International Journal of AI & Data Science
Review Volume 2, Issue 1, 2026

Indirect Area Estimates of Disease Prevalence: Bayesian Evidence Synthesis with an Application to Coronary Heart Disease

Ahmadi Livani
Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran

Article Info

Received: 08-09-2025 Revised:11-10-2025 Accepted:22-12-2025 Published:04-01-2026

***Abstract*:** The incidence of many chronic illnesses, including as cancer, diabetes, asthma, mental illness, and coronary heart disease, varies greatly across regions due to substantial correlations between risk factors and socioeconomic and ethnic groupings. When evaluating health care demands and comparing health care provision (such as surgical intervention rates) to health needs, variations in prevalence are crucial. In order to provide indirect prevalence estimates for geographic areas where prevalence statistics are not otherwise accessible, this work focuses on predicting the prevalence of coronary heart disease using a Bayesian technique to synthesise information of various sorts. Regression techniques may be used to get gradients by age, sex, ethnicity, wide area, and socioeconomic level. One source is information on prevalence risk gradients from national health survey data, which often only offer regional identifiers (for confidentiality reasons). A number of health surveys are often accessible, and by using data on prevalence gradients from previous surveys, one may think about pooling strength across surveys (e.g. via a power prior technique). The second source of data is population totals by age, sex, ethnicity, etc. from intercensal population estimates or censuses, to which prevalence rates derived from surveys are added. Since there is a significant link between the frequency of heart disease and death from some major chronic illnesses across regions, data on area mortality is another possible source of information. Using (a) data from the Health Surveys for England for 2003 and 1999, (b) population data from the 2001 UK Census, and (c) area mortality data for 2003, a case study examines the growth of estimates of coronary heart disease prevalence in 354 English regions.

Key words: Coronary heart disease, deprivation, ethnicity, power prior, prevalence. mortality, small area.

Introduction: Need for Spatially Disaggregated Prevalence Estimates For important illnesses, local area prevalence statistics are often either not collected at all or, if they are, are skewed by administrative and measurement biases. Nonetheless, a few of nations regularly conduct national health surveys that reveal national patterns in prevalence. While health managers need data at a much higher geographically disaggregated scale and for those strata (age, sex, ethnicity) by which area populations are recorded—in censuses or by intercensal population estimates—such surveys usually only provide broad regional identifiers. Additionally, the estimations should include how socioeconomic variables affect the prevalence of chronic diseases. In geographic applications, deprivation indices—a term used to describe suffering brought on by low income, substandard housing, high unemployment rates, etc.—are indicators of the socioeconomic situation of the local population. The approach for evaluating neighbourhood disadvantage has advanced significantly in the UK (e.g. Noble et al., 2000; Bailey et al., 2003), particularly in small areas with 1500–2000 residents, of which there are around 32500 in England (ONS, 2006).

In order to establish prevalence estimates for chronic diseases for 354 English regions, with an emphasis on heart disease, this research presents a Bayesian technique. National health surveys are the initial source of data. The Swedish National Public Health Survey, the Italian National Health

Survey, and the Taiwan National Health Interview Survey are just a few examples of the many health surveys that are accessible in various countries. One may think of pooling strength over surveys. The 2003 Health Survey for England is used in this research, with historical data from a 1999 survey provided using a power prior technique (Chen et al., 2000). The spatial scale used in the two surveys consists of nine government regions (the North East of England, the North West, Yorkshire & Humberside, the East Midlands, the West Midlands, Eastern England, London, South East England, and South West England), with the exception of the neighbourhood deprivation category (the quintile rank among 32500 neighbourhoods, with no further identifying information). Table 1 summarises regional differences. Survey data on gradients by age, sex, ethnicity, broad area, and neighbourhood disadvantage are modelled using a binomial regression.

Survey-based prevalence rates should be applied to population totals by age, sex, ethnicity, etc. from censuses or intercensal population estimates as the second source of data. Intercensal population estimates by age, sex, and ethnicity are not created in the UK, although they are in other nations like the US (Smith, 1998). The populations used here are particular for age, sex, and ethnic group and are taken from the UK 2001 Census.

In contrast to prevalence, death statistics are usually well-recorded at a disaggregated geographical level, making them the third source of relevant information. There is evidence of a positive association between the prevalence of heart disease and death. This suggests that survey-based estimates of area prevalence should be adjusted to account for proxy information on prevalence supplied by mortality data across the 354 regions. Other sources of local data, such as hospital admissions or referrals to community care, might be used as proxies for morbidity when death is only seldom associated with a specific form of morbidity (such as asthma or mental disease).

Therefore, the methodology offers a method for indirect prevalence estimation by applying survey-based gradients for heart disease over those stratifiers by which populations for areas are available (e.g., age, sex, ethnic group), while also accounting for neighbourhood deprivation and proxy data on prevalence (from mortality) at the necessary area level. The approach uses a completely Bayesian approach, updating prior densities on parameters based on the probability of the observed data. The WINBUGS programme (Spiegelhalter et al., 2003) employs iterative Monte Carlo Markov Chain methods (Gelfand and Smith, 1990) for model estimation.

The survey-based portion of the prevalence estimate process is described in the next four parts. A section on the combined modelling of area mortality and prevalence, which allows prevalence estimates to take into account geographical mortality patterns, follows them. The last part examines potential advancements in the technique.

Survey Model: Populations, Survey Variables and Choice of Binomial Link

Equivalent variables must be provided for respondents in national health surveys and Census populations (or intercensal population estimates) in order to utilise survey evidence on disease gradients to estimate prevalence in local populations. Numerous nations provide population statistics broken down by age, sex, and ethnicity. The UK's 2001 Census, for instance, contains a tabulation of inhabitants by these factors.

Socioeconomic gradients in illness prevalence, such as those based on an individual's employment standing or the degree of deprivation in their area, are usually also important to consider. This implies that populations by age, sex, ethnicity, and profession, or populations by age, sex, ethnicity, and degree of neighbourhood disadvantage, would be optimal. However, populations are not known to this degree of detail in many nations. For

For instance, there is no four-way disaggregation by age, sex, race, and neighbourhood disadvantage in the UK Census. In these situations, it is suggested that a survey-based prevalence gradient over deprivation levels be used to scale prevalence estimates by age, sex, and ethnicity.

