

## Spatial mapping of hepatitis C prevalence in recent injecting drug users in contact with services

R. J. HARRIS<sup>1\*</sup>, V. D. HOPE<sup>1,2</sup>, A. MORONGIU<sup>1</sup>, M. HICKMAN<sup>3</sup>, F. NCUBE<sup>1</sup>  
AND D. DE ANGELIS<sup>1,4</sup>

<sup>1</sup> Health Protection Agency, Centre for Infections, London, UK

<sup>2</sup> Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK

<sup>3</sup> School of Social and Community Medicine, University of Bristol, Bristol, UK

<sup>4</sup> MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK

(Accepted 3 August 2011; first published online 30 August 2011)

### SUMMARY

In developed countries the majority of hepatitis C virus (HCV) infections occur in injecting drug users (IDUs) with prevalence in IDUs often high, but with wide geographical differences within countries. Estimates of local prevalence are needed for planning services for IDUs, but it is not practical to conduct HCV seroprevalence surveys in all areas. In this study survey data from IDUs attending specialist services were collected in 52/149 sites in England between 2006 and 2008. Spatially correlated random-effects models were used to estimate HCV prevalence for all sites, using auxiliary data to aid prediction. Estimates ranged from 14% to 82%, with larger cities, London and the North West having the highest HCV prevalence. The methods used generated robust estimates for each area, with a well-identified spatial pattern that improved predictions. Such models may be of use in other areas of study where surveillance data are sparse.

**Key words:** England, hepatitis C, prevalence, sentinel surveillance, substance abuse (intravenous).

### INTRODUCTION

The main route of transmission of the hepatitis C virus (HCV) in developed countries is through injection drug use, which accounts for up to 80% of infections in England and Wales [1]. Overall it is estimated that HCV prevalence is about 50% in England, although in other European Union countries national prevalence of HCV infection in injecting drug users (IDUs) ranges from 12% to >75% [2]. There is also likely to be great variation of HCV prevalence within individual countries as well as between them. In the UK, serosurveillance studies suggest a substantial regional

variation [3] with a greater than threefold difference in HCV prevalence between individual sites, e.g. from 60% in Manchester to <20% in North East England and South Wales [4, 5]. Similar variations have been reported within other countries, e.g. in Italy prevalence in IDUs ranges from 31% to 87% [2]. Clearly, the risk of HCV infection can be very different between areas (with estimates of HCV incidence in IDUs in the UK ranging from <5/100 person-years to 40/100 person-years [4, 6, 7]), and therefore opportunities for prevention and scale of the intervention coverage required to reduce transmission also vary [8, 9].

IDUs are a difficult to reach population, partially covered and represented by a mixture of data sources (such as data from needle exchanges, specialist drug treatment, prisons) [10]. Information on HCV

\* Author for correspondence: Mr R. J. Harris, Statistics Unit, Health Protection Agency, Centre for Infections, 61 Colindale Avenue, London NW9 5EQ, UK.  
(Email: ross.harris@hpa.org.uk)

prevalence in IDUs is not available through routine laboratory surveillance of diagnostic tests, because reporting of exposure categories is incomplete or missing and because not all those at risk come forward for testing. Public health surveillance of HCV infection, therefore, often relies upon sentinel and other surveys of IDUs that purposively recruit through community services and settings [11]. In England, the unlinked anonymous monitoring survey (UAM) of IDUs [12, 13] monitors the prevalence of antibodies to HCV (anti-HCV) and HIV (anti-HIV) in those attending a national sample of specialist services for drug users in about one-third of 149 local Drug Action Teams (DATs) responsible for commissioning services in England. It is impractical and too costly to undertake such surveys in every local area. Further, some of the samples are too small to provide direct estimates of local area anti-HCV prevalence in the sampled areas as the survey's focus is on producing national data [13]. Statistical models are therefore required to use the available data from the survey and other auxiliary sources to derive local estimates.

Synthetic estimation, which uses simple regression models of prevalence in terms of chosen auxiliary variables in the sampled areas, is commonly applied in situations such as these to obtain estimates in areas that have not been sampled [14–16]. However, this approach assumes a relationship between the outcome and the auxiliary (predictor) variables that may be incorrect, relatively weak and subject to systematic (e.g. spatial) variation; and does not correctly account for uncertainty in the estimated regression coefficients. Random-effects models [16, 17] are often used to reduce the uncertainty of estimates by borrowing strength via shrinkage, but do not account for the uncertainty of estimates in the non-sampled areas, which reduce to the simple synthetic estimator where the random effect is assumed to be zero.

