Estimating Prevalence of Injecting Drug Users and Associated Death Rates in England Using Regional Data and Incorporating Prior Information

Ruth King

School of Mathematics and Statistics, University of St Andrews, St. Andrews, UK.

Sheila M. Bird

Medical Research Council, Cambridge and University of Strathclyde, Glasgow, UK.

Gordon Hay

Centre for Drug Misuse Research, University of Glasgow, Glasgow, UK.

Sharon J. Hutchinson

University of Strathclyde and Health Protection Scotland, Glasgow, UK.

Summary. Injecting drug users (IDUs) have a direct social and economical impact, yet can typically be regarded as a hidden population within a community. We estimate the size of the IDU population across the nine different Government Office Regions of England in 2005/6 using capture-recapture methods with age (ranging from 15-64) and gender as covariate information. We consider a Bayesian model-averaging approach using log-linear models, where we are able to include explicit prior information within the analysis in relation to the total population size (elicited from the number of drug-related deaths and injectors' drug-related death rates) and the male to female ratio of IDUs. Estimating the data at the regional level allows for regional heterogeneity and was aggregated to obtain an estimate at the England level with posterior mean of 194600 and 95% credible interval (180350, 208800), estimated to nearest 50. The results show significant regional variability in the estimated prevalence of current IDUs (with posterior means ranging from 3 to 9 per 1000 of population aged 15-64) and injecting drug-related death rates across the gender $\times$ age cross-classifications.

Keywords: Drug-related deaths; Log-linear models; Population size; Injecting drug users; Model-averaging; Prior information

1. Introduction

We focus on estimating the prevalence of current injecting drug users (IDUs) of opiates and/or crack cocaine in England, and at the Government Office region level. England's population of injectors rose epidemically in the (late) 1980s (de Angelis *et al*, 2004), several years later than Scotland's (Hutchinson *et al*, 2006). In addition, England's quality-assurance in methadone prescribing was somewhat later than Scotland's but was achieved prior to 2004, see Strang *et al* (2010). A major public health reason to engage injectors in methadone-substitution therapy is to reduce their dual risk of injecting-associated harms - blood-borne virus transmission and drug-related death (DRD). Methadone clients may continue to inject but, typically, their number of injections of illicit heroin reduces considerably

E-mail: ruth@mcs.st-and.ac.uk

(Hutchinson *et al*, 2000). In terms of DRDs, in the decade 2000-2009, 78% (3,125/4,012) of Scotland's DRDs were opiate-related (heroin and/or methadone), while the proportion was lower for England and Wales in the decade 1998-2007 at 63% (9,931/15,795). In England and Wales, as in Scotland, the proportion of DRDs that are opiate-related was higher for males than for females: Scotland - males 81%, females 67% (520/780); England and Wales - males 67%, females 46% (1,490/3,205) as reported by Bird *et al* (2010).

Not all opiate-related DRDs occur in injectors, but the majority does. Official statistics do not report whether the deceased had a history of injection drug use, let alone whether s/he was a current injector, and so we cannot know which opiate-related DRDs occurred in current injectors. Instead, as a reasonable approximation, we shall count, or attribute, all heroin-related DRDs (but no methadone-only DRDs) as having occurred in current injectors and report an 'injecting DRD rate'. Our use of this term denotes heroin-related DRDs per 100 current injectors. This definition differs from the approximation that King *et al* (2009) used for Scottish injectors' DRD rate, which was made before having access to more detailed toxicology on a decade of DRDs. Our analyses of Scottish injectors' DRD rates discovered that the DRD rate was markedly higher for young male than for young female injectors but that the female advantage was not sustained in older injectors (35-64 years). Greater Glasgow was one of the first areas in Scotland identified with high injecting prevalence, but there was later both diffusion of injecting to regions outwith Glasgow and earlier desistence in Glasgow which dissuaded younger recruits from injecting. A crude measure of males' desistence between regions is the ratio of male current injectors by age-group (15-34 to 35-64), with higher ratios the more worrying from a public health perspective.

We shall investigate the extent to which English Government Office Regions, notably London and the North West where injecting was established early, demonstrate similar trends as in Glasgow. We use capture-recapture methods to estimate the numbers of IDUs at the regional level. These are then combined to obtain an estimate at the England level, while still accounting for regional heterogeneity. English regions differ in size and so we also relate our regionally estimated number of current injectors to the region's mid-2005 population (aged 15-64 years).

Capture-recapture data for closed populations have a long history in application to both ecological (Otis *et al*, 1978) and epidemiological populations (Hook and Regal, 1995). Within epidemiological studies, capture-recapture data are used in a variety of situations including the estimation of hidden populations (Mastro *et al*, 1994; Frischer *et al*, 1993; Beynon *et al*, 2001; King *et al*, 2009) and disease prevalence (Hook *et al*, 1980; Madigan and York, 1997; Chao *et al*, 2001). Capture-recapture data involve a number of different data sources. Individuals are uniquely identifiable from each source which allows for the construction of a contingency table where each cell entry corresponds to the number of individuals observed by each distinct combination of sources. However, there is an unobservable cell corresponding to the number of individuals who belong to the population but were not observed by any source. Thus, failing to estimate this cell entry can potentially significantly underestimate the true population size, particularly with difficult to reach populations. To estimate the unobservable cell, a model is fitted to the observed data. We consider the commonly used log-linear models and apply a Bayesian approach that permits the use of a model-averaged estimate of the population size, accounting for both parameter and model uncertainty (Madigan and York, 1997; King and Brooks, 2001).

Additional covariate information can often be collected corresponding to individual characteristics, such as gender, location, age, marital status etc. Individuals with different characteristics may have different propensities to be observed by different combinations of

sources (King *et al*, 2009). Discrete covariates can be introduced as additional factors within the analysis to account for covariate heterogeneity. For example, King *et al* (2005) allowed for two-level covariate factors within their analysis. For the nine Government Office Regions of England, we adopt a similar approach considering two demographic characteristics, each with two levels: sex and age-group, (15-34 years, 35-64 years), by which DRDs are also cross-classified. Note that we do not include the region itself as a discrete covariate, but analyse the regional data independently of each other. This permits a direct comparison of important interactions identified for each region and avoids additional complex computational issues. Of particular interest is not only the estimates of IDUs within and across regions, but also injectors' DRD rates. We use expert prior information on the injecting DRD rate, combined with information on the regional number of heroin-related DRDs, to elicit an informative prior on total number of injectors. The DRDs are themselves provided across the different covariate levels, permitting the estimation of injecting DRD rates for the different joint covariate levels.

In Section 2 we describe the regional capture-recapture data and introduce the notation that we use throughout the paper before describing the Bayesian approach that we implement to analyse the data in Section 3. Section 4 provides the results obtained from the analysis, with particular focus on the number of injecting drug users and associated DRD rates. Capture propensities are also discussed. We conclude with a discussion in Section 5.

2. Regional Data

Data used within the capture-recapture analyses were collected nationally across England. These data can be disaggregated to the Drug Action Team (DAT) area level, a total of 149 within England. We consider the data collected in the financial year 2005-6. For each DAT area, the same four sources are used to identify IDUs uniquely from which we can construct a 2^4 contingency table with a single unknown cell. These sources collect data from (1) probation; (2) Drug Intervention Programme (DIP) prison assessments; (3) drug treatment; and (4) DIP community assessments. DIPs are a crime reduction initiative which works across different organisations, including criminal justice bodies such as police, prison and probation, and drug treatment services. Assessments which record an individual's current drug using and drug injecting status are carried out at various points in their journey through the criminal justice system and into treatment. In England, there has been major investment in DIPs, both in prisons and in the community, with the aim of engaging in assessment and drug treatment those caught up in the criminal justice system who test positive for opiates or cocaine. Regions where connections across services are made successfully would be revealed by the same clients tending to feature on more than one data-source and perhaps by lower injecting DRD rates if current injectors are successfully engaged in opiate substitution therapy, which is mainly by methadone in the UK.