Model-based rates of heart disease prevalence are obtained by applying a binomial regression to data from the 2003 and 1999 Health Surveys for England. If a survey participant reported experiencing angina or a heart attack during the last year, as verified by a physician, they are considered to have coronary disease. Age ($a = 1, ..., 7$, i.e., ages 0-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+), sex ($s = 1, 2$; i.e., male, female), ethnicity ($e = 1, ..., 4$; i.e., white, black, south Asian, and all other ethnic groups), and

regions (r = 1, ..., 9), as shown in Table 1. The neighbourhood deprivation quintile (d = 1,..., 5, with d = 5 for the most impoverished) is also included in the 2003 survey. Risk category cells are used to group respondents; they are referred to as separate covariate patterns by Greenland (2001). Thus, after accounting for survey weighting for differential non-response, the observations become numbers at risk, naserd, and ill individuals yaserd (JHSU, 2004).

Table 1: Age standardised chd prevalence (ages 35+) with 95% confidence intervals by sex & government office region 2003 Health Survey for England.

	North East	North West	Yorkshire & Humberside	East Midlands	West Midlands
Males	7.5	9.4	11.6	9.2	10.5
2.5%	4.2	7.2	8.7	6.4	7.8
97.5%	10.8	11.6	14.5	12.0	13.2
Females	9.9	6.6	7.9	7.1	7.4
2.5%	6.4	4.8	5.6	4.7	5.2
97.5%	13.4	8.4	10.2	9.5	9.6

	East	London	South East	South West	England
Males	7.3	8.1	8.2	6.6	8.8
2.5%	5.1	5.9	6.3	4.5	8.0
97.5%	9.5	10.3	10.1	8.7	9.6
Females	3.9	4.6	4.3	3.9	5.8
2.5%	2.3	3.0	2.9	2.3	5.2
97.5%	5.5	6.2	5.7	5.5	6.5

Applying a log-binomial regression, commonly known as the log-linear binomial (Greenland, 2004), permits inferences on prevalence pro-portion ratios rather than prevalence odds ratios (Skov et al., 1998; Zocchetti et al. 1997). For instance, although logit coefficients only yield relative risks under the premise of a rare illness, this relationship allows the estimate of the prevalence relative risk gradient across neighbourhood deprivation quintiles. Cell probabilities were limited to a maximum of 0.999 in order to prevent probability exceeding 1. Only the first 100–200 iterations of MCMC sampling yielded this default value.

The prevalence model for the 2003 survey data incorporates main effects in all variables and second order interactions for which there is evidence in the literature on health outcomes, not necessarily heart disease, in order to evaluate potential interactions between risk factors. With the exception of removing the primary deprivation impact and any interactions involving deprivation, the historical data model (for the 1999 survey) is identical. Age-sex, sex-region, sex-ethnicity, sex-deprivation, age-deprivation, age-ethnicity, and ethnicity-deprivation are among the second-order interactions that are considered. While several gender-age heart disease risk profiles and sex-ethnicity interactions have been documented, Table

1 suggests the sex-region relationship (Primatesta and Brookes, 2000). Some studies of cardiovascular outcomes that incorporate area deprivation and area type show few interactions as meaningful, even if completeness in modelling terms may suggest incorporating multiple interactions (e.g. Martinez et al., 2003). In order to achieve parsimony, the age and deprivation variables are reframed as binary when included in interactions:

A model including main effects and the above mentioned interactions is then

$$\begin{aligned} y_{aserd} &\sim \text{Bin}(n_{aserd}, \rho_{aserd}) \\ \log(\rho_{aserd}) &= \alpha_{1s} + \alpha_{2a} + \alpha_{3e} + \alpha_{4d} + \beta_{1es} + \beta_{2a^*s} \\ &\quad + \beta_{3ea^*} + \beta_{4a^*d^*} + \beta_{5d^*s} + \beta_{6ed^*} + \gamma_{rs} \end{aligned} \quad (3.1)$$

where parameters treated as fixed effects $\{\alpha_{2a}, \alpha_{3e}, \alpha_{4d}, \beta_{1es}, \beta_{2a^*s}, \beta_{3ea^*}, \beta_{4a^*d^*}, \beta_{5d^*s}, \beta_{6ed^*}\}$ are subject to corner constraints, except for the gender terms α_{1s} that are both taken as The priors on $\{\alpha, \beta\}$ in (3.1) and (3.2) are derived from accumulated epidemiological data, such as the prevalence of treated heart disease found in UK studies. Heart disease prevalence rates range from 0.1 per 1000 for males and females under 34 to 205/1000 for males and 172/1000 for females over 85, according to data from the Key Health Statistics from General Practice (ONS, 2000). Normal priors N (−9, 5) and N (0, 5) are chosen for α1s and α2a, respectively, due to this broad range in risk. According to accumulated evidence (e.g., Hoare, 2003), N (0, 1) priors will cover likely ranges in relative risk for the remaining risk factors (for ethnic and deprivation categories) and the interactions. This is consistent with an earlier estimate that, with 95% confidence, the related relative risks will fall between 0.14 and 7.1. When a log link is employed in the binomial regression, it is easy to incorporate more informative elicited priors on relative risk, which may be justified (e.g. Greenland, 2001).

The regional effects γ_{rs} are treated as random and follow a bivariate spatial conditional autoregressive prior (see Appendix 1), with the multivariate CAR

precision matrix Φ^{-1} assumed to follow a Wishart prior with 2 degrees of freedom and identity scale matrix. Reasons for expecting spatial correlation in regional relativities include the north-south contrast in prevalence (Table 1), as well as environmental factors, such as water hardness (Shaper *et al.*, 1980; Catling *et al.*, 2005).

