

EDITORIAL

Multi-level modelling, the ecologic fallacy, and hybrid study designs

Jon Wakefield

Accepted 10 March 2009

The ecologic fallacy

In this issue, Robinson's highly influential paper is reprinted,¹ along with a paper advocating the use of multi-level thinking by Subramanian *et al.*², and commentaries by Oakes³ and Firebaugh.⁴

On re-reading Robinson's paper, I was again struck by the clarity of the basic take-home message: ecological data can estimate individual associations in only very rare situations. Robinson illustrated the ecologic fallacy using correlation coefficients applied at different levels of aggregation, whereas more recent work has focused on loglinear models.^{5,6} For common (in a statistical sense) outcomes, such as the illiteracy-race example considered in Robinson's paper, a logistic form is more appropriate (and is used by Subramanian *et al.*) but this form is less amenable to analytic study.⁷ There has been an abundance of work on the myriad causes of ecologic bias on estimates of individual-level associations, which include within-area variability in exposure, and within-area confounding.^{8–13} One might think that the estimation of contextual associations can be carried out with ecologic data alone, but Greenland¹⁴ shows this is not the case. A key point is that contrary to what is claimed by some authors,¹⁵ ecologic data alone do not allow one to determine whether ecological bias is likely to be present in a particular dataset. The only solution to the ecologic fallacy is to supplement the ecologic data with individual-level data^{8,13}, a subject that we now briefly review.

Hybrid designs

Prentice and Sheppard¹⁶ describe a very powerful method for overcoming ecologic bias using within-area covariate samples. When the outcome is not rare, combining data is straightforward.¹³ When the model is linear, one may analytically evaluate which areas

should be sampled in order to maximize information.¹⁷ Work on developing methods for combining ecologic and case-control data has also been carried out.^{18–20} The ecologic data provide power, while the case-control data provide identifiability, and hence overcome ecologic bias, and only very small case-control samples are required. A closely related approach is to use two-phase methods^{21–25} in an ecologic context.²⁶ A number of alternative models have also been proposed.^{27,28}

Multi-level models

Subramanian *et al.*² argue for the use of multi-level models and provide a range of interesting analyses of the Robinson data. Multi-level models are increasing in popularity but I would like to stress the importance of model checking (as does Oakes) and prior choice. Multi-level models are very flexible and allow the specification of complex nested and crossed structures, but as model complexity increases, in tandem increase the number of assumptions that need verification.

The random effects in multi-level models allow dependencies in data to be acknowledged but their use requires care. Outside of a linear-mixed effects model, little theory exists to support the reliability of estimation when violations of assumptions occur. Typically, random effects are assumed to be normally distributed and are required to be independent of any covariates in the model. An important decision is whether the random effect variances depend on covariates, since misspecification can lead to serious bias.²⁹ Numerical investigations of deviations from normality of random effects have been carried out^{30,31} and indicate that for estimation of fixed effects misspecification may not be fatal. Serious bias can arise when the random effects are correlated with covariates, however.³² McCulloch *et al.*³³ provide a recent summary of what is currently known on the effect of misspecification of the random effect distribution. To summarize, it is crucial to assess the

Departments of Statistics and Biostatistics, University of Washington, Box 357232, Seattle, WA 98195-7232, USA.
E-mail: jonno@u.washington.edu

many layers of assumptions upon which reliable inference rests. With respect to prior choice, the specification for the variance components can be very important, as we illustrate shortly.

Exploratory data analysis

If possible, it is preferable to use the raw data to assess the assumptions required for the multi-level models that we are contemplating, before modelling has begun. After a model has been fitted, estimates of random effects reflect both the data and the assumed random effects distribution and so can be difficult to interpret.

To illustrate we calculate the empirical log odds for the data analysed by Subramanian *et al.*, which consist of state-level counts of the number illiterate (cannot read and over 10 years of age), along with denominators, for each of the 49 states and three races (native-born white, foreign-born white, black). These data provide a relatively easy case study for model assessment since the counts are large, and the models envisaged are relatively simple. For each of the states, we also have a binary indicator that denotes whether Jim Crow laws were present in that state. These laws enforced racial segregation in all public places.

