Health and Safety Executive

Projection of mesothelioma mortality in Great Britain

Prepared by **Health and Safety Laboratory** for the Health and Safety Executive 2009

RR728
Research Report

Health and Safety Executive

Projection of mesothelioma mortality in Great Britain

Emma Tan & Nick Warren
Health and Safety Laboratory
Harpur Hill
Buxton
Derbyshire
SK17 9JN

There has been an increase in mesothelioma mortality in Great Britain, with 1705 deaths recorded in 2006. In 2005, a statistical model was developed based on a simple birth-cohort model, which assumes that the risk of mesothelioma depends on age and years of exposure and that an individual's asbestos exposure depends on the year of exposure. An optimisation technique was used to fit the model and a profile of the population exposure was estimated. Projections of the future burden of mesothelioma mortality were calculated, however statistical uncertainties in the formulation of the model could not be taken into account. In this report, the model has been refined and refitted using the MATLAB's fminsearch function and the Metropolis-Hastings algorithm, a Markov Chain Monte Carlo technique. Credible intervals for model parameters as well as prediction intervals for future cases of mortality amongst males are presented. Mortality amongst all males is expected to keep increasing, reaching a peak at around 2,040 deaths in the year 2016, with a rapid decline following the peak year. Around 91,000 deaths are predicted to occur by 2050 with around 61,000 of these occurring from 2007 onwards.

This report and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect HSE policy.

HSE Books

First published 2009

Acknowledgements:

The authors would like to thank Andrew Darnton and John Hodgson (CSAG Epidemiology, HSE) for their valuable contribution to this report.

EXECUTIVE SUMMARY

Aims

This report presents a Bayesian statistical analysis of mesothelioma mortality in Great Britain between the years 1968 and 2006. This report updates previous work carried out by HSE Statistics Branch, using Bayesian Markov Chain Monte Carlo methods.

The aims of the statistical analysis were:

- Using suitable software, to construct a more efficient and statistically rigorous algorithm for model parameter estimation;
- To refit the collective population dose model to data up to and including 2006 and incorporate terms for background mesothelioma cases not caused by asbestos exposure;
- To test the adequacy of the models by running projections based on data up to earlier years and to assess the fit in later observed years; and
- To produce updated estimated annual mesothelioma deaths to 2050 with confidence and prediction intervals.

Main Findings

- The expected number of mesothelioma cases amongst males is projected to increase to a peak of 2038 (90% prediction interval [1929, 2156]) in the year 2016 (90% prediction interval [2015,2016]), decreasing thereafter and eventually reaching a point where the majority of deaths are 'background cases'. This is consistent with previous HSE work.
- The non-clearance model (with a clearance half-life of 1,000,000 years) provided a better fit to the data than a clearance model with a shorter half-life.
- Males aged 20 to 49 years were most likely to be exposed to asbestos.
- Estimated population exposure to asbestos increased rapidly from the 1930s to the late 1960s, reaching a global maximum year of exposure in 1963. There were also two periods around 1930 and 1950 where population exposure briefly reached local peaks. These peaks do not appear to be statistical artefacts. They may be related to events which occurred around the time of the peaks. The first coincides with the introduction of the Asbestos Industry Regulations in the UK in 1931 as well as the Great Depression. The second occurs just after World War II after which shipyard activity – especially in naval yards - will have reduced.
- The background rate was estimated at approximately 1.08 (90% C.I. [0.71, 1.51]) cases per million amongst males, suggesting that there are a small number of cases (about 23 per year) that are not caused by exposure to asbestos.

Limitations

- A comparison of predictions made by the model with selected early cutoffs for the input data (using data up to 1987, 1992 and 2002) with the observed data in later years suggested that the model does not systematically under- or over-predict the scale of mesothelioma mortality in later years. However care must be taken when making projections based on available data; any outlying data for the most recent years available may have high leverage and thus have a greater influence on the fit of the model.