Let $\theta = \{\alpha, \beta, \gamma, \Phi_\gamma\}$ parameters, and $0 \leq \delta \leq 1$ be a precision parameter (with beta prior) that weights the historical data D_h relative to the likelihood of the current study data D. Following Chen *et al.* (2000, p. 124) the power prior

takes the form

$$\pi(\theta.\delta|D_h) \propto [P(D_h|\theta)]^\delta \delta^{a_\delta - 1}(1-\delta)^{b_\delta - 1} \quad (3.3)$$

where $P(D_h|\theta)$ is the binomial likelihood, and (a_δ, b_{delta}) are pre-specified beta density hyperparameters. With δ an unknown the joint posterior density for (θ, δ) is then

$$P(\theta, \delta||D, D_h) \propto P(D|\theta)[P(D_h|\theta)]^\delta \delta^{a_{delta}-1}(1-\delta)^{b\delta-1}. \quad (3.4)$$

For the current analysis there is expected to be considerable continuity between the two surveys in prevalence differentials and the 1999 survey data includes rel-evant information on ethnic prevalence gradients; on the other hand, the model forms for 2003 and 1999 differ (because only the 2003 model includes neigh-bourhood deprivation) and so some downweighting is appropriate. Here three alternative beta priors for δ are considered, namely Be(250,1), Be(100,1) and Be(50,1).

Survey Model Specification and Pooling over Surveys

the top two deprivation quintiles (d∗ = 2) are compared with the lowest three (d∗ = 1), and ages up to 64 (a∗ = 1) are compared with ages 65 and above (a∗ = 2). There is substantial evidence to support such a contraction; for instance, impoverishment mostly affects mortality and early illness (e.g. Barnett et al., 2001).

Table 2: Parameter summaries with 95% credible intervals, survey models

Prior on δ		$\delta \sim$ Be(250,1)		$\delta \sim$ Be(100,1)		$\delta \sim$ Be(50,1)	
Fit		$p_e = 26.6, DIC = 1561$		$p_e = 28.7, DIC = 1563$		$p_e = 31, DIC = 1566$	
Risk group	Parameter	Mean	95% Interval	Mean	95% Interval	Mean	95% Interval
	δ	0.97	(0.9,1)	0.42	(0.34,0.51)	0.20	(0.15,0.26)
Males	α_{11}	-7.06	(-7.59,-6.35)	-6.90	(-7.47,-6.45)	-6.77	(-7.43,-5.94)
Females	α_{12}	-7.70	(-8.29,-6.95)	-7.55	(-8.12,-7.07)	-7.46	(-8.15,-6.66)
Ages 35-44	α_{22}	2.43	(1.65,3.1)	2.27	(1.69,2.87)	2.15	(1.3,2.91)
Ages 45-54	α_{23}	3.81	(3.1,4.38)	3.58	(3.08,4.15)	3.43	(2.61,4.11)
Ages 55-64	α_{24}	4.85	(4.13,5.41)	4.70	(4.23,5.27)	4.58	(3.8,5.24)
Ages 65-74	α_{25}	5.54	(4.83,6.08)	5.41	(4.94,6)	5.28	(4.46,5.9)
Ages 75-84	α_{26}	5.84	(5.14,6.39)	5.72	(5.23,6.3)	5.59	(4.75,6.23)
Ages 85+	α_{27}	5.95	(5.26,6.53)	5.84	(5.34,6.46)	5.72	(4.89,6.38)
Black	α_{32}	-0.54	(-1.09,-0.01)	-0.44	(-1.14,0.2)	-0.31	(-1.1,0.43)
South Asian	α_{33}	0.43	(0.17,0.7)	0.33	(-0.05,0.68)	0.15	(-0.33,0.59)
Other	α_{34}	-0.52	(-1.69,0.55)	-0.67	(-1.98,0.53)	-0.72	(-2.08,0.39)
Neighbourhood Deprivation							
Quintile 1	α_{51}	-0.30	(-0.46,-0.14)	-0.31	(-0.47,-0.16)	-0.32	(-0.48,-0.16)
Quintile 2	α_{52}	-0.07	(-0.22,0.07)	-0.09	(-0.25,0.06)	-0.09	(-0.23,0.05)
Quintile 3	α_{53}	-0.10	(-0.26,0.05)	-0.12	(-0.27,0.04)	-0.12	(-0.28,0.02)
Quintile 4	α_{54}	0.08	(-0.08,0.25)	0.11	(-0.08,0.29)	0.11	(-0.06,0.28)
Quintile 5	α_{55}	0.39	(0.22,0.56)	0.42	(0.26,0.59)	0.42	(0.25,0.59)
Interactions							
Ethnic-Sex	β_{122}	0.86	(0.28,1.48)	0.80	(0.07,1.49)	0.77	(-0.11,1.64)
	β_{122}	-0.11	(-0.52,0.3)	-0.04	(-0.57,0.49)	0.06	(-0.57,0.69)
	β_{122}	-0.32	(-1.83,0.96)	-0.11	(-1.65,1.25)	-0.04	(-1.54,1.32)
Age-Sex	β_{122}	0.04	(-0.2,0.29)	0.05	(-0.23,0.34)	0.09	(-0.18,0.38)
Ethnic-Age	β_{122}	-0.76	(-1.35,-0.16	) -0.83	(-1.62,-0.12)	-0.87	(-1.77,-0.04)
	β_{122}	-0.40	(-0.77,-0.04	) -0.34	(-0.78,0.13)	-0.19	(-0.78,0.37)
	β_{122}	0.56	(-0.63,1.77	) 0.60	(-0.78,1.77)	0.71	(-0.56,2.02)