Instead, we propose to overcome this problem by assuming that neighbouring areas tend to be more similar than distant ones, i.e. there is spatial correlation. The inclusion of spatially correlated random effects then allows the estimation of random effects in non-sampled areas, also incorporating an appropriate degree of uncertainty in the prevalence estimate (A. Saei & R. Chambers, University of Southampton, unpublished data). In this paper we propose models of this form for the estimation of local HCV prevalence in recent IDUs in contact with services in England, extending them to a Bayesian framework as described by Gómez-Rubio (N. Gómez Rubio,

S. Best, *et al.*, unpublished data) and, for example, Best *et al.* [18].

## MATERIALS AND METHODS

### Data sources

The principal data source used is the UAM survey of IDUs, which samples from harm reduction and treatment services in selected DATs, these being chosen to give a broad range of geographical and urban/rural settings. Respondents voluntarily and anonymously provide an oral fluid sample for testing [19], and complete a brief questionnaire. The survey received multi-centre ethics committee approval. In the analyses here only those that have injected in the last year are considered (termed 'recent injectors'). Data are in the form of a numerator and denominator for participants from each DAT area, year, age group and sex combination; with age split into three age groups: 15–29, 30–44 and 45–59 years, and for the years 2006–2008 (Table 1).

Auxiliary information on crime, uptake of benefits, population density and age structure, deprivation and health is available from the Office of National Statistics [20] (ONS) at the Local Authority District (LAD) level. ONS data in terms of counts or numbers were summed to DAT-level totals if the DAT covered more than one LAD. Per-person rates or proportions for each DAT were calculated using ONS estimates of population size [21], and log rates (or logit of proportions) were used as covariates in analyses to aid the prediction of prevalence estimates.

The final data source is the National Drug Treatment Monitoring System [22] (NDTMS) which collects data on drug users receiving addiction treatment. This is used to relate prevalence estimates from the UAM data to the treated population, under the assumption that those in treatment are representative of those in contact with all services. These data consist of numbers of recent injectors (within the last year) in treatment by age and sex for each DAT (Table 1).

### HCV prevalence model

Observed HCV prevalence  $\hat{p}_{d,t,a,s}$  for DAT  $d$ , year  $t$ , age  $a$  and sex  $s$  was modelled via logistic regression, similar to that described by Besag *et al.* [23], with the following basic form:

$$\text{logit}(\hat{p}_{d,t,a,s}) = \alpha_{a,s} + \beta_t t + \beta_{\text{pop},1} \text{pop} \cdot 1_{d,t} \dots + \beta_{\text{pop},n} \text{pop} \cdot n_{d,t} + v_d + u_d, \quad (1)$$

Table 1. Summary of data on recent injectors from the National Drug Treatment Monitoring System (NDTMS) in 2008 and unlinked anonymous monitoring (UAM) survey of injecting drug users, England, 2006–2008 aggregated

	Recent injectors, NDTMS (%)	Recent injectors, UAM (%)	UAM observed anti-HCV (prevalence)
Age (yr)			
15–29	19 080 (35.0 %)	2808 (41.6 %)	836 (29.8 %)
30–44	31 189 (57.2 %)	3471 (51.4 %)	1734 (50.0 %)
45–59	4279 (7.8 %)	473 (7.0 %)	310 (65.5 %)
Males	41 733 (76.5 %)	5095 (75.5 %)	2209 (43.4 %)
Females	12 799 (23.5 %)	1657 (24.5 %)	671 (40.5 %)
Total	54 341	6752	2880 (42.7 %)

Anti-HCV, Antibodies to hepatitis C virus.

where  $\alpha_{a,s}$  is an age-sex specific intercept,  $\beta_t$  is the effect of year, and  $\beta_{\text{pop},i}$  ( $i = 1, 2, 7 \dots, n$ ) are population-level effects from auxiliary ONS data (e.g. log rate of drug offences). The component  $v_d$  is an unstructured random effect at the DAT level, distributed independently as  $v_d \sim N(0, \sigma_v^2)$  if the area is ever sampled in 2006–2008 and zero otherwise. The  $u_d$  are conditionally autoregressive terms for the spatially correlated random effects at the DAT level. These are conditional on the group of surrounding neighbours (not including  $d$ ),  $u_{-d}$ , with the form

$$u_d | u_{-d} \sim N\left(\sum_{j \in \delta_d} \frac{u_j}{|\delta_d|}, \frac{\sigma_u^2}{|\delta_d|}\right), \tag{2}$$

where  $\delta_d$  is the group of contiguous neighbours that share a boundary with area  $d$  and  $|\delta_d|$  the number of neighbours, and the sum of the random effects  $u_d$  is constrained to equal zero. Deviation from the prevalence estimated by auxiliary information is therefore shared between the unstructured and spatially correlated random effects: the stronger the spatial correlation, the more variation being explained by the  $u_d$ .

The oral fluid test for antibodies to HCV (anti-HCV) has a sensitivity of 91.7% and a specificity of 99.2% [19]. The observed prevalence  $\hat{p}_{d,t,a,s}$  is related to the true prevalence  $p_{d,t,a,s}$  as described in Sweeting *et al.* [3].

