Estimating hepatitis C prevalence in England and Wales by synthesizing evidence from multiple data sources. Assessing data conflict and model fit

M. J. SWEETING*
MRC Biostatistics Unit, Institute of Public Health, Robinson Way,
Cambridge CB2 0SR, UK
michael.sweeting@mrc-bsu.cam.ac.uk

D. DE ANGELIS
MRC Biostatistics Unit, Institute of Public Health, Robinson Way,
Cambridge CB2 0SR, UK and Health Protection Agency Centre for Infections, Statistics Unit,
61 Colindale Avenue, London NW9 5EQ, UK

M. HICKMAN
Department of Social Medicine, University of Bristol, Canynge Hall,
Whiteladies Road, Bristol BS8 2SP, UK

A. E. ADES
MRC Health Services Research Collaboration, University of Bristol,
Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK

SUMMARY
Multiparameter evidence synthesis is becoming widely used as a way of combining evidence from multiple and often disparate sources of information concerning a number of parameters. Synthesizing data in one encompassing model allows propagation of evidence and learning. We demonstrate the use of such an approach in estimating the number of people infected with the hepatitis C virus (HCV) in England and Wales. Data are obtained from seroprevalence studies conducted in different subpopulations. Each sub-population is modeled as a composition of 3 main HCV risk groups (current injecting drug users (IDUs), ex-IDUs, and non-IDUs). Further, data obtained on the prevalence (size) of each risk group provide an estimate of the prevalence of HCV in the whole population. We simultaneously estimate all model parameters through the use of Bayesian Markov chain Monte Carlo techniques. The main emphasis of this paper is the assessment of evidence consistency and investigation of the main drivers for model inferences. We consider a cross-validation technique to reveal data conflict and leverage when each data source is in turn removed from the model.

*To whom correspondence should be addressed.

© The Author 2008. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org.
1. INTRODUCTION

The hepatitis C virus (HCV) is transmitted through blood and can lead to chronic liver disease and death. In England and Wales, the total number of people infected with HCV is unknown and currently object of debate. The difficulty with estimating the magnitude of the infected population is that no nationwide surveillance has been carried out, and populations at high risk of infection, such as injecting drug users (IDUs), are not easy to study (Health Protection Agency and others, 2006).

The prevalence of HCV antibodies (anti-HCV) has been estimated in specific groups of the population such as IDUs attending treatment clinics, genitourinary medicine (GUM) clinic attenders, women attending antenatal clinics, and individuals donating blood. These group-specific estimates cannot be used to estimate population prevalence unless some additional information is known on the composition of each group in terms of HCV risk groups and the prevalence of the risk groups in the population. Traditionally, the way of proceeding in this situation has been either to assume that these studies provide unbiased estimates of anti-HCV prevalence in the population or to make informal judgments on the representativeness of each study in terms of risk for HCV and naively introduce ad hoc adjustments to the prevalence to reflect such judgment (Department of Health, 2002). The effect of these adjustments could be assessed via sensitivity analysis, with no possibility of a formal incorporation of uncertainty in the final estimates of the number of infected individuals.

In this paper, we propose a Bayesian multiparameter evidence synthesis approach to the estimation problem, where information from each available study is linked to 3 risk groups for HCV infection: current IDUs, ex-IDUs, and non-IDUs. We introduce a number of key parameters relating to either the prevalence of HCV in each risk group or the prevalence of the risk groups in the population. All the available data provide information on a complex function of these parameters by defining each study in terms of a mixture of the risk groups. We incorporate additional, perhaps uncertain, data on the composition of these mixtures and consider the modeling of potential biases. The approach allows an evidence-based estimation of anti-HCV prevalence in each of the 3 risk groups. The resulting prevalence estimates, combined with estimates of the size of each risk group in the population, will produce an estimate of the overall prevalence of anti-HCV in the country, with the uncertainty in each study propagated through to these final estimates.

Multiparameter evidence synthesis has been used in the medical literature to estimate HIV prevalence (Goubar and others, 2006) and toxoplasmosis incidence (Welton and Ades, 2005) and for decision modeling of prenatal testing for HIV (Ades and Cliffe, 2002). The approach is part of the confidence profile method (Eddy and others, 1990), which also encompasses conventional fixed- and random-effects meta-analysis and mixed treatment comparisons. A thorough review of the area has been undertaken by Ades and Sutton (2006).

One major issue that arises when synthesizing multiple data sources is the consistency of the evidence. Goodness-of-fit assessment has generally been conducted through the use of posterior mean deviances and posterior predictive p-values (Goubar and others, 2006; Welton and Ades, 2005; Ades and Cliffe, 2002). In addition, a cross-validation approach can be implemented, which is useful in identifying conflicts between specific data sources. In this paper we reestimate the model leaving out each data source in turn and study effects on the posterior mean deviance contributions. The cross-validation also allows assessment of data sources that are influential to the synthesis. Without such an approach it is difficult in a complex modeling process to satisfactorily explain the main “drivers” of the conclusions. We compare the relative fit of a number of models using the deviance information criterion (DIC) (Spiegelhalter and others, 2002) and absolute fit using posterior predictive p-values.

Section 2 presents the underlying epidemiological model and the quantities that we wish to estimate. Section 3 outlines the available data and the limitations of each data source. Section 4 presents in
more detail the model that has been developed and the functional relationships between key quantities of interest, while Section 5 describes how we undertake the evidence synthesis and assess Bayesian model fit. Section 6 presents the results of our investigations, concentrating in particular on model assessment.

2. THE EPIDEMIOLOGICAL MODEL

The primary route of transmission for HCV infection is injecting drug use (Ramsay and others, 1998); thus, the population of England and Wales at time $T$ is divided into 3 mutually exclusive risk groups, denoted by $g$, with $g = \text{CUR}$ for current IDUs, $g = \text{EX}$ for ex-IDUs, and $g = \text{NON}$ for non-IDUs. Here, we define by “current” IDU an individual who has injected within the last year and by “ex-IDU” someone who last injected more than a year ago. The non-IDU risk group will include individuals exposed to infection through other means such as blood transfusions before screening for anti-HCV was introduced in 1991.

Let $\rho_{\text{CUR}}(r, s, a)$, $\rho_{\text{EX}}(r, s, a)$, and $\rho_{\text{NON}}(r, s, a)$ represent the prevalence (i.e. the proportion) of the current IDU, ex-IDU, and non-IDU risk groups in the population at time $T$ for region $r$, gender $s$, and age group $a$. Note that stratification by region, gender, and age group will be used to explain heterogeneity in the data. We consider 3 regions: London ($r = 0$), North West England ($r = 1$), and rest of England and Wales ($r = 2$); 3 age groups: $[15, 29]$ ($a = 0$), $[30, 44]$ ($a = 1$), and $[45, 59]$ ($a = 2$) and take $s = 0$ and $s = 1$ to indicate the female and male gender, respectively. By definition

$$\rho_{\text{NON}}(r, s, a) = 1 - \rho_{\text{CUR}}(r, s, a) - \rho_{\text{EX}}(r, s, a)$$

and let

$$\rho_{\text{IDU}}(r, s, a) = \rho_{\text{CUR}}(r, s, a) + \rho_{\text{EX}}(r, s, a)$$

be the prevalence of ever injecting drugs in the population given $(r, s, a)$. Let $\pi_{\text{CUR}}(r, s, a)$, $\pi_{\text{EX}}(r, s, a)$, and $\pi_{\text{NON}}(r, s, a)$ denote the prevalence of anti-HCV in the current IDU, ex-IDU, and non-IDU risk groups, respectively. Then, the linear combination

$$\pi(r, s, a) = \rho_{\text{CUR}}(r, s, a)\pi_{\text{CUR}}(r, s, a) + \rho_{\text{EX}}(r, s, a)\pi_{\text{EX}}(r, s, a) + \rho_{\text{NON}}(r, s, a)\pi_{\text{NON}}(r, s, a)$$