Notationally, we label the sources $S1, S2, S3$ and $S4$, using the same order as above. We label each cell in the 2^4 contingency table by $\boldsymbol{k} \in \{0,1\}^4$ which represents the combination of sources that an individual is observed by. For example, cell $\boldsymbol{k} = \{0,1,0,0\}$ corresponds to being observed by only source 2 (DIP prison assessments). Covariate data corresponding to gender and age on each individual observed are also recorded in the data collection process. We cross-classify the observed individuals into gender and age-group (15-34 and 35-64), allowing us to construct four 2^4 regional contingency tables, which can be written as a 2^6 contingency table with each cell corresponding to the number of individuals that

Table 1. Number of unique injecting drug users observed in each region and each cross-classification of gender and age.

	Region	Total	Male 15-34	Female 15-34	Male 35-64	Female 35-64
1	East of England	3408	1574	605	962	267
2	East Midlands	5717	3365	963	1117	272
3	London	8198	2687	1062	3492	957
4	North East	5585	2944	858	1643	140
5	North West	11309	4678	1756	3904	971
6	South East	5444	2065	940	1498	401
7	South West	8767	4091	1580	2405	691
8	West Midlands	6627	3886	1081	1332	328
9	Yorkshire and Humber	11231	6412	2221	2089	508
	England	66286	32243	11066	18442	4535

are observed by each combination of four sources for each gender × age-group classification. These contingency table data at the DAT area level were aggregated to the 9 Government Office regions for transmission to, and analysis by, us. For each of the 9 regional contingency tables, there are 4 unknown cell entries, corresponding to the number of individuals not observed by any of the sources for each gender × age-group classification.

For a given region, we let $\boldsymbol{n}_{obs}$ and $\boldsymbol{n}_{unobs}$ denote the set of observed and unobserved cell entries, respectively, and $\boldsymbol{n} = \{\boldsymbol{n}_{obs}, \boldsymbol{n}_{unobs}\}$. Further, for each individual region, we let $n_{(i,j)}$ denote the *observed* number of individuals of gender i in age-group j; and $n_{(i,j):\boldsymbol{k}}$ the number of individuals of gender i in age-group j that belong to cell $\boldsymbol{k} \in \{0,1\}^4$ for $i \in \{M, F\}$ (M = male; F = female) and $j \in \{15-34, 35-64\}$. Thus, $n_{(i,j):\mathbf{0}} = n_{(i,j):\{0,0,0,0\}}$ denotes the number of individuals of gender i in age-group j that are not observed (i.e. the missing cell for the given cross-classification). We let $N_{(i,j)}$ denote the total number of individuals of gender i in age-group j for $i \in \{M, F\}$ and $j \in \{$15-34, 35-64$\}$; and $\boldsymbol{N} = \{N_{(i,j)} : i \in \{\mathrm{M}, \mathrm{F}\}; j \in \{15-34, 35\text{-}64\}\}$, so that,

$$N_{(i,j)} = n_{(i,j)} + n_{(i,j):\mathbf{0}} = \sum_{\boldsymbol{k} \in \{0,1\}^4} n_{(i,j):\boldsymbol{k}}.$$

We let $N_{tot} = \sum_{ij} N_{(i,j)}$ denote the total number of IDUs in the given region. Finally, for notational purposes, we let $N_{(i)} = \sum_j N_{(i,j)}$ (corresponding to the total number of individuals of gender $i \in \{M, F\}$) and $N_{(j)} = \sum_i N_{(i,j)}$ (corresponding to the total number of individuals in age-group $j \in \{15-34, 35-64\}$). To provide a brief summary of the data, we present the observed number of unique individuals identified in each region in Table 1 along with the corresponding number observed for each combination of gender and age (i.e. $\sum_{ij} n_{(i,j)}$ and $n_{(i,j)}$ for $i \in \{M, F\}$ and $j \in \{$15-34, 35-64$\}$) for each region.

3. Analysis

The observed contingency table for each region is analysed independently of all other regions. We consider the set of log-linear models initially introduced by Fienberg (1972), where the log of the cell probabilities, $\boldsymbol{p}$, are a linear sum of main effects and interaction terms between the sources and/or covariates (and normalised so that the sum of the cell probabilities equals unity). We restrict the set of possible interactions to that of two-way interactions

corresponding to source × source (6 in total), source × covariate (8 in total) and covariate × covariate (only 1) interactions.

We let the set of all possible log-linear parameters be denoted by $\boldsymbol{\theta}$. The corresponding set of cell entries, $\boldsymbol{n}$, has a Multinomial distribution with parameters, N_{tot} and $\boldsymbol{p}$. For further details see for example King *et al* (2005, 2009) who consider similar models in relation to the number of IDUs for Scotland, there treating region as an additional two-level factor.

3.1. Bayesian approach

We consider a Bayesian approach and analyse the data from each region independently of all other regions so that, without loss of generality, we condition on any given region. For a given log-linear model, m, (in terms of the log-linear parameters present in the model) we let the corresponding set of log-linear parameters be denoted by $\boldsymbol{\theta}_m$. We then form the joint posterior distribution over the set of log-linear parameters and total number of individuals in each gender × age-group cross-classification,

$$\begin{aligned}\pi(\boldsymbol{N},\boldsymbol{\theta}_m|\boldsymbol{n}_{obs}) &\propto f(\boldsymbol{n}_{obs}|\boldsymbol{N},\boldsymbol{\theta}_m)p(\boldsymbol{N},\boldsymbol{\theta}_m)\\ &\propto N_{tot}!\prod_{i\in\{M,F\}}\prod_{j\in\{15-34,35-64\}}\frac{1}{(n_{(i,j):\mathbf{0}}!)}\prod_{\boldsymbol{k}\in\{0,1\}^4}p_{ij:\boldsymbol{k}}^{n_{ij:\boldsymbol{k}}}p(\boldsymbol{N},\boldsymbol{\theta}_m)\end{aligned}$$

where $p_{ij:\boldsymbol{k}}$ denotes the probability that an individual of gender i in age-group j is observed in contingency table cell $\boldsymbol{k}$ and is a deterministic function of the $\boldsymbol{\theta}_m$ parameters. The first terms in the posterior distribution correspond to the Multinomial likelihood component and $p(\boldsymbol{N},\boldsymbol{\theta}_m) = p(\boldsymbol{N})p(\boldsymbol{\theta}_m)$ the prior on the total population counts for each gender × age-group cross-classification and log-linear parameters that are assumed to be independent of each other.

We do not specify the log-linear model *a priori*, in terms of the log-linear interaction terms that are present in the model, but consider a model discrimination approach. Within the Bayesian framework we follow the approaches of Madigan and York (1997) and King and Brooks (2001) and extend the posterior distribution to include the model space. In other words, we treat the model itself to be a discrete parameter, given the observed data, and form the joint posterior distribution over both the model and parameter space. The (marginal) posterior model probability for model m, given the data, can be expressed in the form,

$$\pi(m|\boldsymbol{n}_{obs}) \propto \int_{\boldsymbol{\theta}_m}\sum_{\boldsymbol{N}}\pi(\boldsymbol{N},\boldsymbol{\theta}_m|\boldsymbol{n}_{obs})d\boldsymbol{\theta}_m,$$

where the denominator once again ensures that the sum of the posterior distribution over admissible models sums to unity. In addition, we are also able to calculate the posterior (model-averaged) distribution of the population sizes, accounting for both parameter and model uncertainty. For example, the posterior model-averaged distribution for the number of IDUs for each gender × age-group cross-classification (and hence the total population size) is given by,

$$\pi(\boldsymbol{N}|\boldsymbol{n}_{obs}) = \sum_m \pi(\boldsymbol{N}|\boldsymbol{n}_{obs},m)\pi(m|\boldsymbol{n}_{obs}),$$

where $\pi(\boldsymbol{N}|\boldsymbol{n}_{obs},m)$ denotes the marginal posterior distribution for $\boldsymbol{N}$ under model m.

Table 2. Estimated average number of injecting drug related deaths (DRDs) per year for each region using data from the four calender years 2004-2007.