Table 2: (Continued) Parameter summaries, survey models

Prior on $\delta\delta \sim$ Be(250,1)		$\delta \sim$ Be(100,1)		$\delta \sim$ Be(50,1)			
Risk group	Parameter	Mean	95% Interval	Mean	95% Interval	Mean	95% Interval
Age-Deprivation	β_{422}	-0.15	(-0.41,0.1)	-0.22	(-0.48,0.04)	-0.23	(-0.49,0.03)
Sex-Deprivation	β_{522}	0.18	(-0.07,0.43)	0.17	(-0.08,0.43)	0.19	(-0.08,0.47)
Ethnic-Deprivation	β_{622}	0.05	(-0.7,0.75)	0.02	(-0.79,0.76)	-0.09	(-0.92,0.7)
	β_{632}	-0.36	(-0.96,0.14)	-0.34	(-0.94,0.2)	-0.26	(-0.89,0.33)
	β_{642}	-0.67	(-2.06,0.56)	-0.58	(-1.95,0.71)	-0.64	(-2.03,0.61)
Region Effects: Males							
N. East	γ_{11}	-0.05	(-0.31,0.2)	-0.07	(-0.35,0.17)	-0.08	(-0.36,0.17)
N. West	γ_{21}	0.04	(-0.13,0.2)	0.04	(-0.15,0.21)	0.04	(-0.14,0.21)
Yorks/Humb	γ_{31}	0.22	(0.04,0.4)	0.22	(0.03,0.4)	0.2	(0.01,0.39)
E. Midlands	γ_{41}	0.02	(-0.16,0.2)	0.03	(-0.16,0.21)	0.03	(-0.17,0.22)
W. Midlands	γ_{51}	0.06	(-0.1,0.23)	0.07	(-0.12,0.24)	0.07	(-0.12,0.25)
East	γ_{61}	-0.06	(-0.25,0.13)	-0.08	(-0.29,0.12)	-0.09	(-0.31,0.12)
London	γ_{71}	-0.07	(-0.24,0.11)	-0.04	(-0.24,0.16)	-0.01	(-0.23,0.19)
S. East	γ_{81}	-0.03	(-0.21,0.15)	-0.01	(-0.18,0.17)	0	(-0.18,0.17)
S. West	γ_{91}	-0.15	(-0.38,0.06)	-0.15	(-0.37,0.06)	-0.16	(-0.38,0.06)
Region Effects: Females							
N. East	γ_{12}	0.35	(0.08,0.6)	0.37	(0.11,0.64)	0.37	(0.08,0.63)
N. West	γ_{22}	0.03	(-0.17,0.22)	0.04	(-0.17,0.23)	0.04	(-0.16,0.24)
Yorks/Humb	γ_{32}	0.19	(-0.03,0.4)	0.21	(-0.01,0.43)	0.22	(-0.01,0.43)
E. Midlands	γ_{42}	0.13	(-0.08,0.35)	0.13	(-0.12,0.36)	0.12	(-0.11,0.35)
W. Midlands	γ_{52}	0.03	(-0.16,0.23)	0.05	(-0.16,0.27)	0.05	(-0.17,0.26)
East	γ_{62}	-0.17	(-0.43,0.07)	-0.2	(-0.47,0.05)	-0.19	(-0.47,0.07)
London	γ_{72}	-0.15	(-0.38,0.06)	-0.18	(-0.43,0.07)	-0.2	(-0.47,0.07)
S. East	γ_{82}	-0.14	(-0.35,0.07)	-0.15	(-0.38,0.05)	-0.16	(-0.39,0.06)
S West	γ_{92}	-0.27	(-0.56,0)	-0.26	(-0.57,0.02)	-0.25	(-0.53,0.03)

Survey Model Results

Iterations 1000–5000 of two chain sampling runs with dispersed starting values are used to draw conclusions; convergence is reached by iteration 1000 using the Gelman-Rubin criteria (Gelman et al., 1995). The deviance information criterion (DIC) of Spiegelhalter et al. (2002) is used to compare models. It consists of the posterior mean deviance plus a complexity measure pe, which is calculated as the difference between D^1 and the deviance, $Dev(\Psi^1)$, at the posterior mean of $\Psi = (\theta, \delta)$.

Thus, $Dev(\Psi^1) + 2pe$ can be used to obtain the DIC. New data (ynew, aserd) are sampled from the model for model checking, and the degree to which 95% intervals for new data include the actual data is used to determine compatibility with actual data (Gelfand, 1996).

Parameter estimates (log relative risks) under various δ pri-ors are displayed in Table 2. Although DICs slightly decrease with larger values of δ due to lower pe, average deviances are similar across the three options on δ. With actual cases yaserd in all cells covered by 95% intervals for replicate data sampled from P (ynew|y), regardless of the prior on δ, Gelfand's (1996) posterior predictive checking procedure is satisfactory.

While ethnic group effects for δ ~ Be(250,1) show lower risk for Blacks and higher risk for South Asians, age and sex effects are significant under all options (cf. Primatesta and Brookes, 2000). The deprivation effects (centred around their average) under all δ priors reveal a relatively flat intermediate effect and a primary contrast between the extremes of neighbourhood deprivation. A north-south prevalence contrast within England is supported by the regional effects; three coefficients are significant under δ ~ Be(250,1), and the posterior means (95% intervals) of the parameter contrasts γ31 − γ91, γ12-γ92, and γ32-γ92 are 0.37 (0.07,0.69), 0.61 (0.19,1.07), and 0.45 (0.08,0.85). The presence of regional effects even after adjusting for neighbourhood deprivation and population composition points to real contextual variation in the risk of heart disease. According to its 95% credible interval under the two more informative priors on δ, β122 of the interactions is significant, indicating a higher prevalence among black women than among men. Under δ ~ Be(250,1), both β322 and β332 are significantly negative, indicating a lower risk for Black and South Asian elders.

Relevant Survey Outputs for Prevalence-Mortality Model

The analysis's objective is to calculate the prevalence of heart disease in 354 English regions where Census population counts by age, sex, and ethnic group are available. There is no additional population disaggregation by neighbourhood deprivation quintile. However, it is possible to obtain population totals to make sure the area prevalence estimates account for neighbourhood deprivation.