(2.1)

defines the anti-HCV prevalence in region $r$, gender $s$, and age group $a$. The weighted average

$$\pi = \frac{\sum_{r}\sum_{s}\sum_{a} N(r, s, a)\pi(r, s, a)}{\sum_{r}\sum_{s}\sum_{a} N(r, s, a)}$$

(2.2)

is the overall anti-HCV prevalence, where $N(r, s, a)$ represents the population in region $r$, sex $s$, and age group $a$. Similarly, the marginal prevalence of intravenous drug use can be calculated for risk group $g$, as can the marginal risk group–specific HCV prevalence as follows:

$$\rho_{g} = \frac{\sum_{r}\sum_{s}\sum_{a} N(r, s, a)\rho_{g}(r, s, a)}{\sum_{r}\sum_{s}\sum_{a} N(r, s, a)}$$

$$\pi_{g} = \frac{\sum_{r}\sum_{s}\sum_{a} N(r, s, a)\rho_{g}(r, s, a)\pi_{g}(r, s, a)}{\sum_{r}\sum_{s}\sum_{a} N(r, s, a)\rho_{g}(r, s, a)}$$

The aim is to estimate $\pi_{g}(r, s, a)$ and $\rho_{g}(r, s, a)$ simultaneously on the basis of evidence from many data sources. Some of this evidence will be in terms of data directly informing $\pi_{g}(r, s, a)$ and/or $\rho_{g}(r, s, a)$, while other evidence will be indirect, that is, constituted by data expressed as complex functions of the quantities of interest. In general, there is much more information on current injecting drug users rather than the ex, for whom the evidence is typically sparse.
3. DATA SOURCES

Each data source provides information either directly or indirectly on the prevalence of IDU or the prevalence of anti-HCV. The data are not all specific to 1 year. However, the majority are conducted in the period 2001–2003, and hence we shall assume  \( T = 2003 \) for all data sources.

3.1 Data informing the prevalence of injecting drug use


*Household surveys.* Four national household surveys contain questions on the practice of injecting drug use and/or the use of heroin. The British Crime Survey (BCS) (Home Office Research Development and Statistics Directorate and BMRB Social Research, 2006) and the Offending Crime and Justice Survey (CJS) (Home Office Research Development and Statistics Directorate and BMRB Social Research, 2005) are 2 Home Office household surveys which report current use of heroin in the past year and ever use of heroin. Data are available for the year 2003. The Survey of Psychiatric Morbidity (Singleton and others, 2001), carried out by the Office of National Statistics between March and September 2000 among adults aged 16–74 years, also contains questions on heroin use in the past year and ever use of heroin. Finally, the National Survey of Sexual Attitudes and Lifestyles (NATSAL) (National Centre for Social Research, 2005) conducted in 2000 in people aged 16–44 years contains questions on non-prescribed injecting drug use in the past year and ever use of non-prescribed injecting drugs.

Multiple response biases in household surveys are likely to cause severe underreporting in the number of both current and ex-IDUs/heroin users (Harrison and Hughes, 1997), and hence IDU prevalence is likely to be underestimated.

3.2 Data informing the prevalence of hepatitis C

*IDUs attending specialist treatment agencies.* The Unlinked Anonymous (UA) Prevalence Monitoring Programme (Unlinked Anonymous Surveys Steering Group, 2002) provides data on prevalence of anti-HCV in current dependent IDUs as well as on ex-IDUs still in contact with treatment services for the years 1998–2003. Estimates are available by region, gender, and age group and for other important predictors of prevalence such as the duration of injecting and the year of starting injecting. In this data source, saliva samples are used to test for HCV antibodies. The test sensitivity of saliva samples for HCV antibodies is approximately 92% (Judd and others, 2003) and must be accounted for in estimates of HCV prevalence (see Section 5.3).

*GUM clinic attendees.* As part of the UA programme, sera have been collected from patients attending GUM clinics (Unlinked Anonymous Surveys Steering Group, 2002). Patients are asked questions about the use of injecting drugs in their lifetime. Archived sera from GUM clinics in 1995 have been tested for anti-HCV (Balogun and others, 2003), in patients who reported never injecting drugs, and more recent samples in 2001 are available in patients who report ever injecting. The prevalence measured in ever injectors can be considered a mixture of anti-HCV prevalence in the current and ex-IDU groups.

*Studies in pregnant women.* Two major studies have been carried out to estimate the prevalence of HCV in pregnant women in 1997/1998. The first tested UA archived specimens from women who attended 1 of 14 antenatal clinics in London and 11 clinics in the Northern and Yorkshire region (Balogun
and others, 2000). In total, 42,613 samples were tested for the presence of anti-HCV antibodies, and the data are stratified by the 2 regions and into the 15–29 and 30–44 age groups. The majority of the 14 London antenatal clinics were situated in Inner London boroughs. The second study tested 126,009 dried blood spot samples from the Neonatal Screening Laboratory serving 29 districts in Inner and Outer London and adjacent nonmetropolitan areas (Ades and others, 2000). Pregnant women are a mixture of current, ex- and non-IDUs, and hence the analysis must account for this.

**Blood donations.** Blood donations are routinely tested for the presence of anti–HCV antibodies (http://www.hpa.org.uk/infections/topics_az/BIBD/menu.htm). Since blood donors are likely to be a healthy subpopulation, we would expect the resulting anti–HCV prevalence estimates from this group to be lower than those in the non-IDU population.

**The tested population.** The Sentinel Surveillance Study of HCV Testing records the prevalence of anti-HCV in individuals being tested in 8 sentinel sites across England (Brant and others, 2007). Data are available by region, sex, and age group for all tested individuals, and main risk group is obtained from a questionnaire carried out in a subsample. The prevalence of HCV in this tested population is likely to be higher than that in the population at large, and any model must account for this.

### 4. Developing the Model

#### 4.1 Relating current and ex-injecting drug use

The available data generally inform about IDU and HCV prevalence of current users. The household surveys provide some information on prevalence of ex-IDUs, but this is likely to lead to biased estimates. To estimate the main quantities of interest described in Section 2, ex-IDU–related quantities must be expressed as functions of current IDU quantities, thus allowing propagation of information on the prevalence of current IDU and HCV in current IDUs through to the ex-user population. This can be achieved by modeling explicitly the process of transition between current and ex-IDU through the concepts of time of starting injecting and duration of injecting. For instance, the prevalence of ex-IDU, \( \rho_{\text{EX}}(r, s, a) \), can be written as

\[
\rho_{\text{EX}}(r, s, a) = \rho_{\text{IDU}}(r, s, a) \kappa_{\text{EX}}(r, s, a),
\]

where \( \kappa_{\text{EX}}(r, s, a) \) is the probability of being an ex-user at time \( T \) conditional on being an ever IDU and the covariates \( r, s, \) and \( a \). Under simplifying assumptions (see Section A of the supplementary material available at Biostatistics online, http://www.biostatistics.oxfordjournals.org), \( \kappa_{\text{EX}}(r, s, a) \) can be expressed in terms of the injecting duration (\( D \)), the time since starting (TSS), defined as the lag between the time of starting injecting and time \( T \), and the age at first use (AAFU) as follows:

\[
\kappa_{\text{EX}}(r, s, a) = \sum_{t=0}^{T} F_{D|r,s,a}(t) f_{\text{TSS}|r,s,a}(t) \]

\[
= \sum_{t=0}^{T} F_{D}(t) f_{\text{TSS}}(t)[F_{\text{AAFU}|r,s}(A_{a+1} - t) - F_{\text{AAFU}|r,s}(A_{a} - t)]
\]

\[
= \sum_{t=0}^{T} f_{\text{TSS}}(t)[F_{\text{AAFU}|r,s}(A_{a+1} - t) - F_{\text{AAFU}|r,s}(A_{a} - t)],
\]