Overall DAT-level prevalence for DAT  $d$  in year  $t$ ,  $p_{d,t}$ , is expressed as a weighted average of the prevalence estimates described above,  $p_{d,t,a,s}$ , with weighting according to the age-sex distribution of recent injectors in treatment from the NDTMS

data, i.e.:

$$p_{d,t} = \left( \sum_{a,s} p_{d,t,a,s} N_{d,t,a,s} \right) / \sum_{a,s} N_{d,t,a,s}, \tag{3}$$

where  $N_{d,t,a,s}$  is the number of recent injectors in DAT  $d$ , year  $t$ , age  $a$  and sex  $s$ . This weighting relates estimates from the UAM survey data to the population of recent injectors in treatment, correcting for any over- or under-sampling of demographic groups in the UAM survey data.

Estimation of  $\hat{p}_{d,t,a,s}$  in equation (1) is performed in a Bayesian framework, requiring specification of prior distributions for each parameter. A uniform prior is placed on standard deviations of unstructured and spatial random effects ( $\sigma_v$  and  $\sigma_u$ ), this form being shown to have good properties for hierarchical models [24]. Inverse gamma priors on variances were also investigated to test robustness under different assumptions. All regression parameters and intercept terms in the logistic model were given flat prior distributions (priors with equal probability on the entire real line).

Posterior distributions were obtained via Markov Chain-Monte Carlo techniques implemented in WinBUGS [25]. Medians and 95% credible intervals (CrIs) were on the basis of 50 000 iterations from two chains running in parallel, following a 5000 iteration ‘burn-in’ period. Convergence was assessed through the use of the Brooks–Gelman diagnostic [26].

### Model comparison

Various potential covariates and alternative model formulations were investigated, with further details

Table 2. Comparison statistics for models of prevalence of antibodies to hepatitis C virus in injecting drug users, England, 2006–2008

Included covariate	$D(\bar{\vartheta})$	$p_D$	DIC	CV deviance	Covariate OR per s.d.	$\sigma_v$	$\sigma_u$
None	2063.6	52.8	2116.4	964	n.a.	0.406	0.829
Proportion aged <35 yr	2062.8	51.7	2114.4	853	1.35	0.326	0.771
Drug crime	2070.9	52.1	2123.0	868	1.25	0.357	0.717
Violent crime	2067.1	53.3	2120.4	926	1.11	0.356	0.861
Sex crime	2061.0	54.1	2115.1	1048	0.94	0.452	0.838
Theft crime	2066.7	52.5	2119.2	875	1.31	0.319	0.849
Population density	2063.5	52.4	2115.9	946	1.31	0.398	0.776
Conception in under-18s	2063.7	52.8	2116.5	940	1.22	0.356	0.870
Income support claimants	2063.1	52.4	2115.4	939	1.38	0.360	0.790
Unemployment benefit claimants	2064.9	52.8	2117.7	956	1.20	0.384	0.839
Environmental deprivation	2063.7	52.9	2116.6	971	1.18	0.427	0.792

$D(\bar{\vartheta})$ , Posterior mean deviance;  $p_D$ , effective number of model parameters; DIC, deviance information criteria ( $D(\bar{\vartheta}) + p_D$ ); CV, cross-validation deviance (measure of predictive performance); OR, odds ratio per standard deviation increase of covariate;  $\sigma_v$ ,  $\sigma_u$ , standard deviations of unstructured ( $v$ ) and spatially correlated random effects ( $u$ ).

shown in the Results section. Models were compared via the deviance information criterion [27] (DIC), a measure of model fit plus a penalty for model complexity, which can be expressed as  $DIC = D(\bar{\vartheta}) + p_D$ , where  $D(\bar{\vartheta})$  is the posterior mean deviance and  $p_D$  is the effective number of parameters in the model. We also assessed the predictive performance of models via leave one out cross-validation (CV) [28–30], where data for each of the  $k$  sampled DATs are omitted in turn and the model is estimated based on data from the  $k-1$  remaining DATs. Measures of discrepancy can then be calculated using observed and predicted values for the omitted DAT, such as the deviance, which can then be summed across all DATs to provide a summary statistic for the model being tested. Given that the data have a binomial distribution, the CV deviance takes the form

$$\text{dev}_{CV} = \sum_{d=1:k} -2[r_d \log p_{d(-d)} + (n_d - r_d) \log(1 - p_{d(-d)})], \quad (4)$$

where  $\text{dev}_{CV}$  is the sum of CV deviances,  $p_{d(-d)}$  is HCV prevalence estimated in DAT  $d$  with data for DAT  $d$  omitted, and  $r_d$  and  $n_d$  the observed numerator and denominator in DAT  $d$ . For clarity, the summation over year, age and sex is not shown in the above equation. Use of equation (4) is intended to provide an assessment of out-of-sample predictive performance, which is a primary goal of the model due to the large number of non-sampled DATs for which estimates are required, and we chose models

that minimize  $\text{dev}_{CV}$  and therefore provide the closest values to ‘unseen’ data.