In Figure 1, we plot the log odds of illiteracy by race with points to the left corresponding to the 27 states without Jim Crow laws and those to the right those 22 states with Jim Crow laws. Within each collection of states the points are ordered by increasing overall illiteracy rate. We clearly see the effect of Jim Crow laws on native-born whites and blacks, and large between-state variability in illiteracy is also evident. We calculate summaries of these log odds by race and by Jim Crow status; the row labelled 'Analysis 1' in Table 1 gives the results. In both states with and without Jim Crow laws, the odds of illiteracy for foreign-born whites are roughly 15 times those of native-born whites in states without Jim Crow laws. Blacks in non-Jim Crow law states have odds that are eight times those of native-born whites, whereas relative to this category the odd ratio for blacks in Jim Crow states is 27. In Table 2, we present the standard deviations for the log odds for each of the 2×3 combinations of absence/presence of Jim Crow laws and race with 'Analysis 1' giving the empirical standard deviations. We see that for foreign-born whites the spread is roughly the same in both types of states, while for the other two races the spread is greater in Jim Crow states.

In Figure 2, we provide normal QQ plots of the log odds and scatterplots of the pairs of log odds by state. In the usual implementation of multi-level models, these log odds are modelled as random effects and assumed to be normally distributed. Panels (a)–(c) give normal QQ plots for each race (after standardization). Points close to the line indicate normality, and nothing appears terribly amiss here. As an aside, we

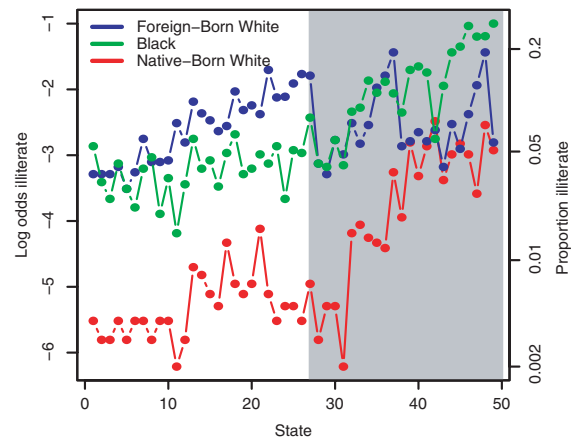


Figure 1 Log odds of illiteracy by state and race, ordered by increasing proportion illiterate within states without Jim Crow laws (background white), and within states with Jim Crow laws (background grey)

Table 1 Odds ratio summaries from three analyses, by Jim Crow status and race

		Race		
		NBW	FBW	B
Non-Jim Crow	Analysis 1	Ref	15.3	8.1
	Analysis 2	Ref	14.5	7.9
	Analysis 3	Ref	15.0	8.2
Jim Crow	Analysis 1	4.5	15.8	27.4
	Analysis 2	4.3	14.5	25.8
	Analysis 3	4.8	16.7	30.0

NBW = native-born white; FBW = foreign-born whites; B = black. Analysis 1 is based on the empirical log odds. Analysis 2 is based on a multi-level model that assumes that the random effect distributions are identical for non-Jim Crow and Jim Crow states. Analysis 3 is based on a multi-level model with distinct random effect distributions for each of non-Jim Crow and Jim Crow states.

Table 2 Standard deviation of log odds from three analyses, by Jim Crow status and race

		Race		
		NBW	FBW	B
Non-Jim Crow	Analysis 1	0.48	0.55	0.40
	Analysis 2	0.81	0.54	0.54
	Analysis 3	0.48	0.56	0.37
Jim Crow	Analysis 1	1.08	0.53	0.68
	Analysis 2	0.81	0.54	0.54
	Analysis 3	1.09	0.53	0.69

NBW = native-born white; FBW = foreign-born whites; B = black. Analysis 1 is based on the empirical log odds. Analysis 2 is based on a multi-level model that assumes that the random effect distributions are identical for non-Jim Crow and Jim Crow states. Analysis 3 is based on a multi-level model with distinct random effect distributions for each of non-Jim Crow and Jim Crow states.