where \( f_{D}, f_{\text{TSS}}, \) and \( f_{\text{AAFU}} \) and \( F_{D}, F_{\text{TSS}}, \) and \( F_{\text{AAFU}} \) denote the probability mass functions and the distribution functions of \( D \), TSS, and AAFU, respectively. Thus, \( \kappa_{\text{EX}}(r, s, a) \) is the sum over all possible TSSs of the probability of starting \( t \) periods ago, at an age consistent with being in the age group \( a \) at \( T \), and injecting for a duration less than \( t \). Conditional on current age group \( a \), the TSS probability function \( f_{\text{TSS}|r,s,a}(t) \) is written in terms of its marginal distribution \( f_{\text{TSS}}(t) \) and the AAFU distribution \( F_{\text{AAFU}} \).
Following the same logic, the prevalence of current IDU, \( \rho_{\text{CUR}} \), can be written in terms of the prevalence of ever IDU. As a result, \( \rho_{\text{CUR}} \) and \( \rho_{\text{EX}} \) can be expressed as

\[
\rho_{\text{CUR}}(r, s, a) = \rho_{\text{IDU}}(r, s, a)(1 - \kappa_{\text{EX}}(r, s, a))
= g_1(\rho_{\text{IDU}}(r, s, a), f_D, f_{\text{TSS}}, f_{\text{AAFU}}), 
\tag{4.2}
\]

\[
\rho_{\text{EX}}(r, s, a) = \rho_{\text{IDU}}(r, s, a)\kappa_{\text{EX}}(r, s, a)
= g_2(\rho_{\text{IDU}}(r, s, a), f_D, f_{\text{TSS}}, f_{\text{AAFU}}), 
\tag{4.3}
\]

where the functions \( g_1 \) and \( g_2 \) can be gleaned from (4.1) and are given in full in Section A of the supplementary material available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org). From (4.2) and (4.3), it is clear that it is sufficient to model only the prevalence of ever IDU, and with knowledge of \( f_D, f_{\text{TSS}}, \) and \( f_{\text{AAFU}} \), the prevalence of current and ex-IDU is derived. In essence, our synthesis will therefore use data on \( D, TSS, \) and \( AAFU \) in order to propagate information on \( \rho_{\text{CUR}}(r, s, a) \) through to \( \rho_{\text{IDU}}(r, s, a) \), which in turn defines \( \rho_{\text{EX}}(r, s, a) \).

The probability mass functions \( f_D, f_{\text{TSS}}, \) and \( f_{\text{AAFU}} \) can also be used, together with the prevalence of HCV in drug users conditional on a time since starting \( TSS = tss \) and injecting duration \( D = d \), to derive the prevalence of HCV in current and ex-IDUs. Indicating by \( \pi_{\text{IDU}}(r, s, a, d, tss) \) the prevalence of HCV in drug users conditional on \( TSS \) and injecting duration, it can be shown that

\[
\pi_{\text{CUR}}(r, s, a) = g_3(\pi_{\text{IDU}}(r, s, a, d, tss), f_D, f_{\text{TSS}}, f_{\text{AAFU}}),
\tag{4.4}
\]

\[
\pi_{\text{EX}}(r, s, a) = g_4(\pi_{\text{IDU}}(r, s, a, d, tss), f_D, f_{\text{TSS}}, f_{\text{AAFU}}),
\tag{4.5}
\]

where (4.4) and (4.5) are explicitly derived in Section B of the supplementary material available at *Biostatistics* online (http://www.biostatistics.oxfordjournals.org). The prevalence of HCV in ex-users can therefore be estimated by the propagation of information from \( \pi_{\text{IDU}}(r, s, a, d, tss), f_D, f_{\text{TSS}}, \) and \( f_{\text{AAFU}} \).

### 4.2 Regression equations

We need to estimate the quantities of interest, \( \rho_g(r, s, a) \) and \( \pi_g(r, s, a) \), for all combinations of \( r, s, \) and \( a \). The dimension of the estimation problem can be reduced by specifying regression equations for 3 of the basic quantities: \( \rho_{\text{IDU}}(r, s, a), \pi_{\text{IDU}}(r, s, a, d, tss), \) and \( \pi_{\text{NON}}(r, s, a) \). The quantities relating to the current and ex-IDU risk groups can then be derived using (4.2–4.5). A separate regression equation is needed for \( \pi_{\text{NON}}(r, s, a) \) since region, sex, and age effects and baseline HCV prevalence will be different in the non-IDU population compared with the IDU population. In general, the regression equations can be written as follows:

\[
\text{logit}(\rho_{\text{IDU}}(r, s, a)) = a_0 + Xa,
\]

\[
\text{logit}(\pi_{\text{IDU}}(r, s, a, d, tss)) = \delta_0 + X\delta_1 + Y\delta_2,
\]

\[
\text{logit}(\pi_{\text{NON}}(r, s, a)) = \gamma_0 + X\gamma,
\]

where \( a_0, \delta_0, \) and \( \gamma_0 \) are the intercept terms, \( X \) is a design vector for the covariates region, sex, and age, \( Y \) is a design vector for the covariates injecting duration and TSS, and \( a, \delta_1, \delta_2, \) and \( \gamma \) are vectors of log-odds ratios of the covariates relative to the baseline. These parameters along with those defining the probability mass functions \( f_D, f_{\text{TSS}}, \) and \( f_{\text{AAFU}} \) are defined as the “basic” parameters of the model and are collectively labeled by \( \theta \). All other quantities may be derived from these “basic” parameters. Figure 1 shows the associations between model parameters and their derivation from the “basic” parameters.
5. Evidence Synthesis

All the prevalence data used in the synthesis can be expressed in the form of a numerator $y_{i,j}$ and a denominator $n_{i,j}$ for data source $i$, data point $j = 1, \ldots, m_i$. Each data point represents a certain covariate combination defined by the vector $(r_{i,j}, s_{i,j}, a_{i,j})$ and also $(d_{i,j}, tss_{i,j})$ for the UA IDU data source (Section 3.2). Within each source, we assume that the data points are independent and $y_{i,j} \sim \text{Binomial}(n_{i,j}, p_{i,j}(\theta))$.

5.1 Relationship between prevalence data and parameters

Each data source provides either direct or indirect information on one or more of the quantities of interest. There are 3 possible ways in which data can contribute to the synthesis. First, the study may provide unbiased information on one of the quantities of interest. If data source $i$ provides an unbiased estimate of current IDU prevalence, then

$$p_{i,j}(\theta) = \rho_{\text{CUR}}(r_{i,j}, s_{i,j}, a_{i,j})$$

for all $j = 1, \ldots, m_i$.

Second, the data source may estimate a given quantity with bias. To make use of such data sources, we shall assume that any bias acts only on the absolute level of the quantity of interest and does not affect estimation of odds ratios between covariate levels. This is achieved by defining a common bias parameter, $b_i$, for data source $i$, which has an additive effect on the logit scale. For example, if data source $i$ is biased in the estimation of current IDU prevalence, then

$$\logit(p_{i,j}(\theta)) = \logit(\rho_{\text{CUR}}(r_{i,j}, s_{i,j}, a_{i,j})) + b_i,$$
and hence it can be shown that

$$\text{OR}(p_{i,j}(\theta), p_{i,k}(\theta)) = \text{OR}(\rho_{\text{CUR}}(r_{i,j}, s_{i,j}, a_{i,j}), \rho_{\text{CUR}}(r_{i,k}, s_{i,k}, a_{i,k}))$$

for all $j, k = 1, \ldots, m_t$ and where $\text{OR}(a, b)$ is the odds ratio of $a$ and $b$. A bias parameter can be estimated for a data source if there is additional information on the absolute level of the quantity under consideration.

The third possibility is that the data source provides information on a mixture of prevalences across risk groups and hence indirect information on quantities of interest. For example, if data source $i$ provides an estimate of HCV prevalence across a mixture of all 3 risk groups, then

$$p_{i,j}(\theta) = \sum_g \omega_{i,g} \pi_g(r_{i,j}, s_{i,j}, a_{i,j}),$$

where $\omega_{i,g}$ is the mixture proportion for risk group $g$ and $\sum_g \omega_{i,g} = 1$. In this situation, additional information is needed on the mixture proportions in order for the data source to play a role in the synthesis.