N_{id}, providing proportions $w_{id} = N_{id}/N_i$ of total area population living in each deprivation quintile. Let $r_i \in \{1, \ldots, 9\}$ denote the region in which the i-th area is located, then the impact of neighbourhood deprivation in area i is based on averaging the probabilities ρ_{aser_id} according to the population split w_{id}. Then age-sex-ethnic specific prevalence rates R_{iase} are estimated as a weighted average

$$R_{iase} = \left(\sum_d w_{id}\rho_{aser_id}\right), \qquad (5.1)$$

and prevalent cases in each area and for age, sex and ethnic groups are estimated as $P_{iase} = R_{iase}N_{iase}$.

Ethnic group-specific mortality statistics are not available for the English area (see section 4).

The Piase are combined across ethnic groups (at each MCMC iteration) to create area-age totals in order to produce a prevalence rate that can be jointly modelled with mortality. Pias that are further subdivided by populations based on age, sex, and location Area-age-sex prevalence rates are provided by Nias. The age-standardized prevalence rates πis for each area by sex (and for ages over 35) are then obtained by applying these to European standard populations Sa (e.g. Hedman et al., 1999). Since the transforms xis = logit(πis) are more likely to be roughly normal than the prevalence rates themselves, they are monitored to provide an appropriate input for the joint prevalence-mortality analysis. Xis(Vis) represents the posterior means and variances of the xis.

Joint Prevalence-Mortality Model

As previously stated, area heart disease mortality provides evidence on variation in area prevalence; mortality is likely to closely reflect prevalence, though other factors may be involved. Heart disease registers linked to a new payment plan for English general practitioners provide evidence supporting a joint analysis (Strong et al., 2006). These data may not always offer a "gold standard" prevalence estimate due to under- or over-registration,

particularly at lower spatial scales. Nonetheless, Table 3 demonstrates a strong regional correlation (Pearson correlation of 0.9) between prevalence and the chosen mortality index.

Let ED and ED represent expected deaths (using England age-specific rates for 2003) and Di1 and Di2 represent area deaths for males and females (ages over 35). Assuming Poisson sampling, one has Dis ~ Po(μisED), where μis are the relative mortality risks for area i and sex s (1=M, 2=F). In order to account for the spatial patterning of these mortality relative risks, survey-based logit prevalence rate estimates xis must be adjusted.

Table 3: Associations between prevalence and mortality, regional level

Government Region	Standard Prevalence Ratio (Data from GP Payment System)	CHD Mortality 2001-3 Directly Standardised rates, Ages < 75
North East	132	76
North West	120	74
Yorkshire & Humberside	119	69
E. Midlands	100	62
W. Midlands	99	66
East	89	50
London	87	61
South East	86	50
South West	90	50

Let z be the underlying true (logits of) prevalence, measured with error. The bivariate model can be seen as following a form

$P(\mu, z|x) = P(\mu|z)P(z|x)P(x)$, (6.1) namely a marginal prevalence model and a model including an impact of preva-lence on mortality. Similar models are mentioned by X_{ia} and Carlin (1998) and Bernardinelli *et al.* (1997), and both of these studies incorporate spatial correla-tion in the underlying rate.

In the application here, spatially correlated effects pool information over areas and genders. There are many reasons to expect unmeasured risk factors to be spatially correlated, for example between adjacent urban as against rural areas, or between adjacent areas in northern as against southern regions of England. Thus, urban air pollution is a risk for heart disease (Chen *et al.*, 2005), while regional differences in smoking and physical activity are reported by Morris *et al.* (2003). Regional differences in drinking-water hardness have also been linked to cardiovascular disease variations (Monarca *et al.*, 2004; Catling *et al.*, 2005).

So for sexes $s = 1, 2$ the following joint model is postulated

$$D_{is} \sim \text{Po}(\mu_{is}E^{D}) \qquad (6.2)$$

$$\log(\mu_{is}) = \eta_s + \omega_s z_{is}\, z_{is} = x_{is} + u_{is}\, x_{is} \qquad {}_{is} \sim N(X_{is}, V_{is})$$