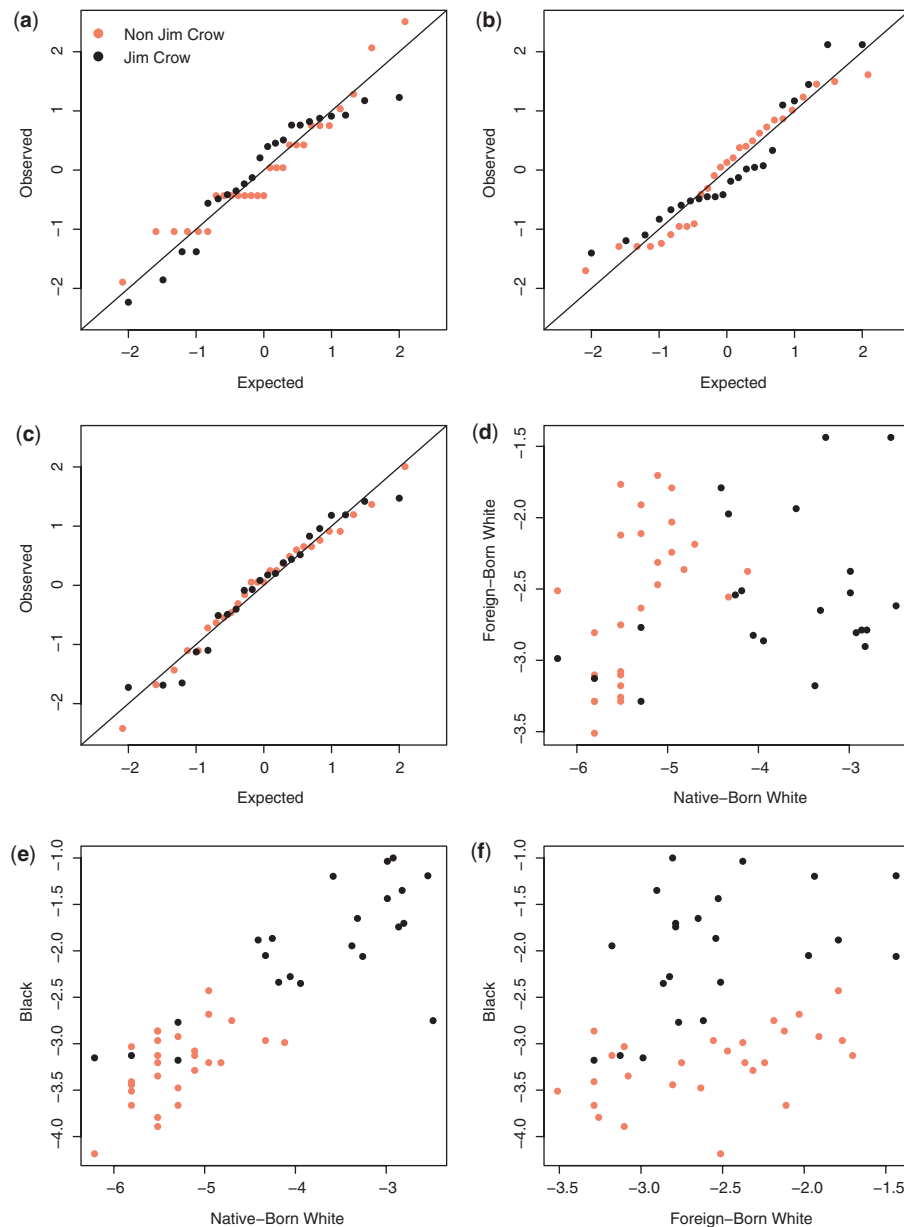


Figure 2 Log odds of illiteracy by Jim Crow status and race. Panels (a)–(c) give normal QQ plots, for native-born whites, foreign-born whites and blacks, respectively (the log odds are standardized). Bivariate scatterplots for: (d) foreign-born white versus native-born white, (e) black versus native-born whites, (f) black versus native-born white. Orange/black points correspond to states without/with Jim Crow laws in place