Table 1 shows how each data source is used in the synthesis. A critical aspect of this synthesis is the lack of unbiased direct data on the absolute prevalence of ex-IDU. The C–R data source (Section 3.1) is assumed to provide unbiased estimates only of current IDU prevalence in London, $\rho_{\text{CUR}}(0, s, a)$. This is then used to derive estimates of $\alpha_0$ and log-odds ratio estimates for male gender and for the 30–44 age group. Meanwhile, the prevalence estimates of current and ex-IDU/heroin use obtained from the household surveys (Section 3.1) are assumed to be biased. The modeling of these biases is carried out as explained above and further explored in Section 6.1.

**Table 1. Relationship between data and model parameters**

<table>
<thead>
<tr>
<th>$i$</th>
<th>Data sources</th>
<th>No. of data points</th>
<th>Unbiased/biased/mixture</th>
<th>Quantity of interest</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C–R† Household surveys</td>
<td>4</td>
<td>Unbiased</td>
<td>$\rho_{\text{CUR}}(0, s, a)$</td>
<td>$\alpha_0, \alpha$</td>
</tr>
<tr>
<td>2</td>
<td>BCS</td>
<td>36</td>
<td>Biased</td>
<td>$\rho_{\text{CUR}}(r, s, a)$</td>
<td>$\alpha, b_{i,\text{CUR}}$</td>
</tr>
<tr>
<td>3</td>
<td>CJS</td>
<td>36</td>
<td></td>
<td>$\rho_{\text{EX}}(r, s, a)$</td>
<td>$\alpha, b_{i,\text{EX}}$</td>
</tr>
<tr>
<td>4</td>
<td>SPM</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>NATSAL‡</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>UA IDUs</td>
<td>364</td>
<td>Unbiased</td>
<td>$\pi_{\text{IDU}}(r, s, a, d, tss)$</td>
<td>$\delta_0, \delta_1, \delta_2$</td>
</tr>
<tr>
<td>7</td>
<td>UA GUM IDUs</td>
<td>8</td>
<td>Mixture</td>
<td>$\pi_{\text{CUR}}(r, s, a)$</td>
<td>$\delta_0, \delta_1, \omega_{\text{GUM,CUR}}$</td>
</tr>
<tr>
<td>8</td>
<td>UA GUM non-IDUs</td>
<td>8</td>
<td>Mixture</td>
<td>$\pi_{\text{EX}}(r, s, a)$</td>
<td>$\delta_0, \delta_1, \gamma_0, \gamma_{\text{GUM,CUR}}, \omega_{\text{GUM,IDU}}$</td>
</tr>
<tr>
<td>9</td>
<td>UA antenatal clinic§</td>
<td>4</td>
<td>Mixture</td>
<td>$\pi_{\text{CUR}}(r, 0, a)$</td>
<td>$\delta_0, \delta_1, \gamma_0, \gamma, \omega_{\text{REG,IDU}}$</td>
</tr>
<tr>
<td>10</td>
<td>Blood donors</td>
<td>18</td>
<td>Biased</td>
<td>$\pi_{\text{NON}}(r, s, a)$</td>
<td>$\gamma, b_{10,\text{NON}}$</td>
</tr>
<tr>
<td>11</td>
<td>Tested population</td>
<td>85</td>
<td>Biased</td>
<td>$\pi_{\text{CUR}}(r, s, a)$</td>
<td>$\delta_1, b_{11,\text{CUR}}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\pi_{\text{EX}}(r, s, a)$</td>
<td>$\delta_1, b_{11,\text{EX}}$</td>
</tr>
</tbody>
</table>

SPM, Survey of Psychiatric Morbidity.
†London and 15–44 ages only.
‡Fifteen to 44 ages only.
§Females, London and rest of England and Wales region, and 15–44 ages only.
Among the data sources informing HCV prevalence, the UA IDU data source (Section 3.2) is assumed to provide unbiased estimates of $\pi_{\text{IDU}}(r, s, a, d, \text{tss})$ and is used to derive estimates of $\delta_0$ and the log-odds ratios $\delta_1$ and $\delta_2$. The UA GUM clinic attendees and the 2 pregnant women studies provide information on HCV prevalence in a mixture of risk groups. The proportion of current IDUs in ever IDU GUM clinic attendees is estimated from NATSAL (National Centre for Social Research, 2005) and denoted $\omega_{\text{GUM,CUR}}$, while in the “non-IDU” GUM data, a proportion $\omega_{\text{GUM,IDU}}$ are assumed actually to be drug users. There are no direct data to inform the proportion $\omega_{\text{GUM,IDU}}$. This will be estimated from the synthesis and can be thought of as an adjustment parameter to resolve possible data conflicts with other data sources that estimate HCV prevalence in non-IDUs. The pregnant women surveys (Section 3.2) are a mixture of current, ex-, and non-IDUs. The proportion of pregnant women who ever injected drugs, denoted $\omega_{\text{PREG,IDU}}$, is also estimated from NATSAL (National Centre for Social Research, 2005). Among the ever IDU pregnant women, the population risk-group proportions are used to further divide women into current and ex-users. Finally, it is assumed that blood donors provide biased estimates of HCV prevalence in non-IDUs, while the tested population provides biased estimates of HCV prevalence separately in current and ex-IDUs. The IDU status of each tested individual is obtained from a questionnaire.

5.2 Auxiliary data

The injecting duration and TSS distributions. Two household studies, the BCS (Home Office Research Development and Statistics Directorate and BMRB Social Research, 2006) and the CJS (Home Office Research Development and Statistics Directorate and BMRB Social Research, 2005), provide information on the date at which ex-users started using heroin and hence on the TSS. For the CJS, data are also available on the time of cessation and hence on the injecting duration for 56 ex-users. Data are binned into 7 categories (0, 1–4, 5–9, 10–14, 15–19, 20–29, and 30–39 years) and are assumed to come from a multinomial distribution.

The duration and TSS distributions estimated from the ex-users will not be the same as the distribution in all users alive at $T$. This is because long-term users are more likely to be still injecting, and hence a sample of ex-IDUs will overrepresent the short-term users. An adjustment method, proposed by Kaplan (1997), accounts for the biases of sampling only current or ex-users. Here, we apply the method to obtain unbiased estimates of both the TSS and the injecting duration distributions for all users alive at the time of the survey (see Section B of the supplementary material available at Biostatistics online, http://www.biostatistics.oxfordjournals.org).

The age of first injecting. The UA programme collects data on injecting drug behavior through questionnaires. In particular, the age at starting injecting drug use is recorded. These data are used to estimate an AAFU distribution by region and gender. For each region and sex combination, ages of first use are binned into the following 6 categories: 10–14, 15–19, 20–24, 25–29, 30–39, and 40–59.

5.3 Likelihood and priors

We adopt a Bayesian approach to the estimation problem. The likelihood is built up as a product of independent binomial likelihoods for the prevalence data and independent multinomial likelihoods for the auxiliary data. The probability mass functions $f_D$ and $f_{\text{TSS}}$ are assumed to be constant in the 7 categories described in Section 5.2 (see “The injecting duration and TSS distributions”), while $f_{\text{AAFU}}$ is assumed constant in the 6 categories presented in Section 5.2 (see “The age of first injecting”). Dirchlet priors are then placed on these probability vectors with all parameters set equal to 1. The regression parameters $\alpha_0$, $\alpha$, $\delta_0$, $\delta_1$, $\delta_2$, $\gamma_0$, and $\gamma$ and all bias parameters are given independent non-informative Gaussian priors with mean 0 and variance $10^5$. The sensitivity of the oral fluid test used in the UA IDU data source is assumed
to be normal on the logit scale with mean 2.40 and variance 0.065, while the specificity is normal on the logit scale with mean 4.82 and variance 0.50, as derived from Judd and others (2003). The parameters \( \omega_{\text{GUM, CUR}} \), \( \omega_{\text{GUM, IDU}} \), and \( \omega_{\text{PREG, IDU}} \) are given Beta(1,1) prior distributions. All other quantities in the model are derived through functional relationships from these basic parameters (as described in Section 4).