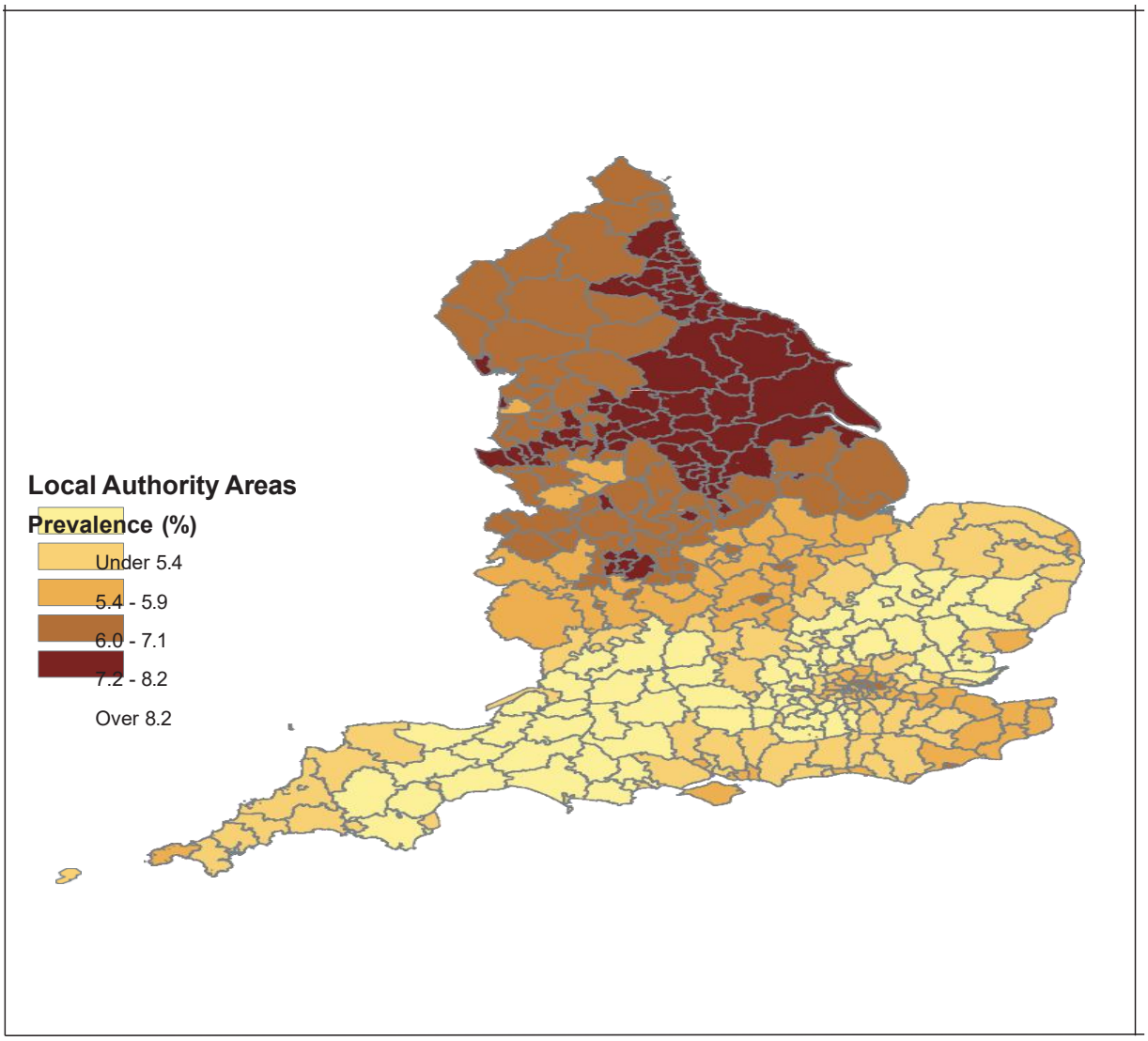


Figure 1: Male prevalence of heart disease

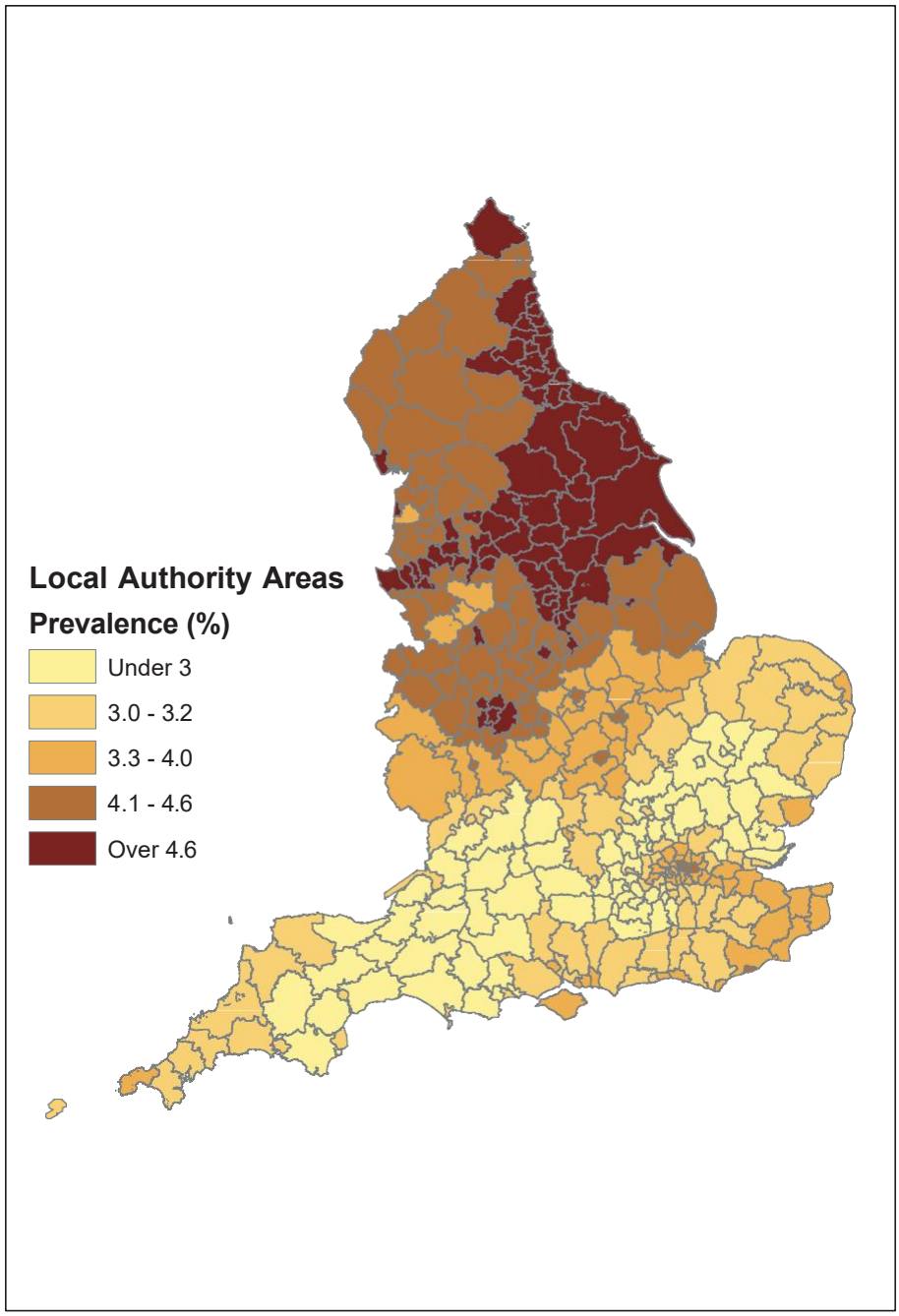


Figure 2: Female prevalence of heart disease

where the effects ui1 and ui2 are centred around their respective means at each iteration, and the spatial correlated errors (ui1, ui2) follow a bivariate conditional autoregressive prior (see Appendix 1). Because many measurement error models involve reverse regression, it is clear from the model's form that z will be impacted by μ and vice versa (Maddala, 2001, Ch 11). N (0, 100) priors are assigned to the intercepts ηs and slopes ωs; the prediction matrix Φ—1 is assumed to follow a Wishart prior with two degrees of freedom and identity scale matrix. The coefficients ω1 and ω2 in model (6.2) are evidently significant in the final 15000 of a two chain run of 25000 iterations, with means (95% intervals) of 0.42 (0.36,0.48) and 0.23 (0.17,0.31). The estimated correlation between (ui1, ui2) is 0.71 (0.54,0.84).