note the collections of horizontal dots that are apparent, particularly in panel (a) for states without Jim Crow laws. On closer examination of the data, we see that many of the population totals for native-born whites end with '000', '333' or '667', which suggests that some form of rounding has been carried out in their calculation. Panels (d)–(f) show the bivariate relationships. Panel (e) shows that the log odds are correlated within states for native-born whites and blacks, suggesting there are unmeasured state-level variables that influence illiteracy rates for these two races.

Prior choice

We first describe issues pertinent to prior choice using a multi-level model, which is identical to model 4 of Subramanian *et al.*

Stage 1: observed data model

Let Y_{i1} , Y_{i2} , Y_{i3} denote the number of native-born white, foreign-born white and black individuals who are illiterate in state i , N_{i1} , N_{i2} , N_{i3} the respective denominators, and $x_i = 0/1$ a state-level indicator for the absence/presence of Jim Crow laws. We define

p_{ij} to be the proportion of illiterate individuals in the population of race j in state i , $i = 1, \dots, 49$, $j = 1, 2, 3$.

Then $Y_{ij} | p_{ij} \sim_{ind} \text{Binomial}(N_{ij}, p_{ij})$ with

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_j + \gamma_j x_i + b_{ij}$$

Stage 2: random effects model

We assume

$$\mathbf{b}_i = \begin{bmatrix} b_{i1} \\ b_{i2} \\ b_{i3} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}\right).$$

We write this latter distribution as $\mathbf{b}_i | \Sigma \sim_{iid} N(\mathbf{0}, \Sigma)$ so that an implicit assumption is that the random effects distribution does not depend upon whether Jim Crow laws are present in state i (which does not look reasonable from the results of Table 2). This model therefore allows the race-specific random effects to be correlated across areas. In a Bayesian analysis, we need to specify a prior distribution for the variance-covariance matrix of the random effects which is a tricky exercise since a variance-covariance matrix needs to be positive definite.

Stage 3: hyperpriors

We assume independent priors:

$$\pi(\beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3, \Sigma) = \pi(\beta_1) \times \pi(\beta_2) \times \pi(\beta_3) \times \pi(\gamma_1) \\ \times \pi(\gamma_2) \times \pi(\gamma_3) \times \pi(\Sigma)$$

with β_j and γ_j , $j = 1, \dots, 3$, assigned flat (improper) priors. One choice of prior over positive definite matrices is the Wishart distribution, which is a multivariate generalization of the gamma distribution. Specifically, we assume that $\Sigma^{-1} \sim \text{Wishart}(r, \mathbf{S})$, where r is a degrees of freedom and \mathbf{S} is a scale matrix (we parameterize the matrix so that $E[\Sigma^{-1}] = r\mathbf{S}$). The specification of these two parameters requires great care.

Subramanian *et al.* report the use of ‘diffuse’ priors in MLwiN. In addition to the above model, they fit a model with independent random effects, i.e. $b_{ij} | \sigma_j^2 \sim_{ind} N(0, \sigma_j^2)$ with $\sigma_{aj}^{-2} \sim \text{Gamma}(a, b)$ and $a = b = 0.001$. The latter is often used but can be influential for some datasets and should be avoided. As pointed out elsewhere,³⁴ under this prior, 99% of the prior mass for σ_j (the standard deviation of the random effects—these are the values that appear in Table 2) lies to the right of 6.4, which is clearly a ridiculous prior. We follow a previously suggested procedure,³⁵ the details of which are in Appendix 1, and give a range for $\exp(b_{ij})$, the residual odds ratio for race j in state i , i.e. the odds of being illiterate compared with the median of the distribution of odds across states for race j . For example, for a range

