APPENDIX B

In hierarchical Bayesian modeling, the parameters at the highest level of the model (β and ζ. in equation 8) are known as hyper-parameters and their prior distribution (π) is known as a hyper-prior. A common two step modeling strategy is to first fit a model with a non-informative hyper-prior, and second replace the non-informative hyper-prior with an informative hyper-prior. When fitting a hierarchical Poisson regression model, the first step can be carried out in R using code written by Christiansen and Morris (PRIMM - (1)). Here we show how to carry out the second step in WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs).

In WinBUGS, Bayesian models are fit using Markov chain Monte Carlo methods with Gibbs sampling the default method (2). With a hierarchical model, Gibbs sampling alternates between sampling first from the conditional posterior distribution of the first level parameters given the data and the hyper-parameters, and then from the conditional posterior distribution of the hyper-parameters given the data and sampled values of the first level parameters (3). After repeating this cycle many times, the sampling process converges to the joint posterior distribution for both first level and hyper- parameters and subsequent cycles produce a Monte Carlo sample from this distribution. Posterior summaries such as the posterior mean and credible interval can be approximated by the corresponding quantities of this Monte Carlo sample.

WINBUGS WITH A NON-INFORMATIVE PRIOR

We first re-fit Christiansen and Morris' hierarchical Poisson regression model with our prior covariate structure to the data in Appendix A using WinBUGS (Figure B1). Our prior structure implies a log-linear model for the expected mortality rate with an intercept (β_1) and terms for age (β_2), sex (β_3), Maori ethnicity (β_4), Pacific Island ethnicity (β_5), educational score (β_6) and interactions terms for age and ethnicity (β_7 and β_8), education and ethnicity (β_9 and β_{10}), and age and education (β_{11}). The data are coded as in Appendix A, except that both age and educational score are centered at values of 50. Centering covariates re-defines the baseline subgroup and a sensible choice of baseline subgroup makes it easier to elicit prior distributions for covariate parameters. Centering covariates can also reduce correlation between covariate parameters so that the Gibbs sampling process converges to the joint posterior distribution in fewer cycles.

The non-informative prior in Christiansen and Morris' model consists of a flat uniform distribution for each β parameter and a 'uniform shrinkage prior' for ζ and these distributions are assumed independent. In WinBUGS, each flat uniform prior is represented by a normal distribution with zero mean and infinite variance. As before, we set d_0, the median of our prior distribution for ζ, at 10 as a measure of confidence in our prior structure.

Fitting the model also requires initial values with which to start the Gibbs sampling process. We use an R add-on package, R2WinBUGS (http://www.stat.columbia.edu/~gelman/bugsR) (4), to pass initial values and data to WinBUGS and then to process the resulting Monte Carlo sample. We run 7 parallel Markov chains in WinBUGS, each of length 11,000 cycles. We discard the first 1000 cycles of each chain, to allow the sampling process to converge to the joint posterior distribution. Gelman-Rubin convergence statistics close to one for each parameter indicate that the chains probably converge to the posterior during the first 1000 cycles (3). To reduce autocorrelation in the posterior sample we then thin the chains, retaining only every 7^{th} value. This gives a posterior sample of 10,000.

WINBUGS WITH AN INFORMATIVE PRIOR

An informative hyper-prior could reflect background knowledge about covariates such as that women have a lower rate of mortality than men, and that the rate of mortality increase with age and with lower socio-economic status (5). In addition, research in New Zealand has consistently found a higher rate of mortality in both Maori and Pacific Island minorities than in the nMnPI majority (6).

Prior information typically comes from studies where rate ratios are reported. So it is convenient to give prior information in terms of percentiles for a distribution of rate ratios and then transform this information to an appropriate distribution for β. For convenience we transform percentile information to independent normal distributions for each of the β parameters, except we transform to a log normal distribution for β_2 so that the parameter for age is always positive.

Our baseline subgroup, to which all other subgroups are compared, is the 50 year old male from the nMnPI majority with a median level of educational achievement. This baseline group is determined by the way data are coded in Appendix A and by centering both age and educational score at 50. Mortality rates are available from previous censuses subdivided by ethnicity, age and sex but not further subdivided by educational achievement. We consider that the mortality rate in our baseline group will be between 400 and 700 deaths per annum per 100,000 with 80% confidence, and that rates outside this interval are just as likely to be high as low. By reversing the usual steps for interval estimation, this interval $[L_k, U_k]$ transforms to a normal distribution for β_k with mean μ_k and standard deviation τ_k (7):

$$\mu_k = \ln\left(\left(L_k . U_k\right)^{1/2}\right), \tag{B1}$$

$$\tau_k = \ln\left(U_k / L_k\right) / 2z, \tag{B2}$$

where z is the standard normal deviate for a two sided confidence interval (ie 1.28, 1.64 or 1.96 for 80%, 90% or 95% confidence respectively). In the same way, we specify rate ratio intervals for female, Maori and Pacific Island subgroups and per 10 unit increase in the educational score and transform these intervals to appropriate normal priors (Table B1).