Posterior distributions of the parameters are obtained using Markov chain Monte Carlo (MCMC) methods in WinBUGS (Spiegelhalter and others, 2003). The first 4000 iterations of the sampler were discarded as a burn-in. Inferences are based on 80,000 samples taken from 2 chains running in parallel. Convergence was assessed through the use of the Gelman–Rubin diagnostic (Brooks and Gelman, 1998).

### 5.4 Assessing Bayesian model fit and conflict

The standardized deviance is the deviance minus the deviance evaluated at the maximum likelihood estimates of the parameters (McCullagh and Nelder, 1983). For binomial data points \( y_{i,j} \), with \( i = 1, \ldots, N \) and \( j = 1, \ldots, m_i \), and model parameters \( \theta \), the contribution of data source \( i \) to the standardized deviance is

\[
d_i(\theta) = -2 \sum_{j=1}^{m_i} \left( y_{i,j} \log \frac{p_{i,j}(\theta)}{y_{i,j}/n_{i,j}} + (n_{i,j} - y_{i,j}) \log \frac{1 - p_{i,j}(\theta)}{1 - y_{i,j}/n_{i,j}} \right),
\]

and since we assume independence between data sources the standardized deviance for the model as a whole is

\[
D(\theta) = \sum_i d_i(\theta).
\]

The posterior mean (standardized) deviance can be used to assess overall goodness-of-fit and will have approximate sampling expectation \( M = \sum_i m_i \) if the model is true (Spiegelhalter and others, 2002). However, this approximation is reliable only if the \( n_{i,j} \) are large enough to provide considerable information about each \( p_{i,j} \) from the likelihood (McCullagh and Nelder, 1983; Kuss, 2002; S. Seaman, D. De Angelis, and A. Presanis, in preparation). Hence, in the case of binomial data where \( n_{i,j} \) are small, the posterior mean deviance is not useful as an absolute goodness-of-fit statistic. However, individual data source contributions to the posterior mean deviance can be of interest. In particular, comparing posterior mean deviance contributions across a number of models may indicate data sources that are conflicting with each other.

To assess potential conflicts using deviance contributions, we consider the following cross-validation approach. The model is refitted where each data source (rather than an individual data point) is, in turn, removed from the model. We then assess the posterior mean deviance contributions from all the remaining data sources. A drop in the posterior mean deviance contribution associated with data source \( a \) when data source \( b \) is removed would imply that the predictive accuracy of data source \( a \) is improved when data source \( b \) is removed from the model. Leaving out whole data sources ensures a less computationally intensive procedure than a full leave-one-out cross-validation where each data point is removed in turn. However, as a result, it is not uncommon to encounter problems of non-identifiability in parts of the parameter space when the data source that is removed is the only one informing those parameters. Due to the complexity of the relationship between the data and the parameters, it is often far from obvious which parameters are identified by which data sources. The proposed cross-validation can therefore reveal which data sources are vital to the synthesis.

Posterior predictive \( p \)-values (Gelman and others, 1996) are used to judge whether a given model is consistent with the data. Posterior predictive \( p \)-values are known to be conservative due to the double use of the data in generating the posterior distribution and calculating the tail area (Bayarri and Berger, 2000). For hierarchical models with random effects, we therefore calculate “mixed” predictive \( p \)-values (Marshall and Spiegelhalter, 2003), which provide a better approximation to full cross-validation \( p \)-values. Finally, we also use the DIC (Spiegelhalter and others, 2002) to choose between competing
models. Models with a lower DIC are better supported by the data and will have better out-of-sample predictive power.

6. Results

The evidence synthesis provides estimates of the prevalence of the risk-group populations, the prevalence of HCV by risk group, and the total prevalence of HCV in the population. Specific results are discussed in detail elsewhere (Health Protection Agency, 2006). Below, we illustrate some of the key model results in relation to the process of fitting and model selection. We start by considering bias modeling in the household surveys and then further investigate our chosen model in terms of goodness-of-fit and data consistency.

6.1 Bias modeling

Table 2 shows the change in posterior mean deviance and the DIC for 5 models that make different assumptions about biases in the household surveys. Model B1 assumes that all household surveys are unbiased in the estimation of current and ex-IDU prevalence. The second scenario (Model B2) assumes that the bias is common across all surveys, risk groups, and covariate combinations. Model B3 allows the bias to be different for estimates of current and ex-IDU prevalence. The fourth model (B4) allows the bias to be survey specific, while the fifth model (B5) uses both survey- and risk group–specific biases.

Model B1 is clearly inferior to the other 4 models that allow for bias in the survey estimates. The best-fitting model when the parameters are set to their mean value is obtained from the most complex model (B5). However, Models B2 and B3 have the lowest posterior mean deviance, which is a better summary of Bayesian model fit since it averages over the posterior distribution of parameter values. Model B2 also has the lowest DIC. In terms of the DIC, there is little to choose between Models B2 and B3. Model B2 estimates the bias in the household surveys, $e_b$, to have a median of 0.25 and 95% credibility interval (CrI) (0.17, 0.38), that is, a reported to true prevalence odds ratio of 0.25. Model B3 estimates $e_{b_{CDU}}$ to have median 0.24 and 95% CrI (0.16, 0.36) and $e_{b_{EX}}$ to have median 0.50 and 95% CrI (0.21, 1.23). The considerable lack of available information on the absolute level of ex-IDU prevalence results in the large uncertainty for $b_{EX}$. Nonetheless, the distributions of $b_{CDU}$ and $b_{EX}$ are very different, and therefore Models B2 and B3 provide epidemiologically diverse results. This is seen in the different estimates of ex-IDU prevalence: 2.7% (95% CrI (1.8, 4.0)) for Model B2 and 1.4% (0.6, 3.2) for Model B3. The higher posterior mean deviance in Models B4 and B5 is possible since the decrease in the classical deviance is not as large as the increase in the number of model parameters. Hence, there is no evidence from the data that the biases are survey specific.

The similarity between the DIC from B2 and B3 suggests that we should choose an appropriate model based on epidemiological interpretations rather than the goodness-of-fit statistics. The sampling frames of the household surveys are less likely to include current IDUs than ex-IDUs (Turner and others, 1992). This is due to the more chaotic lifestyles of current users and the higher propensity of being incarcerated, living in a hostel/unstable accommodation, or being homeless. This would induce a differential bias in the calculation of current and ex-IDU prevalence in households. Therefore, B3 seems the more plausible model, and we concentrate on this model for the remainder of the paper. The overall prevalence $\pi$ calculated from B3, using (2.1) and (2.2), is estimated as 0.60% (95% CrI (0.39, 0.97)).

6.2 Goodness of fit

A QQ plot of the posterior predictive $p$-values for Model B3 against the quantiles of a Uniform(0,1) distribution shows a reasonable fit to most of the data sources (Figure 2). There is slight evidence of lack
Table 2. Goodness-of-fit statistics and results for 5 models comparing the use of bias parameters in the household surveys