Region	Total	Male 15-34	Female 15-34	Male 35-64	Female 35-64
1	60.8	25.4	4.7	26.3	4.4
2	50.8	23.9	3.7	19.6	3.6
3	104.4	41.4	6.2	49.2	7.6
4	115.3	48.4	8.7	48.4	9.8
5	45.7	27.1	3.0	12.8	2.8
6	114.3	48.0	8.3	48.7	9.3
7	78.3	34.9	4.7	32.6	6.1
8	71.7	33.1	4.4	29.5	4.7
9	95.0	54.5	6.8	29.0	4.7
England	736.1	336.7	50.4	296.0	53.0

3.2. Prior expert information

There is external information available that can be combined with expert prior beliefs to provide an informative prior in relation to the total number of IDUs for each gender $\times$ age-group cross-classification. These relate to the total number of IDUs and the male to female ratio of IDUs. In particular, we have independent data relating to the number of heroin-related DRDs for each region between 2004-2007 and prior beliefs relating to the annual DRD rate for injectors. The totality of DRDs includes those with any combination of heroin/morphine, methadone, cocaine, benzodiazapines and alcohol in their systems at time of death. We assume that current IDUs are only those with any heroin/morphine in their system (irrespective of any other drugs identified). The corresponding estimated numbers of injecting DRDs are provided in Table 2 for each region (including at the gender $\times$ age-group cross-classification) and aggregated to the England level. To form the prior on the total population size, we couple this information with the prior beliefs relating to the annual injecting DRD rate. We specify a symmetric 90% interval for IDUs' annual injecting DRD rate of (0.3%, 1.2%) with median of 0.6% (this prior was informed by the analysis of Merrall *et al* (2010) of drug-related death rate for drug treatment clients in Scotland from 1996-2006 and for Scotland's injectors as analysed by King *et al* (2009)). Finally, following King *et al* (2005), we specify a symmetric 80% prior interval for the male to female ratio of (3:2, 9:1).

3.3. Prior distributions

We initially specify priors on the log-linear parameters where we do not have any prior information, before we consider the parameters on which there are some expert prior beliefs, relating to the population size and male to female ratio. We complete the prior specification with the prior model probabilities in terms of the interactions present in each model. For each individual region and each possible log-linear model we follow King and Brooks (2001) and specify a hierarchical $N(\mathbf{0}, \sigma^2 I)$ distribution on the set of log-linear parameters present in the model and use the noninformative prior $\sigma^2 \sim \Gamma^{-1}(0.001, 0.001)$.

We now consider the informative prior beliefs. To represent the expert prior information we specify a log-Normal prior on the total population satisfying the prior beliefs for the total population size (independently over models). For example, suppose that, for a given region, the estimated annual number of DRDs of injectors is x. We specify a prior on the log of the

total number of IDUs for the region to be normally distributed with mean $\log(x/0.6\%)$ (so that the prior median is accurately reflected) and variance 0.1776 (to reflect the specified prior 90% interval). Similarly, we set the log of $R = N_{(M)}/N_{(F)}$ to be normally distributed with mean $\log(3.6742)$ and variance 0.489. This provides a prior median of 3 for the male to female ratio with 80% symmetric interval of (1.5, 9). Finally, without any remaining prior beliefs on the age distribution, we specify Uniform[0,1] priors on the proportion of young, conditional on being male or female, which we denote by P_1 and P_2, so that,

$$P_1 = \frac{N_{(M,15-34)}}{N_{(M)}} \qquad \text{and} \qquad P_2 = \frac{N_{(F,15-34)}}{N_{(F)}}.$$

Finally, we specify a prior over the model space. We define the set of possible models to be those models with a maximum of second-order interaction terms (i.e. two-way factors). This significantly reduces the number of possible hierarchical log-linear models and aims to focus on the most important direct interactions between the different sources and/or covariates. Without any strong prior information relating to the two-way interactions that may be present we specify a prior probability of 0.5 that each interaction is present in the model. This induces an equal prior probability for each possible model in the set of plausible models.

3.4. (Reversible jump) Markov chain Mote Carlo algorithm

The posterior distribution is defined over both parameter and model space, so that we implement a reversible jump (RJ) Markov chain Monte Carlo (MCMC) algorithm (Green, 1995) to explore the posterior distribution, since the posterior distribution is multi-dimensional (the number of parameters differs between different models). Within the algorithm, we use a two-step procedure:

Step 1 Conditional on the model, we cycle through each individual parameter in turn and propose to update the parameter using a Gibbs or Metropolis-Hastings (MH) step;
Step 2 Update the model using a reversible jump step by adding or removing a log-linear interaction term from the model.

We consider each step in turn.

3.4.1. Step 1: Updating the parameters

We update σ^2 using a Gibbs step, since the posterior conditional distribution is of standard form (i.e. inverse Gamma) and a single-update random walk MH algorithm for all the other log-linear parameters and population sizes for each gender $\times$ age-group cross-classification. See Brooks (1998) for a general description of these algorithms and King and Brooks (2001) for the specific application to the log-linear parameters. However, we consider the updating of the total number of individuals in each gender $\times$ age-group cross-classification, $\boldsymbol{N}$, (or equivalently the missing cell entries) in further detail due to the complexity associated with the informative prior specified.

The prior relating to the population size is specified in the form $p(N_{tot}, R, P_1, P_2) = p(N_{tot})p(R)p(P_1)p(P_2)$. However, the model is defined in terms of the total number of individuals in each gender $\times$ age-group classification (or equivalently each unobserved cell entry) and it is these parameters, $\boldsymbol{N} = \{N_{(M,15-34)}, N_{(F,15-34)}, N_{(M,35-64)}, N_{(F,35-64)}\}$, that are

updated using a single-update algorithm within the MCMC algorithm. To demonstrate the issue that arises in this case, without loss of generality, consider the MH random walk to update the parameter $N_{(M,15-34)}$. We propose the candidate value,

$$N'_{(M,15-34)} = N_{(M,15-34)} + \epsilon,$$

where $\epsilon \sim U[-d, d]$ for some prespecified value of d, chosen via pilot tuning. We let $\boldsymbol{N}' = \{N'_{(M,15-34)}, N_{(F,15-34)}, N_{(M,35-64)}, N_{(F,35-64)})$. The acceptance probability is given by $\min(1, A)$, where A reduces to,

$$\begin{aligned} A &= \frac{f(\boldsymbol{n}_{obs}|\boldsymbol{N}', \boldsymbol{\theta}_m)p(\boldsymbol{N}')}{f(\boldsymbol{n}_{obs}|\boldsymbol{N}, \boldsymbol{\theta}_m)p(\boldsymbol{N})} \\ &= \frac{N'_{tot}!N_{(M,15-34)}!}{N_{tot}!N'_{(M,15-34)}!} p^{\epsilon}_{(M,15-34):\mathbf{0}} \frac{p(\boldsymbol{N}')}{p(\boldsymbol{N})}, \end{aligned}$$

where $N'_{tot} = N'_{(M,15-34)} + N_{(F,15-34)} + N_{(M,35-64)} + N_{(F,35-64)}$ and $p_{(M,15-34):\mathbf{0}}$ denotes the current value of the probability an individual is male, in age-group 15-34 and unobserved. The acceptance probability is a function of the joint prior distribution $p(\boldsymbol{N})$. However, the prior on the population size is provided in the form $p(N_{tot}, R, P_1, P_2)$. We calculate the corresponding prior distribution on $p(\boldsymbol{N}) = p(N_{(M,15-34)}, N_{(F,15-34)}, N_{(M,35-64)}, N_{(F,35-64)})$ using a transformation of variable argument. In particular, we have that,

$$p(\boldsymbol{N}) = p(N_{tot}, R, P_1, P_2) \frac{N_{tot}}{N_{(M)} N^3_{(F)}},$$

where the final term corresponds to the Jacobian (see appendix for derivation). We note that an alternative approach, in principle, would be to propose to update the parameters N_{tot}, R, P_1 and P_2 within each iteration of the Markov chain, since the priors are specified directly on these parameters. However, the proposal distribution of the corresponding MH algorithm would be significantly more complex, since this involves simultaneously updating multiple cell entries within a single step. For example, updating P_1 will involve simultaneously updating $N_{(M,15-34)}$ and $N_{(M,35-64)}$; P_2 updating $N_{(F,15-34)}$ and $N_{(F,35-64)}$; and R and N updating all missing entries simultaneously, so that the corresponding proposal distribution is considerably more complex.