of residual odds of $[\exp(0.1), \exp(10)]$ we obtain $a = 0.5, b = 0.0164$. This gives 2.5, 50 and 97.5% quantiles for σ_j of (0.08, 0.26 and 5.76), which are far more prudent. For the more general dependent case with 3×3 variance-covariance matrix Σ , Subramanian *et al.* choose the default in MLwiN which, according to the manual, is a data-dependent prior in which the matrix \mathbf{S} is chosen based on the data, which though often not fatal, is not strictly legal since it is using the data twice. Appendix 1 contains details of how r and \mathbf{S} may be chosen in this more general case. We again take the range of residual odds to be $[\exp(0.1), \log(10)]$ with correlations of 0 (an alternative would be to pick a correlation > 0 to reflect the belief of shared unmeasured predictors of illiteracy across all races), along with an integer value of r which, in one sense, gives the most conservative prior (the variance of the prior decreases with increasing r , so a lower value of r gives larger variance). These choices give $r = 3$ and \mathbf{S} a diagonal matrix with diagonal entries $S_{jj} = 30.45$, $j = 1, 2, 3$.

Hence, we see that prior specification is not straightforward, but is important since it may influence the results, particularly when the number of units (here, states) is small. As a minimum, priors should be clearly specified, along with the estimation method used. Ideally data and code should be made available on an author’s web site, or as Supplementary Material.

Interpretation

The interpretation of parameters in multi-level models requires great care. To illustrate, consider the model for native-born whites only:

$$\log\left(\frac{p_{i1}}{1-p_{i1}}\right) = \beta_1 + \gamma_1 x_i + b_{i1},$$

with $b_{i1} \sim_{iid} N(0, \sigma_1^2)$. This model implies that the odds of exposure are log-normally distributed:

$$\frac{p_{i1}}{1-p_{i1}} = \exp(\beta_1 + \gamma_1 x_i + b_{i1}) \\ \sim \text{Log Normal}\{\beta_1 + \gamma_1 x_i, \sigma_1^2\}$$

For this model:

- $\exp(\beta_1 + b_{i1})$ is the odds of illiteracy for native-born whites in state i ;
- $\exp(\beta_1)$ is the median illiteracy for native-born whites for states without Jim Crow laws. An alternative definition is the odds of illiteracy for native-born whites in a ‘typical’ state without Jim Crow laws; ‘typical’ here means $b_{i1} = 0$;
- $\exp(\beta_1 + \gamma_1)$ is the median odds of illiteracy for native-born whites in states with Jim Crow laws;
- $\exp(\beta_1 + \sigma_1^2/2) = E[\exp(\beta_1 + b_{i1})]$ is the average odds of illiteracy for native-born whites across states without Jim Crow laws, and

- $\exp(\beta_1 \pm 1.96 \times \sigma_1)$ is a 95% interval for the odds of illiteracy for native-born whites across states without Jim Crow laws. Hence, this function gives an indication of the variability in the odds of illiteracy across states without Jim Crow laws.

In Table 1, Analyses 2 and 3, we report the median odds ratios. For native-born whites in states without Jim Crow laws the median odds of illiteracy is estimated as 0.005. A 95% interval for these odds across states without Jim Crow laws is estimated as (0.002, 0.0123), so that the odds of illiteracy range between 1 in 500 and 1 in 80.

In Table 1, we report analyses with a common random effects distribution across all states ('Analysis 2' and model 4 of Subramanian *et al.*) and a model with distinct random effects for each category of states (absence/presence of Jim Crow laws); this model ('Analysis 3') is equivalent to separate fits to each category of states. We see that this final analysis is more appropriate, and essentially recovers the point estimates from the analysis with the empirical log odds, since here the denominators are large. Since the random effects are so well estimated, and the number of states is not small, the particular Wishart specification assumed here is not influential.

We might ask what multi-level Analysis 3 has added here, when compared the analysis based on the raw empirical log odds, since the point estimates are virtually identical. However, correct standard errors for odds ratios require an explicit model for the dependence between the counts within each state, and this is provided by the multi-level approach, but not by the empirical log odds calculations.