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>$\hat{D}$</th>
<th>$\hat{D}$</th>
<th>$p_D$</th>
<th>DIC</th>
<th>$\rho_{\text{CUR}}$ (%)</th>
<th>$\rho_{\text{EX}}$ (%)</th>
<th>$\pi_{\text{CUR}}$ (%)</th>
<th>$\pi_{\text{EX}}$ (%)</th>
<th>$\pi_{\text{NON}}$ (%)</th>
<th>$\pi$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>No bias</td>
<td>875</td>
<td>949</td>
<td>74</td>
<td>1022</td>
<td>0.26</td>
<td>0.73</td>
<td>33.7</td>
<td>19.9</td>
<td>0.094</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>$b_{i,j} = 0$</td>
<td>(0.22, 0.30)</td>
<td>(0.65, 0.81)</td>
<td>(30.3, 37.3)</td>
<td>(17.2, 22.8)</td>
<td>(0.048, 0.152)</td>
<td>(0.27, 0.39)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>Common bias across surveys and data points</td>
<td>827</td>
<td>902</td>
<td>75</td>
<td>976</td>
<td>0.67</td>
<td>2.69</td>
<td>32.7</td>
<td>18.9</td>
<td>0.098</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>$b_{i,j} = b$</td>
<td>(0.49, 0.93)</td>
<td>(1.83, 4.04)</td>
<td>(29.2, 36.5)</td>
<td>(16.3, 21.7)</td>
<td>(0.048, 0.157)</td>
<td>(0.60, 1.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>Risk group–specific bias</td>
<td>826</td>
<td>902</td>
<td>76</td>
<td>978</td>
<td>0.68</td>
<td>1.41</td>
<td>33.0</td>
<td>19.7</td>
<td>0.091</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>$b_{i,j} = \begin{cases} b_{\text{CUR}}, &amp; \text{estimate } j \text{ is for current IDUs} \ b_{\text{EX}}, &amp; \text{estimate } j \text{ is in ex-IDUs.} \end{cases}$</td>
<td>(0.49, 0.96)</td>
<td>(0.58, 3.19)</td>
<td>(29.3, 37.2)</td>
<td>(16.8, 22.7)</td>
<td>(0.046, 0.150)</td>
<td>(0.39, 0.97)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>Survey-specific bias</td>
<td>826</td>
<td>904</td>
<td>78</td>
<td>981</td>
<td>0.70</td>
<td>2.79</td>
<td>32.6</td>
<td>18.8</td>
<td>0.098</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>$b_{i,j} = b_i$</td>
<td>(0.50, 1.01)</td>
<td>(1.84, 4.27)</td>
<td>(29.2, 36.5)</td>
<td>(16.2, 21.6)</td>
<td>(0.049, 0.158)</td>
<td>(0.60, 1.21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B5</td>
<td>Survey- and risk group–specific bias</td>
<td>823</td>
<td>905</td>
<td>82</td>
<td>986</td>
<td>0.69</td>
<td>1.45</td>
<td>33.2</td>
<td>19.6</td>
<td>0.091</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>$b_{i,j} = \begin{cases} b_{i,\text{CUR}}, &amp; \text{estimate } j \text{ is for current IDUs} \ b_{i,\text{EX}}, &amp; \text{estimate } j \text{ is in ex-IDUs.} \end{cases}$</td>
<td>(0.49, 0.96)</td>
<td>(0.61, 3.31)</td>
<td>(29.4, 37.3)</td>
<td>(16.7, 22.7)</td>
<td>(0.046, 0.151)</td>
<td>(0.39, 0.99)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\hat{D} = \text{deviance evaluated at parameter means}$, $\hat{D} = \text{posterior mean deviance}$, $p_D = \text{effective number of parameters}$, and DIC = deviance information criterion. Posterior medians and 95% CrIs are presented for key parameters.
Fig. 2. Uniform QQ plots of posterior predictive $p$-values.
of fit in the UA GUM attenders and tested population data sources, suggesting greater variation in these sources than provided by the binomial distribution alone. Conversely, there is an indication of over-fitting in the IDU prevalence part of the model for the C–R and household survey data. Here, there are more \( p \)-values close to 0.5 than expected. This, however, may be an artifact due to the conservative nature of posterior predictive \( p \)-values rather than a deficiency of the model.

### 6.3 Cross-validation

A cross-validation whereby either the UA IDU or the blood donor data sources are removed causes poor mixing in the MCMCs of some regression parameters. The UA IDU data source is the only source to inform HCV prevalence log-odds ratios associated with injecting duration and TSS. Without this source, these parameters become only weakly identifiable (Gelfand and Sahu, 1999), mixing of the chains is poor, and convergence is not achieved, as detected by the Gelman–Rubin diagnostic (Brooks and Gelman, 1998). Hence, some statistics are not shown for cross-validations where the UA IDU or blood donor data sources are removed from the model.

Table 3 shows the deviance contributions for each data source for each cross-validation analysis. The first row of the table shows the deviance contributions from the model with all data sources included. The contributions vary in magnitude due to the different number of data points in each source. The table is divided into those sources that inform IDU prevalence (sources 1–5) and those that inform HCV prevalence (sources 6–11). In general, deviance contributions do not change dramatically as each data source is removed from the model, suggesting that there is little evidence of conflict between most sources. In particular, data on IDU prevalence (sources 1–5) are not found to be in conflict with data on HCV prevalence (sources 6–11), and with the current evidence structure this is not surprising. The 2 blocks of data are connected in the synthesis via their joint use of \( f_D \), \( f_{TSS} \), and \( f_{AAFU} \) (see Figure 1); yet these 3 distributions do not cause data conflict since there are no direct data on IDU or HCV prevalence for ex-users to constrain the synthesis. Direct data on ex-IDUs are therefore a necessary condition for assessing consistency between the 2 types of data. However, there is some evidence of conflict within the data blocks, such as between data sources 8 and 9 (the pregnant women sources). Here, the deviance contribution for one source decreases as the other is removed from the model, and vice versa. These contributions are shown

<table>
<thead>
<tr>
<th>Data source left out</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (Model B3)</td>
<td>2.6</td>
<td>46</td>
<td>45</td>
<td>25</td>
<td>24</td>
<td>503</td>
<td>27</td>
<td>7.9</td>
<td>9.6</td>
<td>21</td>
<td>141</td>
</tr>
<tr>
<td>1 (C–R)</td>
<td>2.5</td>
<td>44</td>
<td>48</td>
<td>26</td>
<td>22</td>
<td>503</td>
<td>27</td>
<td>8.0</td>
<td>9.4</td>
<td>21</td>
<td>140</td>
</tr>
<tr>
<td>2 (BCS)</td>
<td>2.1</td>
<td>45</td>
<td>25</td>
<td>24</td>
<td>503</td>
<td>27</td>
<td>8.0</td>
<td>9.5</td>
<td>21</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>3 (CJS)</td>
<td>3.7</td>
<td>45</td>
<td>26</td>
<td>24</td>
<td>503</td>
<td>27</td>
<td>7.8</td>
<td>9.6</td>
<td>21</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>4 (SPM)</td>
<td>2.7</td>
<td>46</td>
<td>46</td>
<td>24</td>
<td>503</td>
<td>27</td>
<td>7.9</td>
<td>9.6</td>
<td>21</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>5 (NATSAL)</td>
<td>2.0</td>
<td>47</td>
<td>44</td>
<td>25</td>
<td>503</td>
<td>26</td>
<td>7.8</td>
<td>9.5</td>
<td>21</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>6 (UA IDUs)</td>
<td>2.5</td>
<td>46</td>
<td>45</td>
<td>25</td>
<td>24</td>
<td>503</td>
<td>27</td>
<td>7.8</td>
<td>9.6</td>
<td>21</td>
<td>141</td>
</tr>
<tr>
<td>7 (UA GUM)</td>
<td>2.4</td>
<td>46</td>
<td>45</td>
<td>25</td>
<td>25</td>
<td>503</td>
<td>27</td>
<td>7.8</td>
<td>9.6</td>
<td>20</td>
<td>141</td>
</tr>
<tr>
<td>8 (UA antenatal)</td>
<td>2.6</td>
<td>46</td>
<td>45</td>
<td>25</td>
<td>24</td>
<td>503</td>
<td>26</td>
<td>5.8</td>
<td>20</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>9 (neonatal sample)</td>
<td>2.5</td>
<td>46</td>
<td>45</td>
<td>25</td>
<td>24</td>
<td>503</td>
<td>25</td>
<td>4.0</td>
<td>20</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>10 (blood donors)</td>
<td>2.5</td>
<td>46</td>
<td>45</td>
<td>25</td>
<td>24</td>
<td>503</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>20</td>
<td>141</td>
</tr>
<tr>
<td>11 (tested population)</td>
<td>2.4</td>
<td>47</td>
<td>45</td>
<td>25</td>
<td>24</td>
<td>503</td>
<td>27</td>
<td>7.8</td>
<td>9.4</td>
<td>21</td>
<td>141</td>
</tr>
</tbody>
</table>

* indicates convergence not achieved as detected by Gelman–Rubin diagnostic. SPM, Survey of Psychiatric Morbidity.
in bold in Table 3. The decrease in deviance is found to be predominantly caused by 2 data points; the estimate of HCV prevalence in pregnant women aged 30–44 years in Northern and Yorkshire region (observed proportion 0.16%) from the antenatal source and the estimate in pregnant women aged 30–44 years in North Thames and Bedfordshire region (0.13%) from the neonatal source. These 2 data points in our model are assumed to estimate the same quantity: prevalence in pregnant women aged 30–44 years in the rest of England and Wales region. The conflict suggests that estimates of prevalence are not homogenous within this region.