3.4.2. Step 2: Updating the model

To update the log-linear interaction terms present within the model we use a reversible jump step (Green, 1995). For a single RJ step, we propose to add or remove a single two-way interaction term (since we only consider models with two-way interactions). We choose each log-linear interaction with equal probability. If the parameter is present in the model, we propose to remove the parameter; if it is not in the model, we propose to add the parameter. Suppose that we propose to add a given two-way interaction parameter. We propose a candidate value from a proposal distribution, q, which in this case is a Normal distribution. The corresponding proposal mean is obtained using the posterior mean of the given parameter from a pilot MCMC run in the model containing all two-way interactions. The proposal variance is chosen via pilot-tuning. See King and Brooks (2001) for further details using an analogous approach. The corresponding acceptance probability simply

reduces to the ratio of the likelihood function using the proposed and current parameter values respectively, multiplied by the ratio of the prior density function to proposal density function for the newly proposed log-linear parameter (the Jacobian is equal to unity).

For each region, the RJMCMC algorithm is run for a total of 2 million iterations with the first 10% discarded as burn-in. Independent replications using over-dispersed starting points obtained similar results (all with the same interpretation) so that we conclude that the algorithm has sufficiently converged. Additionally, using the Brooks-Gelman-Rubin statistic on the missing cell entries provided no evidence for lack of convergence.

4. Results

4.1. Estimating the Number of Injecting Drug Users

Figure 1 provides plots of the prior and (model-averaged) marginal posterior distributions for the number of IDUs in each region. For regions 2, 3, 5, 7, 8 and 9, the priors generally appear to underestimate the number of IDUs in the different regions. The most significant difference between the prior and posterior distributions is clearly for region 5 (North West) with virtually no overlap between the prior and posterior distributions. This would potentially suggest, for these regions, and particularly region 5, that (i) the estimate of the number of injecting DRDs is an underestimate and/or (ii) the injecting DRD rate is lower than the prior expert beliefs. We return to this issue below when discussing the posterior injecting DRD rates. We note that specifying alternative (uninformative priors) on the total population size (namely a Uniform prior on each missing cell) resulted in very similar posterior distributions for the total population size, providing evidence that the posterior for $\boldsymbol{N}$ is data-driven.

Table 3 provides the posterior estimates for the total population size and each combination of gender $\times$ age-group cross-classifications for each of the regions, in addition to the corresponding population sizes for England (i.e. posterior estimates summed over each region). The posterior mean of the total current injector population for England can be easily calculated as the sum of the posterior means of the estimates for each region. However, the corresponding credible intervals (CIs) at the England level cannot be obtained directly from the credible intervals for each individual region. For example, summing the 2.5% quantiles (used for the lower bound of the 95% credible interval) over all regions will not give the corresponding 2.5% quantile for England (the value obtained would be for a much lower quantile for the total population size for England). We obtain the 95% credible interval at the national level by considering a Monte Carlo approach. Recall that the regional datasets are analysed independently of each other, so that the posterior (marginal) distributions of the population sizes are independent across regions. To obtain a sample observation from the posterior distribution of the population size for England, we simply take a sample observation of the number of IDUs from each region and sum these values. This Monte Carlo approach is also used to obtain the gender $\times$ age-group population size estimates and the injecting DRD rate.

From Table 3 we see that three regions (3, 5 and 9) appear to have significantly higher absolute number of IDUs than the other regions. These regions correspond to London, North West and Yorkshire and Humber. In addition, there is consistently a larger estimated number of males than females in each region for each age-group considered. Overall, the posterior mean ratio of males to females (aggregated to the England level) is 3.32 with

Fig. 1. The posterior distribution for the total population size for each region (in black) and the corresponding prior distribution (in grey).

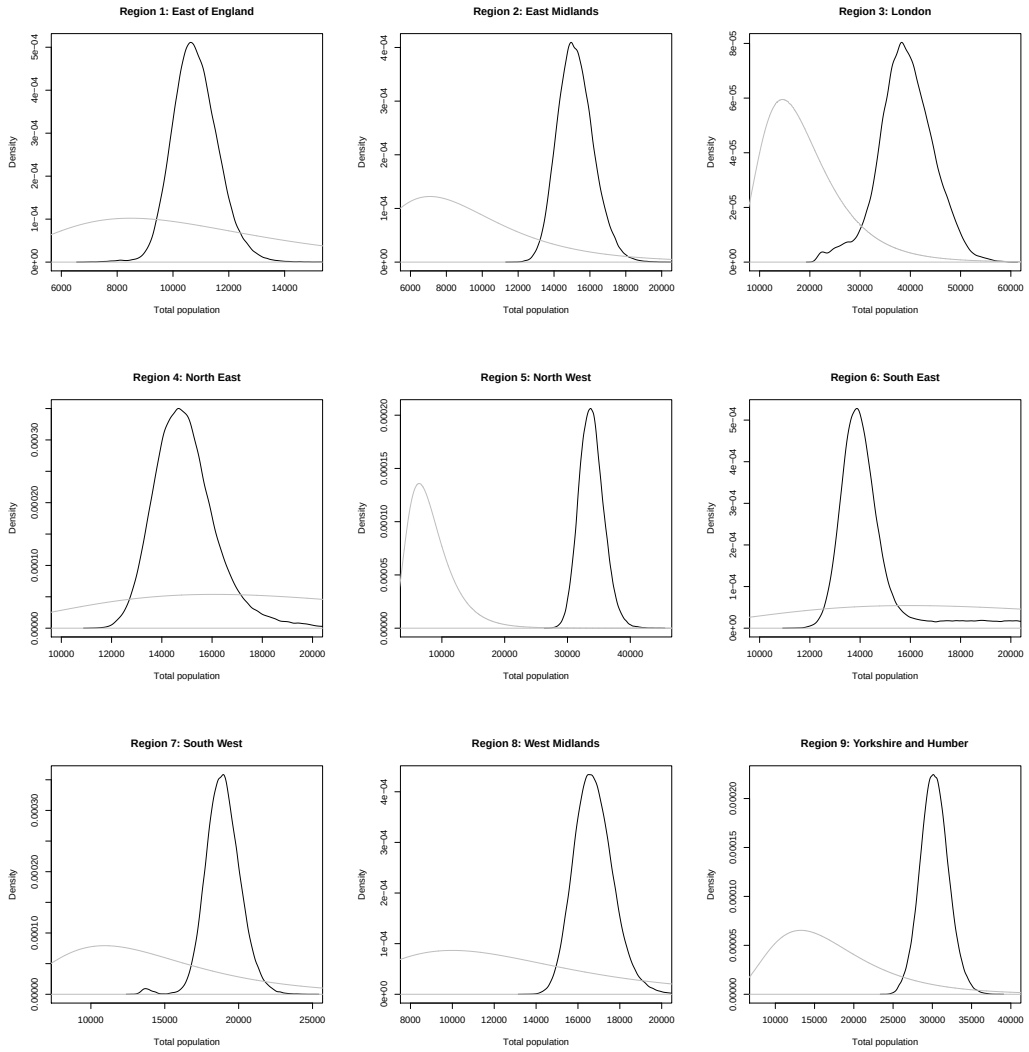

Table 3. Posterior mean and 95% symmetric credible interval (in brackets) for the total number of IDUs in each region and each cross-classification of gender and age and aggregated to the England level using a Monte Carlo approach (rounded to nearest 10).

