and VAN DER LAAR, M. J. W. (2005). HIV-surveys among high risk populations in Amsterdam 2003–2004. Technical report, National Institute for Public Health and the Environment, Bilthoven, Netherlands. (In Dutch.)

VAN VEEN, M. G., PRESANIS, A. M., CONTI, S., XIRIDOU, M., RINDER STENGAARD, A., DONOGHOE, M. C., VAN SIGHEM, A., VAN DER SANDE, M. and DE ANGELIS, D. (2011). National estimate of HIV prevalence in The Netherlands: Comparison and applicability of different estimation tools. *AIDS* **25** 229–237.

WALKER, N., STOVER, J., STANECKI, K., ZANIEWSKI, A. E., GRASSLY, N. C., GARCIA-CALLEJA, J. M. and GHYS, P. D. (2004). The workbook approach to making estimates and projecting future scenarios of HIV/AIDS in countries with low level and concentrated epidemics. *Sexually Transmitted Infections* **80** i10–i13.

WELTON, N. J. and ADES, A. E. (2005). A model of toxoplasmosis incidence in the UK: Evidence synthesis and consistency of evidence. *J. Roy. Statist. Soc. Ser. C* **54** 385–404. MR2135881

WHITEHEAD, A. (2002). *Meta-Analysis of Controlled Clinical Trials*. Wiley, Chichester.

WOLPERT, R. L. and MENGERSEN, K. L. (2004). Adjusted likelihoods for synthesizing empirical evidence from studies that differ in quality and design: Effects of environmental tobacco smoke. *Statist. Sci.* **19** 450–471. MR2185626

S. CONTI
STATISTICS UNIT
CENTRE FOR INFECTIONS
HEALTH PROTECTION AGENCY
61 COLINDALE AVENUE
LONDON NW9 5EQ
UNITED KINGDOM
E-MAIL: stefano.conti@hpa.org.uk

A. M. PRESANIS
MRC BIOSTATISTICS UNIT
INSTITUTE OF PUBLIC HEALTH
UNIVERSITY FORVIE SITE
ROBINSON WAY
CAMBRIDGE CB2 0SR
UNITED KINGDOM

M. G. VAN VEEN
M. XIRIDOU
NATIONAL INSTITUTE FOR PUBLIC HEALTH
AND THE ENVIRONMENT
PO BOX 1
3720 BA BILTHOVEN
THE NETHERLANDS

M. C. DONOGHOE
A. RINDER STENGAARD
WORLD HEALTH ORGANIZATION
REGIONAL OFFICE FOR EUROPE
SCHERFIGSVEJ 8
DK-2100 COPENHAGEN Ø
DENMARK

D. DE ANGELIS
STATISTICS UNIT
CENTRE FOR INFECTIONS
HEALTH PROTECTION AGENCY
61 COLINDALE AVENUE
LONDON NW9 5EQ
UNITED KINGDOM
AND
MRC BIOSTATISTICS UNIT
INSTITUTE OF PUBLIC HEALTH
UNIVERSITY FORVIE SITE
ROBINSON WAY
CAMBRIDGE CB2 0SR
UNITED KINGDOM