The cross-validation also identifies data sources with high leverage. Table 4 shows the posterior medians and corresponding standard errors for $\rho_{\text{CUR}}$, $\rho_{\text{EX}}$, $\pi_{\text{CUR}}$, $\pi_{\text{EX}}$, $\pi_{\text{NON}}$, and $\pi$. The first row shows the results from the prior model, without any data sources. The use of non-informative priors throughout ensures that our prior inferences about the parameters are vague. Without the C–R data source, all the main quantities of interest have large standard errors. This is due to no information being available on the absolute size of the IDU epidemic and hence on some of the mixture proportions that relate HCV prevalence in the surveillance data to $\pi_{\text{CUR}}$, $\pi_{\text{EX}}$, and $\pi_{\text{NON}}$. The UA IDU data source is critical to the estimation of HCV prevalence in current and ex-IDUs, and without it some of the HCV IDU regression parameters are unidentifiable. This causes problems in estimating $\pi_{\text{CUR}}$, $\pi_{\text{EX}}$, and hence $\pi$. The blood donor data source is also critical to estimation of some of the HCV non-IDU regression parameters and hence $\pi_{\text{NON}}$. Without the NATSAL data source, the posterior median estimate of ex-IDU prevalence increases from 1.41% to 1.73% and the corresponding standard error increases from 0.67% to 0.94%. The UA pregnant women and neonatal sources are also important in reducing the uncertainty in the estimates of HCV prevalence in non-IDUs.

### 6.4 Extending the model

Model B3 can be extended to allow for unexplained heterogeneity in the data, in the form of extra binomial variation. This can be achieved by introducing random-effect error terms $\epsilon_{i,j}$ with mean 0 and variance $\sigma_{i,j}$ for data sources $i = 1, \ldots, 11$ and data points $j = 1, \ldots, m_i$:

$$y_{i,j} \sim \text{Binomial}(n_{i,j}, p_{i,j}(\theta)),$$

$$\logit(p_{i,j}(\theta)) = \mu_{i,j}(\theta) + \epsilon_{i,j},$$

$$\epsilon_{i,j} \sim N(0, \sigma_{i,j}).$$

Here, $\mu_{i,j}(\theta)$ is a function of the model parameters as shown in Table 1. Two random-effects models that make different assumptions concerning the exchangeability of the $\epsilon$ parameters are investigated. Model R1 allows exchangeable residual error terms between and within all data sources, that is, $\sigma_{i,j} = \sigma$ for all $i = 1, \ldots, 11$ and $j = 1, \ldots, m_i$. Model R2, however, assumes that the residual error terms are exchangeable only with other similar estimates, as follows:

$$\sigma_{i,j} = \begin{cases} 
\sigma_1, & \text{if estimate is of IDU prevalence,} \\
\sigma_2, & \text{if estimate is of HCV prevalence in high-risk population,} \\
\sigma_3, & \text{if estimate is of HCV prevalence in low-risk population.} 
\end{cases}$$

Therefore, residual error terms are considered exchangeable between all estimates of IDU prevalence, all estimates of HCV prevalence in high-risk populations, and all estimates of HCV prevalence in low-risk populations.

Table 5 shows that both random-effects models dramatically decrease the posterior mean deviance. The additional number of effective parameters for Model R1 compared with the fixed-effect model (B3) is 59, but this additional complexity is justified by a much lower DIC statistic. Model R2 provides no
Table 4. Influence of each data source on inferences about prevalence of IDU and HCV using a cross-validation technique. Shown are posterior medians and standard errors. Model B3 is the initial model from which data sources are subsequently removed.

<table>
<thead>
<tr>
<th>Data source left out</th>
<th>$\rho_{CUR}$</th>
<th>$\rho_{EX}$</th>
<th>$\pi_{CUR}$</th>
<th>$\pi_{EX}$</th>
<th>$\pi_{NON}$</th>
<th>$\pi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (prior model)</td>
<td>0.281 [0.213]</td>
<td>0.190 [0.156]</td>
<td>0.452 [0.340]</td>
<td>0.456 [0.332]</td>
<td>0.455 [0.375]</td>
<td>0.505 [0.268]</td>
</tr>
<tr>
<td>None (Model B3)</td>
<td>0.0068 [0.0012]</td>
<td>0.0141 [0.0067]</td>
<td>0.330 [0.020]</td>
<td>0.197 [0.015]</td>
<td>$9.1 \times 10^{-4}$ [2.7 $\times 10^{-4}$]</td>
<td>0.0060 [0.0015]</td>
</tr>
<tr>
<td>1 (C–R)</td>
<td>0.1025 [0.1430]</td>
<td>0.2821 [0.3686]</td>
<td>0.338 [0.190]</td>
<td>0.186 [0.112]</td>
<td>$5.7 \times 10^{-4}$ [5.2 $\times 10^{-4}$]</td>
<td>0.0998 [0.1344]</td>
</tr>
<tr>
<td>2 (BCS)</td>
<td>0.0067 [0.0012]</td>
<td>0.0136 [0.0070]</td>
<td>0.343 [0.022]</td>
<td>0.200 [0.017]</td>
<td>$9.1 \times 10^{-4}$ [2.7 $\times 10^{-4}$]</td>
<td>0.0060 [0.0016]</td>
</tr>
<tr>
<td>3 (CJS)</td>
<td>0.0067 [0.0012]</td>
<td>0.0135 [0.0066]</td>
<td>0.329 [0.020]</td>
<td>0.194 [0.015]</td>
<td>$9.1 \times 10^{-4}$ [2.8 $\times 10^{-4}$]</td>
<td>0.0058 [0.0014]</td>
</tr>
<tr>
<td>4 (SPM)</td>
<td>0.0064 [0.0011]</td>
<td>0.0134 [0.0063]</td>
<td>0.333 [0.020]</td>
<td>0.196 [0.016]</td>
<td>$9.1 \times 10^{-4}$ [2.8 $\times 10^{-4}$]</td>
<td>0.0057 [0.0014]</td>
</tr>
<tr>
<td>5 (NATSAL)</td>
<td>0.0078 [0.0017]</td>
<td>0.0173 [0.0094]</td>
<td>0.323 [0.020]</td>
<td>0.198 [0.015]</td>
<td>$9.3 \times 10^{-4}$ [2.8 $\times 10^{-4}$]</td>
<td>0.0069 [0.0021]</td>
</tr>
<tr>
<td>6 (UA IDUs)</td>
<td>0.0069 [0.0012]</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>7 (UA GUM)</td>
<td>0.0069 [0.0012]</td>
<td>0.0152 [0.0070]</td>
<td>0.326 [0.020]</td>
<td>0.191 [0.016]</td>
<td>$8.4 \times 10^{-4}$ [2.7 $\times 10^{-4}$]</td>
<td>0.0060 [0.0015]</td>
</tr>
<tr>
<td>8 (UA antenatal)</td>
<td>0.0069 [0.0012]</td>
<td>0.0142 [0.0070]</td>
<td>0.332 [0.020]</td>
<td>0.196 [0.015]</td>
<td>$12.4 \times 10^{-4}$ [3.8 $\times 10^{-4}$]</td>
<td>0.0063 [0.0016]</td>
</tr>
<tr>
<td>9 (neonatal sample)</td>
<td>0.0068 [0.0012]</td>
<td>0.0123 [0.0067]</td>
<td>0.333 [0.020]</td>
<td>0.199 [0.016]</td>
<td>$26.2 \times 10^{-4}$ [14.3 $\times 10^{-4}$]</td>
<td>0.0075 [0.0020]</td>
</tr>
<tr>
<td>10 (blood donors)</td>
<td>0.0068 [0.0012]</td>
<td>0.0119 [0.0060]</td>
<td>0.334 [0.020]</td>
<td>0.200 [0.015]</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>11 (tested population)</td>
<td>0.0068 [0.0011]</td>
<td>0.0148 [0.0070]</td>
<td>0.343 [0.021]</td>
<td>0.191 [0.016]</td>
<td>$9.0 \times 10^{-4}$ [2.8 $\times 10^{-4}$]</td>
<td>0.0061 [0.0015]</td>
</tr>
</tbody>
</table>

* indicates convergence not achieved as detected by Gelman–Rubin diagnostic. SPM, Survey of Psychiatric Morbidity.
Table 5. DIC and results associated with the fixed-effect Model B3 (risk group-specific biases) and 2 random-effects models. Model R1 assumes exchangeability of residual variation across all data points and sources. Model R2 assumes that residual variation is exchangeable only within and between similar data sources (namely, estimates of IDU prevalence, estimates of HCV prevalence in high-risk populations, and estimates of HCV prevalence in low-risk populations).