For age, we consider a negative effect or no effect as biologically implausible. We therefore represent our prior expectation with a log normal distribution. We set our prior median rate ratio at a 7% increase. An increase of 10% or more per year of age implies a doubling or more of the mortality rate with every 10 years of age and we consider this unlikely with a probability of only 0.10. These considerations lead to a normal prior for log β_2 with mean $\mu_2 = \ln(\ln(1.07))$ and standard deviation $\tau_2 = [\ln(\ln(1.1)) - \mu_2]/z$ where z = 1.28. These prior considerations could be improved by carefully studying life tables.

Parameters associated with covariate interactions have a more complex interpretation. These β parameters represent the extent to which a rate ratio for one variable is modified by a second variable, when all other variables are set to zero. Our prior expectation was that these interactions would not be large but otherwise we were quite uncertain about the magnitude or direction of these effects. Consequently we specify standard normal priors for all interaction hyper-parameters. This implies a prior probability of 0.025 for an interaction rate ratio either above 7.1 or below 0.14. Interaction effects of this magnitude would be surprising.

We fit a model with an informative prior by replacing the last section of code in Figure B1 (the k loop) with the code in Figure B2. We then repeat the Gibbs sampling process described in the previous section.

RESULTS

With a non-informative prior, contour plots of the posterior mortality rate appear identical whether the model is fit in PRIMM (Figure 5 in the main text) or in WinBUGS (Figure B3). Even the informative prior leads to apparently identical plots (Figure B4). The informative prior has little influence on estimates of the effect of education on mortality (Table B2). This suggests that the data are still far more informative than this prior. Certainly the informative prior does not weaken our view that the protective effect of education differs between ethnic groups.

REFERENCES

1. Christiansen CL, Morris CN. Hierarchical Poisson regression modeling. J Am Stat Assoc 1997;92:618-632.
2. Spiegelhalter DJ, Thomas A, Best NG, Lunn D. WinBUGS user manual version 1.4. Cambridge: Medical Research Council Biostatistics Unit, 2003.
3. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. London: Chapman and Hall, 1995.
4. Sturtz S, Ligges U, Gelman A. R2WinBUGS: a package for running WinBUGS from R. J Stat Softw 2005;12.
5. Blakely T, Fawcett J, Atkinson J, Tobias M, Cheung J. Decades of disparity II: socioeconomic mortality trends in New Zealand 1981-1999. Wellington, New Zealand: Ministry of Health and University of Otago, 2005.

6. Ajwani S, Blakely T, Robson B, Tobias M, Bonne M. Decades of disparity: ethnic mortality trends in New Zealand 1980-1999. Wellington, New Zealand: Ministry of Health and University of Otago, 2003.
7. Greenland S. Putting background information about relative risks into conjugate prior distributions. Biometrics 2001;57:663-670.

TABLE B1. Prior assumptions and implied prior distributions for covariate hyper-parameters.

Hyper-parameter	Prior Assumption*	Implied prior*
Intercept	p(4.E-3 < exp(β_1) < 7.E-3) = 0.80	β_1 ~ normal [-5.24, 0.22^2]
Age (per year)	median exp(exp(β_2)) = 1.07 and p(exp(exp(β_2)) > 1.10) = 0.10	log β_2 ~ normal [-2.69, 0.27^2]
Female	p(0.50 < exp(β_3) < 0.75) = 0.80	β_3 ~ normal [-0.49, 0.16^2]
Maori	p(2 < exp(β_4) < 3) = 0.80	β_4 ~ normal [0.90, 0.16^2]
Pacific Island	p(1.5 < exp(β_5) < 2) = 0.80	β_5 ~ normal [0.55, 0.11^2]
Educational score	p(0.80 < exp(10.β_6) < 0.95) = 0.60	β_6 ~ normal [-0.014, 0.010^2]
Interactions	Uncertain but unlikely to be large	β_7, β_8, β_9, β_{10}, and β_{11} ~ normal [0, 1]

* p( ), probability; x ~ D[a,b], x distributed D with mean a and variance b.