Region	Total	Male 15-34	Female 15-34	Male 35-64	Female 35-64
1	10810	5040	1640	3370	760
	(9350, 12530)	(4310, 5900)	(1370, 1990)	(2780, 4000)	(620, 950)
2	15220	8890	2220	3400	700
	(13440, 17340)	(7830, 10150)	(1890, 2730)	(2930, 3960)	(570, 880)
3	39130	14400	4530	16710	3490
	(26870, 49850)	(9660, 18710)	(3080, 66200)	(11460, 21480)	(2440, 4790)
4	14970	7410	2250	4890	420
	(12880, 18080)	(6390, 8970)	(1900, 2750)	(3770, 6130)	(310, 550)
5	33830	12910	5410	12240	3280
	(30260, 37970)	(11570, 14450)	(4800, 6120)	(10900, 13800)	(2870, 3740)
6	14690	6690	2530	4290	1180
	(12750, 21980)	(5810, 9900)	(2170, 3810)	(3660, 6500)	(980, 1810)
7	18930	8520	3470	5360	1570
	(16730, 21350)	(7540, 9610)	(2880, 3980)	(4700, 6120)	(1290, 1830)
8	16760	9390	2850	3540	980
	(15100, 18780)	(8460, 10510)	(2500, 3250)	(3130, 4020)	(820, 1150)
9	30280	16530	6250	5960	1550
	(26860, 33960)	(14690, 18500)	(5490, 7070)	(5230, 6770)	(1330, 1790)
England	194610	89760	31150	59760	13940
	(180340, 208810)	(83840, 95890)	(29050, 33510)	(53980, 65380)	(12660, 15430)

corresponding 95% symmetric CI (3.12, 3.46). The posterior mean male to female ratio over the different regions ranges from 2.76 (South West) to 4.61 (North East). These values all lie around the prior median for the male to female ratio.

For comparison with the estimate of the number of IDUs in England in Table 3 by aggregating the posterior regional estimates, we perform a further analysis where we aggregate the raw data across the Government Office regions and analyse the resulting contingency table using the same Bayesian approach. This aggregation of the raw data over the regions removes a potential source of heterogeneity across the different regions. To analyse these data, we use the same prior beliefs as before, which provides a prior median for the total population size of 121,848 with 90% interval (60924, 243696). This lower bound is actually less than the number of observed IDUs (see Table 1). The corresponding posterior mean (rounded to nearest 10) of the total population size is 209,120 with 95% symmetric CI of (197570, 221470). Thus the regionally-derived England estimate (i.e. obtained by aggregating the posterior regional estimates) is generally lower, although there is some overlap between the credible intervals, than that obtained when analysing the data without heed to the regional component (i.e. aggregating at the data level). If we consider the corresponding estimates for the cross-classifications when aggregating at the data level we obtain posterior means and 95% symmetric CIs (rounded to nearest 10) for males 15-34 of 96110 (90770, 101780); females 15-34 of 36760 (34220, 39210); males 35-64 of 59700 (56130, 63590); and females 35-64 of 16540 (15320, 17760). The posterior estimate for males 35-64 is fairly consistent with the regionally-derived England estimate, but with higher estimates for the other gender $\times$ age-group cross-classifications. In other words, allowing for heterogeneity at the regional level results in lower estimates of female and younger male IDUs.

A previous estimate obtained for England by Hay *et al* (2009) is significantly smaller, corresponding to a point estimate of 129,977 with 95% confidence interval (125786, 137034). They also provide estimates for each of the same 9 Government Office regions, these are again typically lower (except for region 6 - South East). These previous estimates were obtained by considering a 2^4 incomplete contingency table (ignoring gender $\times$ age-group cross-classifications) for each individual DAT area and aggregating the estimates (using a bootstrap approach to obtain the confidence interval) to either the regional or England level. We note that Hay *et al* (2009) considered a reduced set of log-linear models, corresponding to those with a maximum of two two-way interactions present (a total of 22 distinct models). Typically the model with lowest AIC value was chosen (although see Hay *et al* (2009) for more specific details) and the corresponding estimate for total population was as given by the chosen log-linear model. By contrast, within our approach, we include the model uncertainty within our estimates (often leading to wider uncertainty intervals to reflect the additional model uncertainty). We return to the underlying reasons for this apparent discrepancy in population estimates between these different approaches in Section 4.3 when we discuss in further detail the interactions identified for each of the Government Office regions.

Finally, Table 4 relates the centrally estimated number of current injectors to regions' mid-2005 population aged 15-64 since the regions differ in population size. England has an estimated 5.8 current injectors per 1,000 of the population aged 15-64 (with 95% symmetric CI 5.4-6.3). The estimated injector prevalence is low (posterior mean around 3) in East England (region 1) and the South East (region 6), high (posterior mean around 7.5) in London (region 3) and the North West (region 5) and very high (posterior mean around 9) for the North East (region 4) and Yorkshire and Humberside (region 9). However, it is an encouraging sign for London and the North West (with high prevalence rates) that their injector prevalences by age-group (15-34 to 35-64) are relatively low compared to England as a whole (posterior mean of 1.61 for males; 2.23 for females; and see Millar *et al* (2006) for further detailed discussion of problem drug use in the North West up to 2001). Regions with high injector ratios by age-group may be a signal of later diffusion with younger injectors predominating. These regions include East and West Midlands (regions 2 and 8), and Yorkshire and Humberside (region 9), the last of which is also beset by the largest overall injector prevalence per 1,000 of the population aged 15-64.

4.2. Injecting Drug-related Death Rates

We obtain a sample from the posterior distribution for the injecting DRD rates by taking the ratio of the estimated number of DRDs (as provided in Table 2) with the total number of IDUs for each gender $\times$ age-group cross-classification at each iteration of the Markov chain. The corresponding posterior mean and symmetric 95% CI of the injecting DRD rates are provided in Table 5. Recall that the prior 80% interval on the injecting DRD rates is (0.3%, 1.2%). We comment first at the England level where the posterior injecting DRD rate is at the lower end of the prior distribution informed by the Scottish analyses. We note that the overall posterior estimate for the injecting DRD rate is lower than that presented by Bloor *et al* (2008) who were investigating the "Scottish effect" of higher DRD rates in Scotland compared to England, obtaining an estimate for Scotland of 0.8% (with 95% uncertainty interval 0.5%-1.2% using data from 2001-5). In addition, the DRD rate in England appears to be significantly lower for younger than older injectors: for males, posterior means of 0.38% for the younger age-group compared to 0.5% for the older age

Table 4. Current injector totals set in context by regions' mid-2005 population aged 15-64 and estimated ratio of young to old (i.e. 15-34 to 35-64) for each gender in each region with symmetric 95% credible intervals.

Region	mid-2005 populations aged 15-64 (in 1000s)	Posterior mean of current injectors (per 1,000) population aged 15-34 (95% CI)	Posterior mean current injectors to nearest 50 (95% CI)	Posterior mean of male injector ratio by age-group (15-34/35-64) (95% CI)	Posterior mean of female injector ratio by age-group (15-34/35-64) (95% CI)
1	3604.0	3.0 (2.6, 3.5)	10800 (9350, 12550)	1.50 (1.36, 1.82)	2.15 (1.88, 2.58)
2	2839.0	5.4 (4.7, 6.1)	152000 (13450, 17350)	2.62 (2.43, 2.85)	3.17 (2.75, 3.62)
3	5269.0	7.4 (5.1, 9.5)	39150 (26850, 49850)	0.86 (0.76, 0.97)	1.30 (1.13, 1.48)
4	1686.1	8.9 (7.6, 10.7)	14950 (12900, 18100)	1.53 (1.36, 1.97)	5.41 (4.41, 7.17)
5	4497.0	7.5 (6.7, 8.4)	33850 (30250, 37950)	1.05 (1.01, 1.10)	1.65 (1.54, 1.77)
6	5338.0	2.8 (2.4, 4.1)	14700 (12750, 22000)	1.56 (1.47, 1.67)	2.15 (1.94, 2.39)
7	3252.7	5.8 (5.1, 6.6)	18950 (16750, 21350)	1.59 (1.48, 1.66)	2.20 (2.02, 2.36)
8	3499.9	4.8 (4.3, 5.4)	16750 (16750, 18800)	2.65 (2.49, 2.86)	2.93 (2.61, 3.36)
9	3325.7	9.1 (8.1, 10.2)	30300 (26850, 33950)	2.78 (2.59, 2.92)	4.05 (3.68, 4.40)
England	33311.4	5.8 (5.4, 6.3)	194600 (180350, 208800)	1.61 (1.56, 1.64)	2.23 (2.14, 2.30)

Table 5. Posterior mean and 95% symmetric credible interval (in brackets) for the drug-related death rates for IDUs, (in %), in each region and each cross-classification of gender and age.