<table>
<thead>
<tr>
<th>Model</th>
<th>$\tilde{D}$</th>
<th>$p_D$</th>
<th>DIC</th>
<th>$\rho_{CUR}$ (%)</th>
<th>$\rho_{EX}$ (%)</th>
<th>$\pi_{CUR}$ (%)</th>
<th>$\pi_{EX}$ (%)</th>
<th>$\pi_{NON}$ (%)</th>
<th>$\pi$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effects</td>
<td>B3</td>
<td>902</td>
<td>76</td>
<td>978</td>
<td>0.68</td>
<td>1.41</td>
<td>33.0</td>
<td>19.7</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.49, 0.96)</td>
<td>(0.58, 3.19)</td>
<td>(29.3, 37.2)</td>
<td>(16.8, 22.7)</td>
<td>(0.046, 0.150)</td>
</tr>
<tr>
<td>Random effects</td>
<td>R1</td>
<td>800</td>
<td>135</td>
<td>935</td>
<td>0.69</td>
<td>1.46</td>
<td>33.5</td>
<td>20.3</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.45, 1.04)</td>
<td>(0.61, 3.55)</td>
<td>(29.6, 37.8)</td>
<td>(17.2, 23.7)</td>
<td>(0.043, 0.158)</td>
</tr>
<tr>
<td></td>
<td>R2</td>
<td>800</td>
<td>137</td>
<td>937</td>
<td>0.68</td>
<td>1.41</td>
<td>33.2</td>
<td>20.3</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.46, 1.00)</td>
<td>(0.63, 3.95)</td>
<td>(29.4, 37.6)</td>
<td>(17.1, 23.5)</td>
<td>(0.044, 0.160)</td>
</tr>
</tbody>
</table>

$p_D =$ Effective number of parameters.

additional improvement over Model R1 according to the DIC. However, this model may have a more natural interpretation since Model R1 allows exchangeable residual error terms across all data sources, which is slightly implausible. Mixed predictive checks calculated in Models R1 and R2 give $p$-values which are similar to posterior predictive $p$-values (data not shown).

All 3 models estimate HCV antibody prevalence to be close to 0.6%, but the uncertainty is larger in both random-effects models. In particular, the upper 95% CrI reaches 1.13% in Model R2 compared with 0.97% in Model B3. Both random-effects models also induce wider CrIs for many other key parameters. The posterior medians increase slightly for $\rho_{EX}$, $\pi_{CUR}$, and $\pi_{EX}$ and decrease for $\pi_{NON}$ from 0.091% (B3) to 0.084% (R1).

7. Discussion

In this paper, we have developed a model for estimating HCV prevalence in England and Wales using available population and surveillance data on IDU and HCV prevalence. We functionally link all data to key parameters of interest using a multiparameter evidence synthesis approach. In developing the model, we have discovered that extra data on major aspects of the IDU epidemic are needed to identify some of these key parameters. In particular, we have used auxiliary data informing the distributions of injecting duration and time of starting injecting in order to estimate the proportion of current to ex-users in the population. This approach builds on work by Kaplan (1997) in which an underlying process that starts at an initial point in time and has a random duration (e.g. the injecting of drugs or HIV infection to AIDS) is only observed at a snapshot in time.

A number of competing models have been investigated using goodness-of-fit statistics such as the posterior mean deviance, the DIC, and predictive $p$-values. The deviance and DIC have been used to compare the relative fit of a number of models and have been useful in eliminating a number of potential models, for example, Model B1. However, using these statistics, we have not been able to conclusively understand how biases in the household surveys are acting on estimates of IDU prevalence. Two competing models, B2 and B3, have been found to be statistically equivalent in terms of fit and yet provide very different estimates of prevalence. We hope to collect further auxiliary data on the IDU epidemic, in terms of injecting duration and time of starting, to allow a critical evaluation of these 2 models in the future.

We have not used the posterior mean (standardized) deviance as an absolute measure of fit since much of our data are sparse, that is, $n_{i,j}$ are small for certain covariate combinations. It has long been known that
residual deviances calculated in a frequentist logistic regression analysis with sparse data are unreliable in assessing absolute fit (Kuss, 2002). Instead, we use posterior predictive $p$-values to determine the absolute fit of a model. The advantage of these $p$-values is that they are easy to calculate in an MCMC framework. However, posterior predictive $p$-values have been criticized as being conservative (suggesting a better model fit than is actually the case) since each data point is used to create its own predictive distribution (Stern and Cressie, 2000; Bayarri and Berger, 2000). An alternative to posterior predictive model checks is to calculate $p$-values from a leave-one-out cross-validation. The obvious disadvantage with this method is the immense computational aspect involved with a complex model such as we have presented in this paper.

We have shown that, through a cross-validatory framework, conflicts between data sources can be assessed. We propose constructing a table of posterior mean deviance contributions for each data source and investigating how these change as other sources are removed from the model. As far as we are aware, this is the first time such a table has been constructed. In particular, our model has shown some lack of agreement for parameters associated with HCV prevalence in non-IDUs. However, in general, there appears to be reasonable consistency between the data sources. The inconsistencies that could arise in this synthesis can be seen in Table 1. For example, each log-odds ratio in the regression of IDU prevalence has up to 5 data sources providing information on it. The fact that no inconsistencies are detected suggests that the data are homogeneous and adds some strength to the estimates we obtain.

This synthesis potentially allows the data sources that inform HCV prevalence in IDUs to affect inferences on IDU prevalence, and vice versa. These 2 “epidemics” are connected via the distributions for TSS, injecting duration, and AAFU. The current evidence structure does not, as it stands, allow us to verify that these 2 sources of evidence are consistent because there are no direct data on either the absolute prevalence of ex-IDU or HCV prevalence in ex-IDUs to validate predictions derived from current users and the TSS and injecting duration distributions. However, the cross-validatory framework has been useful in determining the influence of each data source on key parameters. We have found that a number of data sources are vital to our synthesis, and without them useful inferences cannot be made unless the model structure is changed. More data are currently being sought to improve accuracy for some of these key parameters, in particular the absolute prevalence of ex-IDU, and the injecting duration distribution. This will then allow us to formally assess the consistency of the 2 main blocks of data, on IDU and HCV prevalence.

**Funding**

UK Department of Health (AIDB 2/29 to M.S.).

**Acknowledgments**

We would like to thank Vivian Hope, Koye Balogun, Lisa Brant, Helen Harris, and Mary Ramsay for the provision of hepatitis C surveillance data and their expertise in the field. The authors would also like to thank those who commented on an earlier version of this paper, which was presented to the workshop “Explaining the results of a complex probabilistic modelling exercise: conflict, consistency and sensitivity analysis,” sponsored by the MRC Population Health Sciences Research Network, in Cambridge, September 7–8, 2006. Conflict of Interest: None declared.

**References**


[Received March 9, 2007; revised December 21, 2007; accepted for publication January 22, 2008]