## RESULTS

Data from 6752 recent injectors participating in the UAM survey during 2006–2008 were used in the analyses. Of these, two-fifths (41.6%) were aged 15–29, and a quarter (24.5%) were female (Table 1). Female participants tended to be younger than males with 51.2% aged 15–29 years compared to 38.4% of males. Overall, 2880 (42.7%) tested positive for HCV antibodies.

Of the 149 DATs, 52 were sampled during 2006–2008. The median sample size taken from each DAT in 2008 was 32, with interquartile range 23–64, and a range from 2 to 206. The small sample size in some of the DATs reflects recruitment only through treatment services where most of those participating were ex-injectors (i.e. had not injected in the last year). Of the nine government office regions, the best represented was the South East, with 10/19 (53%) DATs sampled, while in Yorkshire and the Humber only 1/15 (7%) was sampled. Other regions ranged from 29% to 42% of DATs being sampled.

### Model choice

Results comparing different auxiliary covariates are shown in Table 2. The model including logit proportion of adults aged <35 years (proportion <35)

Table 3. Comparison statistics for alternative models of prevalence of antibodies to hepatitis C virus in injecting drug users, England, 2006–2008

Description	$D(\bar{\vartheta})$	$p_D$	DIC	CV deviance	Age < 35, OR per s.d.	Drug crime, OR per s.d.	$\sigma_v$	$\sigma_u$
Final model	2064.9	51.7	2116.6	840	1.30	1.08	0.283	0.781
Omissions								
Omission of spatial effect	2065.8	53.0	2118.7	1007	1.30	1.14	0.629	n.a.
Omission of unstructured random effect	2070.4	53.9	2124.3	795	1.31	1.09	n.a.	0.924
Omission of spatial and unstructured random effect	2450.9	10.0	2460.9	918	1.11	1.50	n.a.	n.a.
Omission of age and sex effects	2208.8	48.0	2256.8	988	1.35	1.07	0.307	0.837
Omission of year effect	2062.8	50.1	2113.0	834	1.32	1.05	0.311	0.775
Extensions								
Separate auxiliary covariate effects for each year	2065.3	56.0	2121.3	835	1.28 1.36 1.21	1.18 1.04 1.10	0.326	0.739
Separate $u$ for each year	2048.7	72.4	2121.1	946	1.24	1.22	0.595	0.273 0.190 0.271
Separate $v$ for each year	2047.7	72.9	2120.6	800	1.25	1.15	0.165 0.123 0.218	0.885
Separate $u$ and $v$ for each year	2046.8	103.3	2150.1	841	1.12	1.40	0.174 0.392 0.456	0.480 0.750 0.701

$D(\bar{\vartheta})$ , Posterior mean deviance;  $p_D$ , effective number of model parameters; DIC, deviance information criteria ( $D(\bar{\vartheta}) + p_D$ ); CV, cross-validation deviance (measure of predictive performance); OR, odds ratio per standard deviation increase of covariate;  $\sigma_v$ ,  $\sigma_u$ , standard deviations of unstructured ( $v$ ) and spatially correlated random effects ( $u$ ).

Models use proportion of adults aged < 35 years and drug crime as covariates.

Where components vary over time, values for years 2006, 2007 and 2008 are shown in order.

gave the best DIC and CV deviance scores, although many covariate choices had a similar DIC score. It is interesting to note that some variables, such as sex offences, gave reasonable DIC scores but poor predictive performance. This can be explained by the increase in  $\sigma_v$ , as the random effects give a smooth fit for sampled DATs but do not aid in the prediction of non-sampled DATs, for which  $v$  reduces to zero. Choosing ‘proportion < 35’ as the principal auxiliary variable, we found that the inclusion of further covariates did not improve the model, with the exception of drug offences. These two covariates were therefore used in subsequent models.

Results comparing alternative model formulations, with the omission and extension of various components, are shown in Table 3. The use of spatially correlated random effects ( $u$ ) greatly improves predictive ability and hence CV deviance increases if this is omitted, although as mentioned above, the unstructured effects ( $v$ ) take up the remaining variation

if this is omitted and this does not translate to a great difference in DIC. The converse was true with the omission of  $v$ : DIC is significantly worse, but there is actually a small improvement in CV deviance. With the omission of both, model fit is worse by both measures; although CV deviance is better than the  $v$  only model, with the covariate for drug offences having a stronger effect.

In terms of extending the model, DIC scores were worse with the addition of further complexity. CV deviance improved a little when allowing separate  $v$  terms for each year, but this may exaggerate spatial effects and over-fit to the sampled areas – the time-varying  $\sigma_v$  are far smaller and there may be identification problems. We therefore retained the basic form of the model; results for model parameters are shown in Table 4.