Region	Total	Male 15-34	Female 15-34	Male 35-64	Female 35-64
1	0.57	0.51	0.29	0.79	0.58
	(0.48, 0.65)	(0.43, 0.59)	(0.24, 0.34)	(0.66, 0.94)	(0.46, 0.71)
2	0.34	0.27	0.17	0.58	0.52
	(0.29, 0.38)	(0.24, 0.31)	(0.13, 0.19)	(0.49, 0.67)	(0.41, 0.63)
3	0.27	0.29	0.14	0.30	0.22
	(0.21, 0.39)	(0.22, 0.43)	(0.10, 0.20)	(0.23, 0.43)	(0.16, 0.31)
4	0.78	0.66	0.39	1.00	2.37
	(0.64, 0.90)	(0.54, 0.76)	(0.32, 0.46)	(0.79, 1.28)	(1.79, 3.14)
5	0.14	0.21	0.06	0.10	0.09
	(0.12, 0.15)	(0.19, 0.23)	(0.05, 0.06)	(0.09, 0.12)	(0.08, 0.10)
6	0.79	0.73	0.33	1.16	0.80
	(0.52, 0.90)	(0.48, 0.83)	(0.22, 0.38)	(0.75, 1.33)	(0.51, 0.94)
7	0.42	0.41	0.14	0.61	0.39
	(0.37, 0.47)	(0.36, 0.46)	(0.12, 0.16)	(0.53, 0.69)	(0.33, 0.47)
8	0.43	0.35	0.15	0.84	0.48
	(0.38, 0.47)	(0.31, 0.39)	(0.14, 0.18)	(0.73, 0.94)	(0.41, 0.57)
9	0.31	0.33	0.11	0.49	0.30
	(0.28, 0.35)	(0.29, 0.37)	(0.10, 0.12)	(0.43, 0.55)	(0.26, 0.35)
England	0.38	0.38	0.16	0.50	0.38
	(0.35, 0.41)	(0.35, 0.40)	(0.15, 0.16)	(0.45, 0.55)	(0.34, 0.42)

group with non-overlapping credible intervals; for females, posterior means of 0.16% to 0.38% for the younger and older age groups, respectively, with non-overlapping credible intervals. We note that the previous analysis of King *et al* (2009), using data from 2003-5, estimated significantly higher injecting DRD rates for the cross-classified groups in Scotland but, unlike this analysis, only identified a lower female DRD rate for young injectors with no gender differential for older injectors. For England, more definitively than for Scotland, we observe that older females' injecting DRD rate is also significantly lower than for older males (posterior mean of 0.38% versus 0.5% with non-overlapping credible intervals). See King *et al* (2009) for further details and results relating to the analyses of the Scottish data.

We now consider the results at the regional level. Comparing the results in Table 5 with the 80% prior interval for DRD rate, it is clear that region 5 (North West) appears to be the most at odds with these prior beliefs, with the upper 97.5% posterior quantiles of injectors' DRD rates all lower than 0.3% (the lower 5% prior quantile) for each gender $\times$ age-group. Comparing the prior and posterior distributions of numbers of IDUs in Figure 1 we see no visible overlap between these distributions. The significantly higher posterior estimate of the population size (compared to the prior specification) consequently produces the lower estimates of the injecting DRD rates.

For all regions, the lowest injecting DRD rates are for females in the younger age-group (15-34), with many regions having an injecting DRD rate in the lower 5% quantile of the prior interval. Overall, the female injecting DRD rates are generally lower than for the males. However, we note that the largest injector DRD rate occurs for females aged 35-64 in region 4 (North East). This also corresponds to the smallest estimated cross-classification population size over all regions. As discussed above, the older age-group (35-64) generally

has a higher injecting DRD rate for both males and females, relative to the younger age-group (15-34), with the exception of region 5 for males.

Three regions in Table 5 (1, East of England; 6, South East; and 4, North East) had particularly high injecting DRD rates, the first two of which (1 and 6) can be seen from Table 4 as regions with the lowest prevalence of current injectors per 1,000 of the population aged 15-64. The North East (region 4) shares with Yorkshire and Humberside (region 9) the burden of equally high injector prevalence but DRD rates that appear to be more than double those of injectors in Yorkshire and Humberside. A possible reason for the successful management of the injecting DRD rates in region 9 becomes apparent in the next subsection, when we consider the interactions present between the sources. Higher injecting DRD rates in low prevalence regions (1 and 6) may reflect a lesser priority accorded to IDUs by dint of lower injector prevalence.

Finally, we note that England's injecting DRD rate, as defined by us, is in line with the DRD rate of 0.36% reported by Merrall *et al* (2010) for all Scotland's drug treatment clients in the five years to the end of March 2006 and considerably lower than the 1% total mortality reported by Cornish *et al* (2010) for 5,577 patients on the UK General Practice Research Database who were prescribed methadone in 1990-2005 and followed-up until one year after expiry of their last prescription for opiate substitution therapy. However, drug treatment clients' total mortality is likely to be twice their DRD rate, see Merrall *et al* (2010).

4.3. Marginal Log-linear Probabilities

The corresponding marginal posterior probability that each covariate is present in the model for each separate region is provided in Table 6. Note that we identify evidence of the presence of an interaction when the posterior model probability is ≥ 0.75, corresponding to a Bayes factor of ≥ 3 (Kass and Raftery, 1995). There are several points of interest. Multiple interactions are clearly important across all regions, namely, $S1 \times S2$ (probation data $\times$ DIP prison assessment data); $S1 \times S3$ (probation data $\times$ drug treatment data); $S2 \times$ Sex; $S2 \times$ Age; $S4 \times$ Age (DIP community assessment data $\times$ Age). For all these interactions, the sign of the interaction is consistent across all regions. In particular, decreased probability of being observed by source $S2$ (DIP prison assessment data) for females and the older age-group; a decreased probability of being observed by source $S4$ (DIP community assessment data) for the older age group; and positive interactions for $S1 \times S2$ and $S1 \times S3$, which are precisely the sort of cross-linkage that policy initiatives had been designed to engender.

There are some discrepancies over the different regions regarding the presence of particular interactions. These include:

- Regions 3 (London) and 9 (Yorkshire and Humber): the only regions to identify the interaction $S2 \times S3$ (DIP prison assessment data $\times$ drug treatment data), despite large investment in the DIP initiative to lead to increased drug treatment. As we would expect, when this interaction is identified, it is estimated to be positive. We note further that the presence of this interaction in region 9 (Yorkshire and Humberside) and absence in region 4 (North East) may explain the lower injecting DRD rate in region 9 compared to region 4 despite both being high injector prevalence areas, as shown in Table 4.

- Region 3 (London): the only region that identifies the interaction $S3 \times$ Age, with older individuals more likely to be observed by the treatment data. However, for this

Table 6. Marginal posterior probability for each two-way interaction being present in the model for each region. Reca
probation data; $S2$ = DIP prison assessment data; $S3$ = drug treatment data; $S4$ = DIP community assessment data.