R and WinBUGS code for the analyses reported here are available from <http://faculty.washington.edu/jonno.cv.html>

Markov chain Monte Carlo

For these data, the Markov chain was very poorly behaved, as shown in Figure 3, which displays the time course of two chains for the log odds of illiteracy for native-born whites in states without Jim Crow laws. The chains were started from 'good' and 'poor' starting points. The good chain had initial values for the fixed effects set at the maximum likelihood estimates (MLEs), while the bad chain had fixed effects set to zero. The bad chain does not 'mix' with the good chain until around 300K iterations have elapsed. Recently,³⁶ a new approximation strategy for Bayesian inference has been described, and is ideally suited to data such as that considered here. The approach is considerably faster than Markov chain Monte Carlo (MCMC), and R code is available.

Diagnosis of convergence in an MCMC context can be very difficult and is a black art, but various checks are available. For these data, we fit the model (with the same first two stages as the Bayesian model

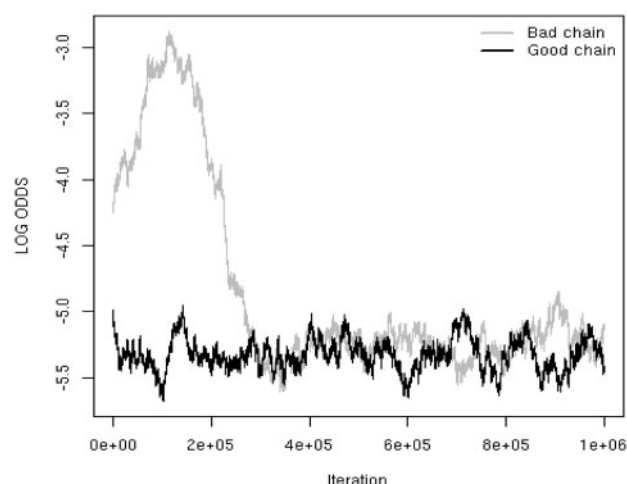


Figure 3 Time course of two Markov chains for the log odds of illiteracy for blacks

Table 3 Comparison of likelihood and Bayesian estimation techniques

Parameter	Likelihood	Bayes
	Estimate (SE)	Estimate (SD)
β_1	-5.32 (0.15)	-5.25 (0.16)
β_2	-2.58 (0.10)	-2.57 (0.10)
β_3	-3.21 (0.10)	-3.19 (0.10)
γ_1	1.52 (0.23)	1.43 (0.25)
γ_2	0.03 (0.15)	0.02 (0.16)
γ_3	1.22 (0.15)	1.18 (0.17)
σ_1	0.79 (—)	0.81 (0.09)
σ_2	0.53 (—)	0.54 (0.06)
σ_3	0.52 (—)	0.54 (0.06)

For the likelihood summaries, we report the MLEs and the asymptotic standard errors, whereas for the Bayesian analysis we report the mean and standard deviation of the posterior distribution.

described above) using maximum likelihood, and then compare the results with the Bayesian analysis. For such abundant data, one would not expect too many differences, and this is confirmed in Table 3 for the model in which a common random effects distribution across all states is assumed.

We end with the usual caveats concerning the analysis of observational data; clearly, we are far from being able to make causal statements for these data since the list of potential confounders is vast. Multi-level models are a useful way of structuring analyses, but their use requires care, and they cannot control for confounding.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

National Institute of Health (grant RO1 CAO95994).

Acknowledgement

I would like to thank Prof Subramanian for supplying the data, and swiftly responding to queries concerning his paper.

Conflict of interest: None declared.