We compared model results using alternative priors for variance components, testing Gamma(0.5, 0.005) priors, as recommended by Kelsall & Wakefield [31],

Table 4. Final model for prevalence of antibodies to hepatitis C virus in injecting drug users, England, 2006–2008, posterior medians and 95% credible intervals

Parameter	Posterior median	95% CrI
<b>Odds ratios for model parameters</b>		
Age (yr)/sex group		
15–29, male	0.58	0.51–0.66
15–29, female	0.51	0.42–0.60
30–44, male	1.00 (baseline)	
30–44, female	1.03	0.86–1.23
45–59, male	1.59	1.25–2.02
45–59, female	1.99	1.26–3.19
Auxiliary variables		
Proportion aged <35 yr	1.30	1.09–1.57
Drug crime	1.08	0.89–1.31
Year effect		
2006	1.00 (baseline)	
2007	0.99	0.86–1.13
2008	0.95	0.82–1.10
Standard deviation of random-effects components		
Unstructured ( $\sigma_v$ )	0.283	0.024–0.549
Spatially structured ( $\sigma_u$ )	0.781	0.442–1.126

CrI, Credible interval.

and flat Gamma(0.1, 0.1) priors. Results were similar with the Gamma(0.1, 0.1) prior ( $\sigma_v=0.351$ ,  $\sigma_u=0.697$ ) but the large probability mass near zero for the Gamma(0.5, 0.005) prior reduced the more weakly identified unstructured random effects, with the spatial effects becoming more pronounced to compensate ( $\sigma_v=0.150$ ,  $\sigma_u=0.801$ ). Prevalence estimates were similar for all choices.

### HCV prevalence estimates

Anti-HCV prevalence estimates ranged from 14% to 82%, with a median of 43% and interquartile range from 32% to 54%. In contrast, observed prevalence ranged from 5% to 77%. Model estimates are more homogeneous due to shrinkage, and slightly higher as they account for the imperfect sensitivity of the test. A forest plot of anti-HCV prevalence estimates and credible intervals for each DAT in 2008 are displayed in Figure 1; and shaded maps of prevalence can be seen in Appendix 1 (available online). Prevalence is generally low in the North East and South West, with high prevalence DATs in and around central London, the North West, and a number of towns and cities

(Brighton, Bristol, Leeds, Nottingham, Reading, Portsmouth).

To examine the effects of model components, the spatially correlated effect  $u$ , unstructured random effect  $v$ , and effects of the auxiliary variables are plotted as shaded maps (see Appendix 1, online). The spatial effects exhibit two broad areas of increased prevalence (beyond that predicted by auxiliary covariates): in the North West, and in London and the areas to its south east; with a broad band of lower prevalence diagonally across the rest of the country. The auxiliary variable for ‘proportion <35’ has a stronger effect than drug offences, although patterns differ in some areas, with ‘proportion <35’ increasing odds of HCV in most large cities, but the variable for drug offences predominantly affecting London and the North West areas. (For full listings of DAT-level model effects see Appendix 2, available online.)

## DISCUSSION

### Main findings

We have shown that there is substantial local variation in anti-HCV prevalence, with areas of high prevalence concentrated in London, the North West, parts of the South East and some major cities. We found a relationship between the proportion of young adults in the adult population and anti-HCV prevalence, although other predictors produced similar patterns of results, which were generally insensitive to model choice. A spatial effect was well-identified in the model, indicating that areas tend to be influenced by their neighbours, beyond what may be predicted from the covariates and demographic variables included in the model.

### Strengths and limitations

This work aimed to provide the best possible local-level estimates of anti-HCV prevalence in recent IDUs in contact with specialist services for drug users across England. This goal was hampered by sparse data, which we sought to overcome by using auxiliary information that exhibits a link with anti-HCV prevalence, and by taking advantage of spatial patterns. Despite these efforts, uncertainty remains in prevalence estimates for non-sampled areas, particularly in areas with unusual covariate levels and contrasting or imprecise spatial effects. As long as this uncertainty at the DAT level is considered we believe