Interactions	Region								
Source × Source	1	2	3	4	5	6	7	8	9
$S1 \times S2$	0.993	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$S1 \times S3$	0.995	1.000	0.950	1.000	1.000	0.996	0.993	1.000	1.000
$S2 \times S3$	0.069	0.049	0.974	0.143	0.053	0.157	0.064	0.073	0.962
$S1 \times S4$	1.000	1.000	1.000	1.000	1.000	0.354	1.000	1.000	1.000
$S2 \times S4$	1.000	0.998	1.000	1.000	1.000	0.085	0.671	1.000	1.000
$S3 \times S4$	0.058	1.000	0.993	1.000	1.000	0.183	0.993	1.000	1.000
Source × Covariate									
$S1$ × Sex	0.047	0.139	0.175	0.031	0.999	0.110	0.995	0.995	0.566
$S2$ × Sex	0.999	1.000	0.793	0.962	1.000	0.998	1.000	1.000	1.000
$S3$ × Sex	0.979	0.977	0.972	0.077	0.029	0.038	0.156	0.064	0.031
$S4$ × Sex	0.105	0.436	0.328	0.823	1.000	0.914	1.000	0.953	1.000
$S1$ × Age	0.678	1.000	0.118	1.000	1.000	0.984	1.000	0.972	0.938
$S2$ × Age	0.933	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
$S3$ × Age	0.110	0.057	0.999	0.308	0.027	0.042	0.058	0.061	0.063
$S4$ × Age	0.997	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Covariate×Covariate									
Age × Sex	0.999	0.786	1.000	1.000	1.000	1.000	1.000	0.500	1.000

region, the interaction $S1$ × Age is not identified whereby, in other regions, fewer younger individuals are identified via source $S1$ (probation data) though the presence of the interaction is not strong in region 1 (East of England).

- Region 6 (South East): does not identify the interactions $S1 \times S4$ and $S2 \times S4$, (between DIP community assessment data and probation data and DIP prison assessment data, respectively) although positive interactions between these sources are strongly identified for all other regions with the exception of region 7 (South West) for the $S2 \times S4$ interaction where there is reasonable uncertainty as to its presence. Once more we note that, as would be expected for the interaction between the two DIP assessments, it is identified as a positive interaction.

- Regions 1 and 6 (East of England and South East): do not identify the interaction $S3 \times S4$ (treatment data and DIP community assessment data), although this is strongly identified as positive in all other regions, which again is a sought-after cross-linkage.

- Region 8 (West Midlands): provides uncertainty regarding the presence/absence of an interaction between Sex × Age, which is identified by the other regions. When present, the interaction corresponds to fewer females being at the older age-group (or conversely more males in the younger age-group).

Finally, we return to the comparison of results obtained within this analysis and those of Hay *et al* (2009). Recall that Hay *et al* (2009) consider the data at the DAT area level, did not include the covariate information and considered only the set of log-linear models with a maximum of two source × source interactions. For all regions, except region 6 (South East), the number of source × source interactions identified in our models typically lies between 4

and 6. Further, all of the source × source interactions that are identified with large posterior support for each region have a posterior mean that is positive. Thus, not including such interactions (as for 8 of the 9 regions) leads to the decreased estimate of population size obtained by the previous analysis of Hay *et al* (2009), rather than differences due to the use of the lower DAT area level data and ignoring the gender and age-group covariate information. For region 6 (South East), Hay *et al* (2009) provide an overall estimate and 95% confidence interval (rounded to nearest 10) of 13270 (10290, 16380). This is reasonably comparable to the estimate provided in Table 2 for this region, with both point estimates contained in the alternative analysis's uncertainty interval but this is not the case for any other region.

5. Discussion

Estimating the number of IDUs and the injecting DRD rate is an inherently difficult problem due to the nature of such hidden populations who, nonetheless, have a clear social and economic impact within society. The use of log-linear models is appealing due to their direct modelling (and interpretation) of interactions between the different data sources and/or covariates which are likely to be present within such complex systems. Using characteristics as covariates (gender and age-group) permitted cross-classified estimates for males and females in each age group (15-34, 35-64) for each region, and the identification of more complex underlying structure and/or patterns. For example, for female IDUs in the North East, an unusually high proportion are younger individuals (15-34) with a posterior mean (95% CI) of 5.41 (4.41, 7.71) of the younger to older (i.e. 15-34 to 35-64) ratio. Conversely, for male IDUs, the mean younger to older ratio is very similar to the overall mean ratio for England. We also note that, consistently within each region, and aggregated to the England level, the younger to older ratio is higher for females than males, indicating that a larger proportion of younger IDUs are female than of older IDUs.

We combine the estimate of number of IDUs with the number of injecting DRDs to obtain an estimate of the injecting DRD rate. Within our analysis, we take the number of injecting DRDs to be the average annual number of heroin-related deaths in each region over the period 2004-7. Thus, we do not include any level of uncertainty on this estimate of injecting DRDs so that all the variability in our estimates of injecting DRDs comes from the uncertainty in the estimates of population size. Similarly the prior interval specified on total population size comes from the uncertainty interval placed on injecting DRD rates. Adding uncertainty to the number of DRDs per region would increase the prior interval on total population size, but this would have little impact on the posterior estimates of prevalence of IDUs since the posterior distributions are largely data-driven (although doing so may create greater overlaps between the prior and posterior estimates of population size). Consequently, assuming a Poisson distribution, say, for the annual number of DRDs (with mean equal to the observed annual mean number of deaths) would result in essentially the same posterior mean for the injecting DRD rate, but with an inflated credible interval width to reflect the additional level of uncertainty incorporated.

The estimates of IDU prevalence and injecting DRD rates are model-dependent. In other words, the interactions present have a direct impact on the estimates obtained for the total population size, and hence on the injecting DRD rates. We used a model-averaging approach to incorporate both parameter and model uncertainty within the estimate for total population size, and associated statistics. However, the underlying model itself is also

of interest in terms of the propensity to be listed on a different source given observation (or not) on another source via source × source interactions. This provides direct insight into cross-linkages between the different data sources, and potentially insight into differences across regions. In particular, missing cross-linkages, which DIPs were designed to facilitate, were identified for regions 1 and 6 (East of England and South East) which have low injector prevalence per 1,000 of the population aged 15-64 but also high injecting DRD rates. Our analysis may lend ecological support to the notion that if the DIP sought-after cross-linkages are not properly in place, regions may experience higher injecting DRD rates. In Yorkshire and Humberside, where cross-linkages between DIP prison assessment and treatment data was identified, injecting DRD rates appeared to be lower than the overall averaged England estimates but the relatively youthful profile of the region's current injectors needs to be addressed separately.

That so many positive source × source interactions were supported by the data from English regions signals the success of cross-departmental initiatives for criminal justice to recognize opiate/cocaine dependency or injecting risk and to encourage relevant arrestees, probationers or prisoners to engage with drug treatment agencies. By the same token, it is unwise for those who commission capture-recapture studies to prescribe how the analysis shall be tackled. Insistence on injector estimates at the level of DAT area means that, were there to be 64 cross-counts as here per DAT area, many cells would be empty and more would have only low counts so that accommodation of many source × source interactions, despite their relevance, becomes both technically and computationally infeasible. To the extent that England's number of current injectors would typically be under-estimated and injecting DRD-rates over-estimated, the impact of opiate-substitute therapy on saving life may be under-estimated. In addition, complacency may be engendered about the transmission risk for blood-born viruses by dint of under-estimating potential transmitters, namely infectious current injectors.

The generally lower injecting DRD rates for England than in Scotland, the basis for our priors, suggests that Scotland could learn from the cross-linkages that England has put in place. Discussion of source × source interactions with regions' criminal justice or drug treatment practitioners may shed further light on their regional implications when local expertise is brought to bear on their interpretation. This analysis is broadly reassuring that criminal justice and drug treatment interventions have delivered but there are concerns also - particularly for those regions in which injector ratios by age-group (15-34 to 35-64) are high and thereby suggest an unwelcome preponderance of younger injectors, which means that greater resistance to injecting needs to be engendered in young people.

Acknowledgements

We would like to thank the Home Office and the research team at the Centre for Drug Misuse Research for providing the requested cross-count contingency table data, Claudia Wells at the Office of National Statistics for the requested number of drug-related deaths by toxicology and cross-classified by region, sex and age-group and Tim Millar for some useful discussions. SMB is funded by the Medical Research Council (WBS number U.1052.00.002.00001.01) and RK was partly funded by the MRC Addiction Research Cluster, NIQUAD (Nationally Integrated Quantitative Understanding of Addiction Harm).