The adjusted prevalences ζis = exp(zis)/(1 + exp(zis)) from the joint model exhibit higher coefficients of variation (i.e., greater inequality) when compared to the mean prevalences πis from the survey model, perhaps because they reflect local variations in mortality relative risks. Quintile maps for the posterior mean prevalence rates ζi1 and ζi2 are shown in Figures 1 and 2, where the North-South contrast is once more evident. The fact that this difference persists even after adjusting for neighbourhood deprivation and ethnodemographic structure implies that environmental factors and health behaviours, such as diet, are also important for variations in prevalence.

The prevalence rates ζis are compared to revascularization rates for the 354 areas in 2002 as an application with policy relevance. There are known differences in the availability of revascularization, such as percutaneous transluminal coronary angioplasty and coronary artery bypass grafts, which other research indicates cannot be accounted for by morbidity (Hippisley-Cox and Pringle, 2000; Payne and Saul, 1997). The correlation between prevalence rates and provision is actually -0.02 for males and 0.08 for females, suggesting that there may be differences in access to surgical care that do not correspond to the need for such care (as indicated by prevalence).

Discussion

There is value in using the survey evidence that is currently available to provide direct estimates of prevalence and morbidity, even though mortality and hospitalisation data are frequently used as proxies for prevalence (morbidity) and thus health need (Ebrahim et al., 2002). A methodology for combining spatially aggregated survey data with data on spatially disaggregated patterns in heart disease mortality that represent regional differences in prevalence has been described in this study (e.g. see Table 3).

This approach can be viewed as a type of meta-analysis across various types of evidence that can be used for different kinds of morbidity. The power prior method can be used to pool data from surveys (in this case, the 1999 and 2003 Health Surveys for England). A different analysis of

The one used in the paper could potentially expand on the idea of evidence pooling by adding more informative priors to the power prior likelihood. More informative priors, particularly on main effects, may be justified by the accumulated epidemiological data on heart disease risk factors. For instance, in the UK, South Asian ethnicity is frequently linked to an increased relative risk of heart disease. For example, according to Greenland (2001, p. 665), one could assume that this group's prior relative risk was between 1 and 3, which would translate into a N (0.55, 0.08) prior. However, using an informative prior approach may also necessitate conducting a sensitivity analysis over various informative priors because there may be comparatively little prior evidence on some interactions (particularly in a specific geographic setting like England with its unique health care system).

Other modelling options could be taken into consideration. Introducing prevalence data from hospital admission records is one choice. Because they reflect the provision of care, these are occasionally dubious as indicators of morbidity; however, they may enhance the estimation of morbidity in cases where hospitalisation is typically inevitable, such as myocardial infarction. In addition to modelling the prevalence of heart disease, one may attempt to jointly model multiple types of prevalence (such as smoking, diabetes, or obesity prevalence) in order to generate indirect area estimates for each. The Health Survey for England tracks the prevalence of these conditions or behaviours in the UK, and they are recognised risk factors for heart disease. Similar opportunities exist for indirectly estimating area prevalence using national health survey data from other nations in conjunction with other pertinent and

locally disaggregated data (e.g. on mortality, hospital admissions).

Appendix. Univariate and Multivariate CAR priors

To explain the form of the CAR prior, first consider a univariate conditional autoregressive prior. Let $(e_1, \ldots, e_n)$ be a vector of effects associated with areas 1,..,n such as relative mortality risks. Then a univariate conditional autoregres-sive prior or CAR prior (Rue and Held, 2005), involves specifying the n full conditionals

$$e_i | e_{[i]} \sim N\Big(\sum_{j=i} c_{ij} e_j / c_{i+},\ \varphi_e / c_{i+}\Big) \qquad (A.1)$$

where $e_{[i]} = (e_1, e_2, \ldots, e_{i-1}, e_{i+1}, \ldots, e_n)$ is the collection of effects excluding area i, $C = [c_{ij}]$ is an n n matrix of spatial interactions c_{ij}, often known but

sometimes involving unknown parameters, the sums

$c_{i+} = \sum_j c_{ij}$ total over rows

in this matrix, and φ_e is a conditional variance. The conditional density uniquely

determines the joint density of the effects $(e_1, \ldots, e_n)$, a feature noted by Besag

and Jin *et al.* (2005). The multivariate normal CAR is the multivariate generalisation of the prior in (A.1). If there are K outcomes, then φ_e is replaced by a $K \times K$ covariance matrix Φ_e. A common practice is to define $c_{ij} = 1$ if areas i and j are adjacent, and $c_{ij} = 0$ otherwise, in which case c_{i+} is the number of areas adjacent to area i.

References

Fitzpatrick, S., Bailey, N., Flint, J., Goodlad, R., Shucksmith, M., and Pryce, G. (2003). Creating a long-term plan for measuring deprivation in Scotland. (http://www.scotland.gov.uk/Publications/2003/09/18197/26538) Scottish Executive

Barnett, S., Martin, D., Diamond, I., and Roderick, P. (2001). A multilevel examination of how social deprivation and rural living affect early-limiting long-term illness.

J. Community Health Epidemiology 55, 44–51.

Bernardinelli, L., Gilks, W., Best, N., and Pascutto, C. (1997). mapping diseases with covariate errors. 16, 741-752; Statistics in Medicine.

J. Besag (1974). Lattice system statistical analysis and spatial interaction.

Royal Statistical Society Journal, Series B 36, 192-236.

Swift, L., Lake, I., Abubakar, I., Catling, L., and Hunter, P. (2005). review of the data supporting the link between water hardness and the prevalence of cardiovascular disease. Contract DWI/70/2/176. Drinking Water Inspectorate, University of East Anglia.