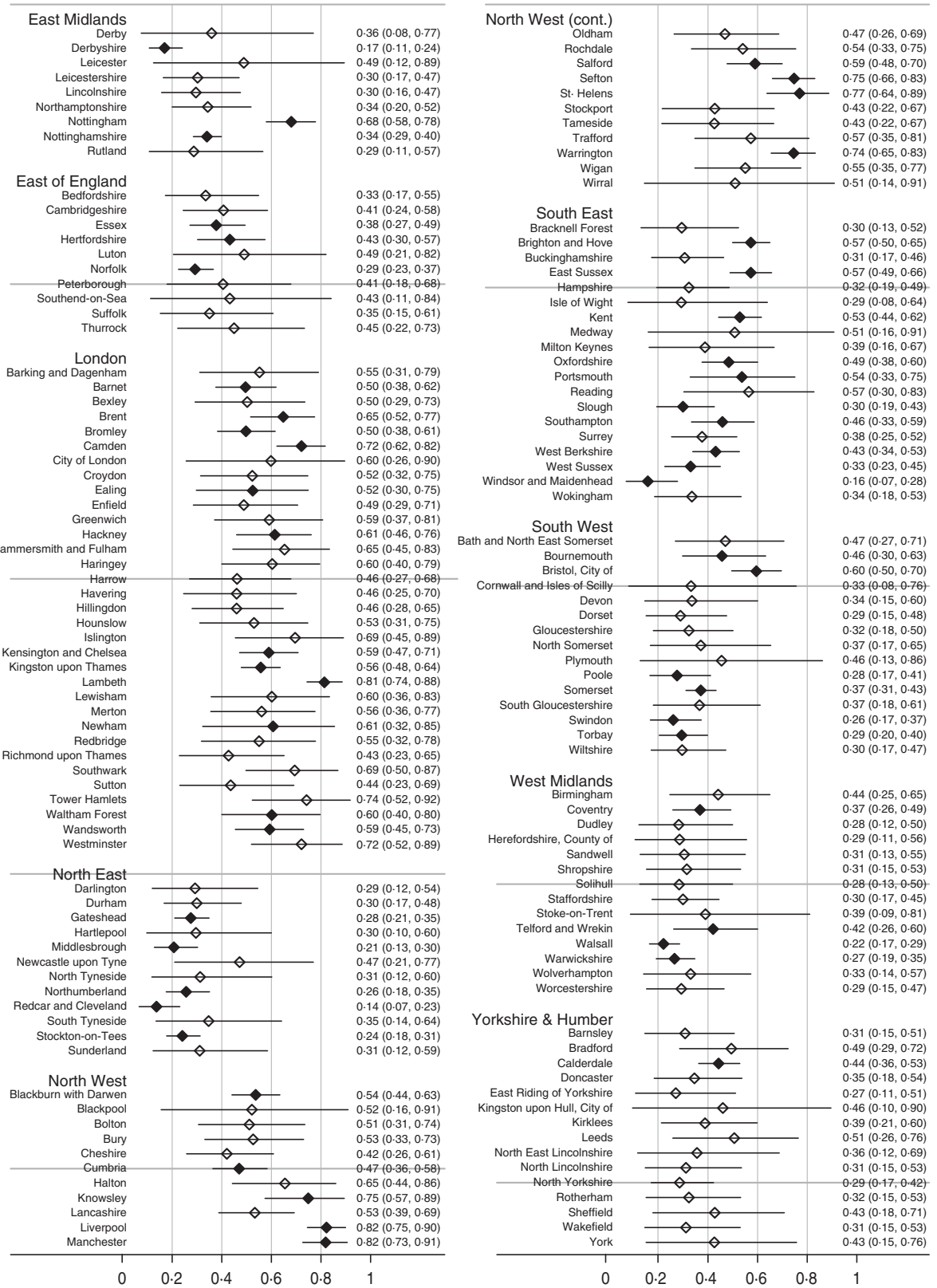


Fig. 1. Prevalence of antibodies to hepatitis C virus in injecting drug users in contact with specialist drug services, England, 2008. Posterior medians and credible intervals for each Drug Action Team (DAT) are displayed by region. Ever-sampled DATs (during 2006–2008) are shown with solid diamonds, non-sampled DATs with hollow diamonds.

that these estimates are a reliable indication of the general patterns in the anti-HCV prevalence in recent IDUs across England.

The DAT-level prevalence estimates are weighted according to the demographic distribution of all recent injectors undergoing treatment for their drug use, and are derived using data obtained from the sample of specialist services collaborating in the UAM survey. However, the UAM survey recruits from both harm reduction services (such as needle exchanges) and drug treatment services (such as substitute prescribing programmes and structured counselling) while only clients of the latter types of service are included in the NDTMS data. The NDTMS data describes those in treatment well; however, it is assumed here that it is generally representative of the characteristics of recent IDUs in contact with all specialist services. This assumption is supported by the fact that three-quarters of the recent IDUs participating in the UAM survey reporting needle-exchange use also reported currently receiving substitute drug treatment (data not shown), while others would be in structured counselling which is not enquired of in the survey. Of course, these DAT-level estimates still do not necessarily reflect prevalence in all recent IDUs as they do not include those not in contact with specialist services. However, the use of specialist services by IDUs in England is very extensive, with considerable overlap in injectors recruited from community and treatment and any differences in HCV prevalence largely due to differences in age or injecting duration [5].

As far as we are aware, the assessment of spatial models when there are non-sampled areas has not been investigated. We argue that the use of 'leave one out' cross-validation provides the best assessment of predictive ability in this setting. The DIC may not be informative for assessing out-of-sample prediction, as variation that is not explained by covariate effects may be taken up by either unstructured or spatially structured random effects, although models that make use of spatial correlation will provide better out-of-sample prediction. Posterior predictive model checks have been suggested as an alternative to full cross-validation [32], but these methods are principally concerned with model fit, and further investigation is required to assess their application when the goal is out-of-sample prediction.