References

- ¹ Robinson WS. Ecological correlations and the behaviour of individuals. *Ame Sociol Rev* 2009;**15**:351–57 (reprinted in *Int J Epidemiol* 2009;**38**:337–41).
- ² Subramanian SV, Jones K, Kaddour A, Krieger N. Revisiting Robinson: the perils of individualistic and ecologic fallacy. *Int J Epidemiol* 2009;**38**:342–60.
- ³ Oakes JM. Commentary: individual, ecological and multi-level fallacies. *Int J Epidemiol* 2009;**38**:361–68.
- ⁴ Firebaugh G. Commentary: 'is the social world flat? W.S. Robinson and the ecologic fallacy'. *Int J Epidemiol* 2009;**38**:368–70.
- ⁵ Richardson S, Stucker I, Hemon D. Comparison of relative risks obtained in ecological and individual studies: some methodological considerations. *Int J Epidemiol* 1987;**16**:111–20.
- ⁶ Wakefield JC, Salway RE. A statistical framework for ecological and aggregate studies. *J Royal Stat Soc [Ser A]* 2001;**164**:119–37.
- ⁷ Salway RA, Wakefield JC. Sources of bias in ecological studies of non-rare events. *Environ Ecol Stat* 2005;**12**:321–47.
- ⁸ Wakefield J. Ecologic studies revisited. *Ann Rev Public Health* 2008;**29**:75–90.
- ⁹ Greenland S, Morgenstern H. Ecological bias, confounding and effect modification. *Int J Epidemiol* 1989;**18**:269–74.
- ¹⁰ Greenland S, Robins J. Ecological studies: biases, misconceptions and counterexamples. *Am J Epidemiol* 1994;**139**:747–60.
- ¹¹ Wakefield JC. Sensitivity analyses for ecological regression. *Biometrics* 2003;**59**:9–17.
- ¹² Piantadosi S, Byar DP, Green SB. The ecological fallacy. *Am J Epidemiol* 1988;**127**:893–904.
- ¹³ Wakefield JC. Ecological inference for 2 x 2 tables (with discussion). *J Royal Stat Soc [Ser A]* 2004;**167**:385–445.
- ¹⁴ Greenland S. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *Int J Epidemiol* 2001;**30**:1343–50.
- ¹⁵ King G. *A Solution to the Ecological Inference Problem*. Princeton: Princeton University Press, 1997.
- ¹⁶ Prentice RL, Sheppard L. Aggregate data studies of disease risk factors. *Biometrika* 1995;**82**:113–25.
- ¹⁷ Glynn A, Wakefield J, Handcock M, Richardson T. Alleviating linear ecological bias and optimal design with subsample data. *J Royal Stat Soc [Ser A]* 2008;**71**:179–202.
- ¹⁸ Haneuse S, Wakefield J. The combination of ecological and case-control data. *J Royal Stat Soc [Ser B]* 2008;**70**:73–93.
- ¹⁹ Haneuse S, Wakefield J. Geographic-based ecological correlation studies using supplemental case-control data. *Stat Med* 2008;**27**:864–87.
- ²⁰ Haneuse S, Wakefield J. Hierarchical models for combining ecological and case-control data. *Biometrics* 2007;**63**:128–36.
- ²¹ Cain KC, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. *Am J Epidemiol* 1988;**128**:1198–206.
- ²² Breslow NE, Holubkov R. Maximum likelihood estimation of logistic regression parameters under two-phase, outcome-dependent sampling. *J Royal Stat Soc [Ser B]* 1997;**59**:447–61.
- ²³ Breslow NE, Holubkov R. Weighted likelihood, pseudo likelihood and maximum likelihood methods for logistic regression analysis of two-stage data. *Stat Med* 1997;**16**:103–16.
- ²⁴ Breslow NE, Chatterjee N. Design and analysis of two-phase studies with binary outcome applied to Wilms tumour prognosis. *Appl Stat* 1999;**48**:457–68.
- ²⁵ Scott AJ, Wild CJ. Fitting regression models to case-control data by maximum likelihood. *Biometrika* 1997;**51**:54–71.
- ²⁶ Wakefield J, Haneuse S. Overcoming ecological bias using the two-phase study design. *Am J Epidemiol* 2008;**167**:908–16.
- ²⁷ Jackson CH, Best NG, Richardson S. Improving ecological inference using individual-level data. *Stat Med* 2006;**25**:2136–59.
- ²⁸ Jackson C, Best N, Richardson S. Hierarchical related regression for combining aggregate and individual data in studies of socio-economic disease risk factors. *J Royal Stat Soc [Ser A]* 2008;**171**:159–78.
- ²⁹ Heagerty PJ, Kurland BF. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* 2001;**88**:973–85.
- ³⁰ Neuhaus JM, Hauck WW, Kalbfleisch JD. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika* 1992;**79**:755–62.
- ³¹ Neuhaus JM. Estimation efficiency with omitted covariates in generalised linear models. *J Am Stat Assoc* 1998;**93**:1124–29.
- ³² Neuhaus JM, McCulloch CE. Separating between and within-cluster covariate effects using conditional and partitioning methods. *J Royal Stat Soc [Ser B]* 2006;**68**:859–72.
- ³³ McCulloch CE, Searle SR, Neuhaus JM. *Generalized, Linear, and Mixed Models*. 2nd edn. New York: John Wiley and Sons, 2008.
- ³⁴ Kelsall JE, Wakefield JC. Discussion of 'Bayesian models for spatially correlated disease and exposure data', by Best et al. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). *Bayesian Statistics 6*. Oxford: Oxford University Press, 1999. p. 151.
- ³⁵ Wakefield JC. Disease mapping and spatial regression with count data. *Biostatistics* 2007;**8**:158–83.
- ³⁶ Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent gaussian models using integrated