Chen, M., Shao, Q., and Ibrahim, J. (2000). For generalised linear models, power prior distributions. Plan, J. Statist. Inf. 84, 121-137.

Chen, L., Knutsen, S., Shavlik, D., Beeson, W., Petersen, F., Ghamsary, M., and Abbey, D. (2005). Are women more vulnerable to the link between ambient particulate air pollution and fatal coronary heart disease? Environmental Health Perspective 113, 1723-1729.

Ebrahim, S., Oliver, S., Shaw, M., Ho, D., Maxwell, R., and Ben-Shlomo, Y. (2002). Coronary revascularization in the South West, 1991-2000: Gender, age, deprivation, and geographic disparities in the use of PTCA and CABG. South West Public Health Observatory and MRC Health Services Research Collaboration, University of Bristol Department of Social Medicine.

A. Gelfand (1996). model determination through sampling-based techniques. Chapter 9 of Markov Chain Monte Carlo in Practice, edited by Gilks, W., Richardson, S., and Spiegelhalter, D. Chapman and Hall.

Smith, A., and Gelfand, A. (1990). methods for calculating marginal densities based on sampling. American Statistical Association Journal, 85, 398-409.

Stern, H., Rubin, D., Gelman, A., and Carlin, J. (1995). Bayesian analysis of data.

Hall and Chapman.

S. Greenland (2001). combining prior distributions with background data on relative risks. Biometrics 57, 663-670.

S. Greenland (2004). model-based estimation of relative risks and other epidemiologic metrics in case-control and common outcome studies. 160, 301-305; Am. J. Epidemiol.

Poussa, T., Nieminen, M., Hedman, J., and Kaprio, J. (1999). Asthma, aspirin intolerance, nasal polyposis, and chronic obstructive pulmonary disease prevalence. International Journal of Epidemiology 28, 717-722.

Pringle, M., and Hippisley-Cox, J. (2000). Deprivation and the location of primary care services are linked to disparities in access to coronary angiography and revascularization. General Practice Br. J. 50, 449-454.

J. Hoare (2003). Comparing disease morbidity and area-based inequality metrics in England from 1994 to 1998. Quarterly Health Statistics 18, 18–24.

Banerjee, S., Jin, X., and Carlin, B. (2005). hierarchical multivariate CAR models that are generalised for areal data. Biometrics 61, 950-961.

Unit for Joint Health Surveys (2004). England Health Survey, 2003. Key findings are summarised. TSO, London.

G. Maddala (2001). Wiley, Introduction to Econometrics, Third Edition.
Martinez, J., Hamel, D., and Pampalon, R. (2003). Stroke mortality and deprivation in Quebec. Can. Chronic Dis. 24, 57-64.
Donato, F., Zerbini, I., and Monarca, S. (2004). A review of epidemiological research on drinking water hardness and cardiovascular diseases, 1979–2004. The World Health Organisation (http://www.who.int/water sanitation health/en/).
Morris, R., Walker, M., Lampe, F., Whincup, P., Emberson, J., and Shaper, A. (2003). Are the same factors responsible for the North-South Gradients in CHD and stroke in Britain? Stroke 34, 2604-2609.
Noble, M., Smith, G., Wright, G., Dibben, C., Owen, T., Lloyd, M., and Penhale, B. (2000). The Indices of Deprivation 2000 are a tool for measuring multiple deprivation at the small area level. Regeneration Research Summary, DETR, Number 37.
National Statistics Office (2000). General Practice's Key Health Statistics (Series MB6 No.2). Office of National Statistics (2006) (http://www.statistics.gov.uk/statbase/Product.asp?vlnk =4863). Super Output Areas (SOAs) Office of Population Censuses & Surveys (1995) http://www.statistics.gov.uk/geography/soa.asp.
General Practice Morbidity Statistics, Fourth National Study, 1991–1992. (http://www.statistics.gov.uk/statbase/Product.asp?vlnk =616&More=n) Payne, N., and Saul, C. (1997). Coronary artery revascularization rates are compared with the prevalence of angina and coronary mortality in order to examine variations in the use of cardiology services within a health authority. 314, 257-61, Brit. Med. J.
Brookes, M., and Primatesta, P. (2000). Heart disease: Risk factors and prevalence. The Health of Minority Ethnic Groups, Health Survey for England, 1999. Chapter 3 of the Joint Health Surveys Unit for the Department of Health, edited by Erens, B., Primatesta, P., and Prior, G.
Held, L., and Rue, H. (2005). The theory and applications of Gaussian Markov random fields.
Chapman and Hall/CRC Press.
Packham, R., Shaper, A., and Pocock, S. (1980). The British Regional Heart Study: water quality and cardiovascular mortality. J. Environ. Pathol. Toxicol. 4, 89–111.
Deddens, J., Skov, T., and Petersen, M. (1998). Estimation and hypothesis testing of prevalence proportion ratios. International Journal of Epidemiology 27, 91-95.
A. Smith (1998). Where are the crossroads between intercensal population estimates and the American community survey? US Bureau of the Census, Washington, D.C., Population Division Technical Working Paper 31. (http://www.census.gov/population/www/documentation/twps0031/twps0031).Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian metrics for model fit and complexity. J. Royal Statistical Society Series B 64, 583-639.
Thomas, A., Best, N., Spiegelhalter, D., and Lunn, D. (2003). Version 1.4 of the WinBUGS User Manual, January 2003. http://www.mrc-bsu.cam.ac.uk/bugs Strong, M., Radford, J., and Maheswaran, R. (2006). Using information from the new UK general practitioner quality and outcomes framework, a practice-level analysis of socioeconomic deprivation, coronary heart disease prevalence, and care quality was conducted in Rotherham. Public Health Journal, 28, 39–42.
Carlin, B., and Xia, H. (1998). Lung cancer mortality mapping in Ohio using spatiotemporal models with covariate errors. Medical Statistics 17, 2025-2043.
Bertazzi, P., Consonni, D., and Zocchetti, C. (1997). Prevalence rate ratios and odds ratios in cross-sectional studies are related. International Journal of Epidemiology 26, 220-223.