Out estimates are based partly on auxiliary data, chosen primarily on the basis of model selection. However, these predictors also have plausible

interpretations: the proportion of the adult population aged < 35 years may be viewed as a measure of urbanicity; and drug crime may act as a proxy for severity of problem drug use. Taken together these factors are surmised to give an indication of which environments are more likely to experience high levels of HCV infection. Prevalence estimates for non-sampled areas may be sensitive to model choice, and we therefore assessed the robustness of prevalence estimates to the use of different auxiliary variables. Estimates were similar for most alternatives, although including only drug offences as a covariate decreased the prevalence estimates for 10 DATs by 5–8% and increased those for City of London and Isle of Wight by 18% and 17%, respectively. This is due to high rates of drug offences relative to the proportion of adults aged < 35 years in these areas.

### Findings in relation to the evidence base

The substantial geographical variation in anti-HCV prevalence observed here is consistent with previous studies, but allows a more formal assessment of this phenomenon via a modelling process that provides for the first time prevalence estimates for all local areas. For some DATs HCV prevalence has been estimated using respondent driven sampling (RDS) [33], which is postulated to be the most reliable way to obtain representative estimates for hidden populations [34]; however, only IDUs that have injected in the last 4 weeks were included. Results are available for Bristol [35], Birmingham and Leeds (M. Hickman, V. Hope, personal communication), showing a prevalence of 57% (95% CI 52–62), 38% (95% CI 33–44) and 58% (95% CI 53–64), respectively, compared to 60% (95% CrI 50–70), 44% (95% CrI 25–61) and 51% (95% CrI 26–76) in this study. Our results are uncertain for the non-sampled areas (Birmingham, Leeds), and the RDS method does not necessarily generate unbiased estimates [36] so no firm conclusions may be drawn, although the results appear to follow a broadly similar pattern for the areas covered. Previously, estimates have only been available at a regional level for England, and have not used the NDTMS data to relate findings to a wider population. These data thus complement regional and national data, as well as strengthening local evidence.

### Implications

The estimates of anti-HCV prevalence for non-sampled areas have substantial uncertainty and need



to be viewed with some caution; however, they will provide DATs with a more robust indication of the HCV prevalence in recent IDUs in contact with specialist services than relying on regional estimates of prevalence. This is especially so for those regions, like the East Midlands, with very marked variations in prevalence at the DAT level. These estimates therefore have public health utility as they permit more informed local commissioning of services to prevent, diagnose and treat HCV infection in this population group. As these estimates use routinely available data they can be repeated over time; doing this would allow the monitoring of local trends in anti-HCV prevalence in IDUs, and so permit service commissioning to adapt to these.

## CONCLUSIONS

The novel approach adopted here has produced local estimates of anti-HCV prevalence of use in informing local public health responses. The limitations to these estimates need to be noted and the methods employed need further validation and refinement. However, there is potential to apply these methods to other issues affecting IDUs, such as hepatitis B infection or even behavioural data, e.g. needle and syringe sharing or the uptake of diagnostic testing. This approach may also be of use in other situations where data are sparse and local-level estimates are needed to inform response; e.g. sexual health indicators in men who have sex with men and sex workers or general health issues in marginalized groups such as the homeless.

## NOTE

Supplementary material accompanies this paper on the Journal's website (<http://journals.cambridge.org/hyg>).

## ACKNOWLEDGMENTS

The authors thank Dr Stefano Conti and Dr Ayoub Saei for their comments and advice; Karl Molden for the provision of National Drug Treatment Monitoring System data; and all the collaborators involved with the unlinked anonymous prevalence monitoring programme, as well as the drug users who gave their time to be interviewed.

This work was supported by the United Kingdom Department of Health (grant number GHP 003/002/024).

## DECLARATION OF INTEREST

None.