nested laplace approximations (with discussion). *J Royal Stat Soc [Ser B]* 2009.

³⁷ Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2006;**1**:515–34.

³⁸ Gustafson P, Hossain S, MacNab YC. Conservative prior distributions for variance parameters in hierarchical models. *Canadian J Stat* 2006;**34**:377–90.

Appendix 1

Prior choice

We begin with the independent random effects model with

$$\begin{aligned} b_{ij}|\sigma_j^2 &\sim_{ind} N(0, \sigma_j^2) \\ \sigma_j^{-2} &\sim \text{Gamma}(a, b) \end{aligned}$$

for $j = 1, 2, 3$. We can average over σ_j^2 to obtain the marginal distribution of $\exp(b_{ij})$, the residual odds, which is a more interpretable quantity and describes the odds for race j in state i , relative to the median odds across states. The marginal distribution for $\exp(b_{ij})$ is a log Student's t -distribution with $d=2a$ degrees of freedom, location zero and scale $\Sigma=b/a$. We choose $a=0.5$ so that the marginal distribution

is a Cauchy distribution, and then choose b so that 95% of the residual odds lie within a range that we specify. In particular, for the range $(1/R, R)$ we use the relationship $\pm t_{0.025}^d \sqrt{\Sigma} = \pm \log R$, where t_r^d is the $100 \times r$ -th quantile of a Student's t random variable with d degrees of freedom, to give $a = d/2$, $b = (\log R)^2 d/2 (t_{1-(1-q)/2}^d)^2$. For example, for a range of $[\exp(0.1), \exp(10)]$, we obtain $b = 0.0164$.

For the case of p dependent random effects, $\mathbf{b}_i = (b_{i1}, \dots, b_{ip})$, we have

$$\begin{aligned} \mathbf{b}_i|\Sigma &\sim_{ind} N_p(\mathbf{0}, \Sigma) \\ \Sigma^{-1} &\sim \text{Wishart}_p(r, \mathbf{S}) \end{aligned}$$

which, on marginalization over Σ , gives \mathbf{b}_i as multivariate Student's t with location $\mathbf{0}$, scale matrix $[(r-p+1)\mathbf{S}]^{-1}$ and degrees of freedom $d = r - p + 1$. The margins of a multivariate Student's t are t also, which allows r and \mathbf{S} to be chosen as in the univariate case. Specifically, the j -th element, b_{ij} follows a univariate student t distribution with location 0, scale $S^{jj}/(r-p+1)$, and degrees of freedom $d = r - p + 1$ where S^{jj} is element (j, j) of the inverse of \mathbf{S} .

For other approaches to prior choice in hierarchical models see Gelman³⁷ and Gustafson *et al.*³⁸