TABLE B2. Mortality rate for male and female 50 year olds with an educational score of zero as a multiple of their mortality rate with a score of 100, New Zealand, 1996-1999: comparison of results for models fit in PRIMM and in WinBUGS.

	PRIMM		WinBUGS			
			Non-informative prior		Informative prior	
Ethnic group	MRR*	95% CI*	MRR*	95% CI*	MRR*	95% CI*
nMnPI*	2.35	1.95, 2.83	2.35	1.94, 2.81	2.38	1.97, 2.88
Maori	1.54	1.19, 2.00	1.54	1.19, 2.02	1.56	1.20, 2.03
Pacific Island	1.37	0.96, 1.95	1.37	0.97, 1.97	1.39	0.98, 2.00

* nMnPI, the non-Maori non-Pacific Island majority; MRR, mortality rate ratio; CI, credible or confidence interval.

FIGURE B1. WinBUGS code for fitting Christiansen and Morris' hierarchical Poisson regression model with a non-informative prior.

FIGURE B2. WinBUGS code for our informative prior for covariate hyper-parameters.

FIGURE B3. Posterior point estimate contours (as in Figure 5 in the main text) for the WinBUGS model with the non-informative prior, New Zealand, 1996-1999 (nMnPI – non-Maori non-Pacific Island majority).

FIGURE B4. Posterior point estimate contours (as in Figure 5 in the main text) for the WinBUGS model with the informative prior, New Zealand, 1996-1999 (nMnPI – non-Maori non-Pacific Island majority).

```
# WinBUGS code for hierarchical Poisson regression model with non-
# informative prior (Christiansen and Morris JASA 1997 92:618-632).
#
# Using the variable names and coding of the data in Appendix A:
# I = 240 is the number of cells; Deaths = Mortality x Risk is
# the number of deaths in a cell; Age and Educational score are both
# centered at 50 so that AgeCen = Age + 2.5 – 50 and EdCen = Ed – 50.
# Note WinBUGS parameterizes the gamma distribution (a,b) such that
# the gamma mean is a/b and variance is a/b**2.

model {
     for (i in 1:I){
          Deaths[i] ~ dpois(theta[i])
          theta[i] <- lambda[i]*Risk[i]
          lambda[i] ~ dgamma(a[i],b[i])
          a[i] <- zeta
          b[i] <- zeta/mu[i]
          mu[i] <- exp( beta[1] + beta[2]*AgeCen[i] + beta[3]*F[i] +
               beta[4]*M[i] + beta[5]*P[i] + beta[6]*EdCen[i] +
               beta[7]*AgeCen[i]*M[i] + beta[8]*AgeCen[i]*P[i] +
               beta[9]*EdCen[i]*M[i] + beta[10]*EdCen[i]*P[i] +
               beta[11]*AgeEd[i] )
          AgeEd[i] <- AgeCen[i]*EdCen[i]
          }

# Specify the uniform shrinkage prior for zeta via a uniform prior
# for shrink0 and then use the relationship between shrink0 and zeta
# (equation 11) to obtain zeta. Median for zeta set to 10.

          shrink0 ~ dunif(0, 1)
          zeta <- 10*shrink0 / (1-shrink0)

# Specify uninformative priors for the beta hyper-parameters using
# normal distributions with large variance. Note WinBUGS parameterizes
# the normal distribution as a mean and precision where the precision
# is the reciprocal of the variance.

          for ( k in 1:11) {
               beta[k] ~ dnorm(0.0, 1.0E-6)
               }
     }


#
```

FIGURE B1. WinBUGS code for fitting Christiansen and Morris' hierarchical Poisson regression model with a non-informative prior.

```
# Informative prior given in Table B1 for the beta hyper-parameters.

           beta[1] ~ dnorm(-5.24,21.0)      #intercept
           beta[2] ~ dlnorm(-2.69,14.0)     #age
           beta[3] ~ dnorm(-0.49,40.0)      #female
           beta[4] ~ dnorm(0.90,40.0)       #Maori
           beta[5] ~ dnorm(0.55,79.4)       #Pacific Island
           beta[6] ~ dnorm(-0.014,9590)     #educational score

# Use standard normal prior for all interactions.

           for ( k in 7:11) {
                beta[k] ~ dnorm(0.0, 1.0)
                }
     }


#
```

FIGURE B2. WinBUGS code for our informative prior for covariate hyper-parameters.