## REFERENCES

1. **De Angelis D, et al.** An evidence synthesis approach to estimating Hepatitis C prevalence in England and Wales. *Statistical Methods in Medical Research* 2009; **18**: 361–379.
2. **EMCCDA.** Statistical Bulletin 2010, Table INF-2. Lisbon: EMCCDA, 2010.
3. **Sweeting MJ, et al.** Hepatitis C infection among injecting drug users in England and Wales (1992–2006): there and back again? *American Journal of Epidemiology* 2009; **170**: 352–360.
4. **Craine N, et al.** Incidence of hepatitis C in drug injectors: the role of homelessness, opiate substitution treatment, equipment sharing, and community size. *Epidemiology and Infection* 2009; **137**: 1255–1265.
5. **Hickman M, et al.** Hepatitis C virus (HCV) prevalence, and injecting risk behaviour in multiple sites in England in 2004. *Journal of Viral Hepatitis* 2007; **14**: 645–652.
6. **Judd A, et al.** Incidence of hepatitis C virus and HIV among new injecting drug users in London: prospective cohort study. *British Medical Journal* 2005; **330**: 24–25.
7. **Hope VD, et al.** Measuring the incidence, prevalence and genetic relatedness of hepatitis C infections among a community recruited sample of injecting drug users, using dried blood spots. *Journal of Viral Hepatitis*. Published online: 27 April 2010. doi: 10.1111/j.1365-2893.2010.01297.x.
8. **Martin N, et al.** Can antiviral therapy for hepatitis C reduce the prevalence of HCV among injecting drug user populations? A modelling analysis of its prevention utility. *Journal of Hepatology*. Published online: 22 October 2010. doi:10.1016/j.jhep.2010.08.029.
9. **Vickerman P, Hickman M, Judd A.** Modelling the impact on Hepatitis C transmission of reducing syringe sharing: London case study. *International Journal of Epidemiology* 2007; **36**: 396–405.
10. **Hickman M, et al.** Injecting drug use in Brighton, Liverpool, and London: best estimates of prevalence and coverage of public health indicators. *Journal of Epidemiology and Community Health* 2004; **58**: 766–771.
11. **Hope VD, et al.** Prevalence of hepatitis C among injection drug users in England and Wales: is harm reduction working? *American Journal of Public Health* 2001; **91**: 38–42.
12. **Health Protection Agency.** (<http://www.hpa.org.uk>). Accessed 13 January 2011.
13. **Health Protection Agency.** Shooting up: infections among injecting drug users in the United Kingdom 2008. London: Health Protection Agency, 2009.
14. **Ghangurde PD, Singh MP.** Synthetic estimation in periodic household surveys. *Survey Methodology, Statistics Canada* 1977; **3**: 152–181.

15. **Rhodes WR.** Synthetic estimation applied to the prevalence of drug use. *Journal of Drug Issues* 1993; **23**: 297–321.
16. **Bajekal M et al.** Synthetic estimation of healthy lifestyles indicators: Stage 1 report. London: National Centre for Social Research, 2004.
17. **Rao JNK.** *Small Area Estimation*. New York: Wiley, 2003.
18. **Best N, Richardson S, Thomson A.** A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research* 2005; **14**: 35–59.
19. **Judd A, et al.** Evaluation of a modified commercial assay in detecting antibody to hepatitis C virus in oral fluids and dried blood spots. *Journal of Medical Virology* 2003; **71**: 49–55.
20. **Communities and Local Government.** (<http://www.fti.communities.gov.uk/DataDownload.aspx>). Accessed 13 January 2011.
21. **Office for National Statistics.** Mid-year population estimates: quinary age groups and sex for local authorities in England and Wales; estimated resident population based on the 2001 Census. London: Office for National Statistics, 2008.
22. **National Drug Treatment Monitoring System.** (<http://www.ndtms.net/>). Accessed 13 January 2011.
23. **Besag J, York J, Mollie A.** A Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991; **43**: 1–59.
24. **Gelman A.** Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2006; **1**: 515–533.
25. **Spiegelhalter D, et al.** *WinBUGS Version 1.4 User Manual*. Cambridge: MRC Biostatistics Unit, 2003.
26. **Brooks SP and Gelman A.** General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998; **7**: 434–455.
27. **Spiegelhalter D, Best NG, Carlin BP.** Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 2002; **64**: 583–639.
28. **Van Houwelingen JC, le Cessie S.** Predictive value of statistical models. *Statistics in Medicine* 1990; **9**: 1303–1325.
29. **Stern HS, Cressie N.** Posterior predictive model checks for disease mapping models. *Statistics in Medicine* 2000; **19**: 2377–2397.
30. **May M, et al.** Development and validation of a prognostic model for survival time data: application to prognosis of HIV positive patients treated with anti-retroviral therapy. *Statistics in Medicine* 2004; **23**: 2375–2398.
31. **Kelsall J, Wakefield J.** Discussion of ‘Bayesian models for spatially correlated disease and exposure data’. In: Bernardo J, Berger J, Dawid A, eds. *Bayesian Statistics 6*. Oxford: Oxford University Press, 1999.
32. **Marshall EC, Spiegelhalter DJ.** Approximate cross-validatory predictive checks in disease mapping models. *Statistics in Medicine* 2003; **22**: 1649–1660.
33. **Heckathorn D.** Respondent-driven sampling: a new approach to the study of hidden populations. *Social Problems* 1997; **44**: 174–199.
34. **Malekinejad M, et al.** Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: a systematic review. *AIDS Behaviour* 2008; **12**: S105–S130.
35. **Hickman M, et al.** Assessing IDU prevalence and health consequences (HCV, overdose and drug-related mortality) in a primary care trust: implications for public health action. *Journal of Public Health (Oxford)* 2009; **31**: 374–382.
36. **Goel S, Salganik MJ.** Assessing respondent-driven sampling. *Proceedings of the National Academy of Sciences USA* 2010; **107**: 6743–6747.