A. Calculation of Jacobian

We provide the calculation of the Jacobian in the transformation of variables from the prior on the parameters N_{tot}, R, P_1, P_2, on which the prior information is specified to the model parameters $\boldsymbol{N} = \{N_{(M,15-34)}, N_{(F,15-34)}, N_{(M,35-64)}, N_{(F,35-64)}\}$. We have the relationship between these parameters of the form.

$$\begin{aligned}
N_{tot} &= N_{(M,15-34)} + N_{(F,15-34)} + N_{(M,35-64)} + N_{(F,35-64)};\\
R &= \frac{N_{(M,15-34)} + N_{(M,35-64)}}{N_{(F,15-34)} + N_{(F,35-64)}} = \frac{N_{(M)}}{N_{(F)}};\\
P_1 &= \frac{N_{(M,15-34)}}{N_{(M,15-34)} + N_{(M,35-64)}} = \frac{N_{(M,15-34)}}{N_{(M)}};\\
P_2 &= \frac{N_{(F,15-34)}}{N_{(F,15-34)} + N_{(F,35-64)}} = \frac{N_{(F,15-34)}}{N_{(F)}}.
\end{aligned}$$

Consequently, the corresponding Jacobian is given by,

$$\begin{aligned}
\left|\frac{d(N_{tot}, R, P_1, P_2)}{d\boldsymbol{N}}\right| &= \begin{vmatrix}
1 & 1 & 1 & 1\\
\frac{1}{N_{(M)}} - \frac{N_{(M,15-34)}}{N^2_{(M)}} & -\frac{N_{(M,15-34)}}{N^2_{(M)}} & 0 & 0\\
\frac{1}{N_{(F)}} & \frac{1}{N_{(F)}} & -\frac{N_{(M)}}{N^2_{(F)}} & -\frac{N_{(M)}}{N^2_{(F)}}\\
0 & 0 & \frac{1}{N_{(F)}} - \frac{N_{(F,15-34)}}{N^2_{(F)}} & -\frac{N_{(F,15-34)}}{N^2_{(F)}}
\end{vmatrix}\\
&= \frac{N_{(M,15-34)} + N_{(F,15-34)} + N_{(M,35-64)} + N_{(F,35-64)}}{N_{(M)}N^3_{(F)}}\\
&= \frac{N_{tot}}{N_{(M)}N^3_{(F)}}.
\end{aligned}$$

This result is easily obtained using an algebraic computer package, such as Maple.

References

Beynon, C., Bellis, M. A., Millar, T., Meier, P., Thomson, R. and Jones, K. M. (2001), Hidden need for drug treatment services: measuring levels of problematic drug use in the North West of England. *Journal of Public Health Medicine* **23** 286–91.

Bird, S. M., Hutchinson, S. J. Hay, G. and King, R. (2010), Missing targets on drug-related deaths, and a Scottish paradox. *International Journal of Drug Policy* **21** 155–159.

Bird, S. M. and Robertson, J. R. and Strang, R. (2010), Drug-related deaths: what now needs to be done. published in *Straight Statistics* (http://www.straightstatistics.org/), 10/11/2010.

Bloor, M., Gannon, M., Hay, G., Jackson, G., Leyland, A. H., and McKeganey, N. (2008), Contribution of problem drug users' deaths to excess mortality in Scotland: secondary analysis of cohort study. *British Medical Journal* **337**:a478.

Brooks, S. P. (1998), Markov chain Monte Carlo method and its application *The Statistician* **47** 69–100.

Chao, A., Tsay, P. K., Lin, S., Shau, W. and Chao, D. (2001), The application of capture-recapture models to epidemiological data. *Statistics in Medicine* **20** 2123–57.

Cornish, R., Macleod, J., Strang, J., Vickerman, P. and Hickman, M. (2010), Risk of death during and after opiate substitution treatment in primary care: prospective observational study in UK General Practice Research Database. *British Medical Journal* **341**:c5475.

de Angelis, D., Hickman, M. and Yang, S. Y. (2004) Estimating Long-term Trends in the Incidence and Prevalence of Opiate Use/Injecting Drug Use and the Number of Former Users: Back-Calculation Methods and Opiate Overdose Deaths. *American Journal of Epidemiology* **160** 994–1004.

Fienberg, S. E. (1972), The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika* **59** 591–603.

Frischer, M., Leyland, A., Cormack, R., Goldberg, D. J., Bloor, M., Green, S. T., Taylor, A., Covell, R., McKeganey, N. and Platt, S. (1993), Estimating the population prevalence of injection drug use and infection with human immunodeficiency virus among injection drug users in Glasgow, Scotland. *American Journal of Epidemiology* **138** 170–81.

Green, P. J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82** 711–32.

Hay, G., Gannon, M., MacDougall, J., Eastwood, C., Williams, K. and Mllar, T. (2009), Capture-recapture and anchored prevalence estimation of injecting drug users in England: national and regional estimates. *Statistical Methods in Medical Research* **18** 323–339.

Hook, E. B. and Regal, R. R. (1995), Capture-recapture methods in epidemiology: Methods and limitations *Epidemiologic Reviews* **17** 243–64.

Hook, E. B., Albright, S. G. and Cross, P. K. (1980), Use of Bernoulli census and log-linear methods for estimating the prevalence of spina bifida in livebirths and the completeness of vital record reports in New York state. *American Journal of Epidemiology* **112** 750–758.

Hutchinson, S. J., Bird S. M., Taylor, A. and Goldberg, D. J. (2006), Estimating the prevalence, incidence and cessation of injecting drug use in Glasgow 1960-2000: combining expert opinion with capture-recapture prevalence data. *International Journal of Drug Policy* **17** 29–34.

Hutchinson, S. J., Taylor, A., Gruer, L., Barr C., Mills, C., Elliott, L., Goldberg, D. J., Scott, R. and Gilchrist, G. (2000), One year follow-up of opiate injectors treated with oral methadone in a GP-centred programme. *Addiction* **95** 1055–1068.

Kass, R. E. and Raftery, A. E. (1995), Bayes factors. *Journal of the American Statistical Association* **90** 773–93.

King, R., Bird, S. M., Brooks, S. P., Hutchinson, S. J. and Hay, G. (2005), Prior information in behavioural capture-recapture methods: demographic influences on drug injectings' propensity to be listed in data sources and their drug-related mortality. *American Journal of Epidemiology* **162** 694–703.

King, R., Bird, S. M., Hay, G. and Hutchinson, S. J. (2009), An update on the estimation of the prevalence of injector-drug users in Scotland via capture-recapture methods. *Statistical Methods in Medical Research* **18** 341–59.

King, R. and Brooks, S. P. (2001), On the Bayesian analysis of population size. *Biometrika* **88** 317–36.

Madigan, D. and York, J. C. (1997) Bayesian methods for estimation of the size of a closed population. *Biometrika* **84** 19–31.

Mastro, T. D., Kitayaporn, D. and Weniger, B. G. (1994) Estimating the number of HIV-infected injection drug users in Bangkok: A capture-recapture method. *American Journal of Public Health* **84** 1094–99.

Merrall E. L. C, Bird S. M., Hutchinson S. J. (2010) Mortality of those who attended drug services in Scotland 1996-2006: record linkage study. *Technical Report*.

Millar, T., Gemmel, I., Hay, G., Heller, R. F. and Donmall, M. (2006) How well do trends in incidence of heroin use reflect hypothesised trends in prevalence of problem drug use in the North West of England? *Addiction Research and Theory* **14** 537–549.

Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978), Statistical inference from capture data on closed animal populations. *Wildlife Monographs* **62** 1–135.

Strang, J., Hall, W., Hickman, M. and Bird, S. M. (2010), Impact of supervision of methadone consumption on deaths related to methadone overdose (1993-2008): analyses using OD4 index in England and Scotland. *British Medical Journal* **341**:c4851.