- The updated model provides a reasonable basis for making relatively short-term projections of mesothelioma mortality in Britain, including the extent and timing of the peak number of deaths. However, longer-term predictions comprise two additional sources of uncertainty which are not captured within the prediction intervals for the annual number of deaths: 1) whether the form of the model is valid for more recent and future exposure contexts, and 2) if the model is valid in such contexts, the uncertainty arising from the particular choice of the population exposure profile beyond 1978.

Recommendations

- Comparisons of the projections with new data should be made in order to further assess the fit and the adequacy of the existing model. The model may also be refitted to obtain updated model parameters and model projections.

- Alternative models where, for example, the risk of mesothelioma levels off with time since exposure, should be investigated.

- Further work should be carried out on female data. Different approaches to fitting models to female data should be considered, in particular, whether to assume a common value for certain parameters for both males and females.

CONTENTS

1 INTRODUCTION

Mesothelioma is a form of cancer that is almost always caused by exposure to asbestos. The principal sites affected are the pleura (the membrane which covers the lungs and lines the internal chest wall) and the peritoneum (the membrane which forms the lining of the abdominal cavity). It may also occur in the heart, the pericardium and the tunica vaginalis. The majority of men who develop mesothelioma have had occupations with significant exposure to asbestos fibres (Rake *et al.*, 2009). Mesothelioma has a long latency period; symptoms usually emerge between 15 and 60 years (with a mean of about 40 years) after exposure to asbestos. Mesothelioma is rapidly fatal; 85% of all deaths have been amongst men. The majority of deaths occur amongst those over 60 years of age, with few deaths occurring amongst those under 50.

Imports of asbestos as well as its use began to increase in the early 1900s. Exposure to asbestos of the UK population is also likely to have increased during the same period, especially amongst those whose occupations involve high exposure to asbestos. The level of asbestos imports reached a peak in the mid-1960s; asbestos was widely used in building materials until the late 1970s, after which its use rapidly decreased. The removal of asbestos subsequently began to increase from 1980 onwards and the use of asbestos in thermal insulation was eventually banned in 1986. Due to the long latency period however, the annual number of deaths caused by mesothelioma has yet to peak, decades after peak usage, and an increase in mortality rates in the next few years is expected. Whilst the majority of cases of mesothelioma are caused by exposure to asbestos, much of which occurred in occupational settings, particularly in men, a small number of cases (which are referred to as background cases) occur spontaneously amongst those with no exposure. Mesothelioma now accounts for over 1% of all cancers.

Attempts to predict the future number of mesothelioma cases have been carried out in several other countries including Denmark (Kjaergaard and Andersson, 2000), Australia (Leigh and Driscoll, 2003) and France (Ilg *et al.*, 1998) where it has been predicted that mesothelioma mortality has yet to peak. In the United States, a peak has been predicted around the years 2000 to 2004 (Price and Ware, 2004). Projections of the future burden of mortality in Great Britain have been published by Health and Safety Executive Statistics Branch and have been widely used both within HSE and externally. Earlier projections made using a simple age-birth cohort model where the annual mesothelioma rate r_{ab} for a particular age is given by the overall mesothelioma death rate for that particular age multiplied by the mesothelioma risk in the appropriate birth cohort

$$r_{ab} = k_a c_b$$

were found to be inadequate since the model assumed that the ratio of death rates at different ages is identical across all birth cohorts (Peto and Hodgson, 1995).

Hodgson *et al.* (2005) developed a more complex model based on the dose-response model for mesothelioma (Heath Effects Institute, 1991). Using this updated model and fitting to observed deaths to 2001, mesothelioma mortality in Great Britain amongst males aged under 90 was predicted to reach a peak at around 1,650 to 2,100 deaths per year some time between 2011 and 2015, followed by a rapid decline.

This report presents a more refined statistical analysis of mesothelioma mortality amongst males in Great Britain based on Markov Chain Monte Carlo (MCMC) methods using a modified form of the model formulated by Hodgson *et al.* (2005). Predictions of mesothelioma mortality and estimates of the peak year (the year at which mesothelioma mortality will peak) and the peak number of deaths are also presented.

1.1 DATA

The number of deaths due to mesothelioma in Great Britain (where mesothelioma was mentioned on the death certificate) is published annually by HSE. In both males and females, 99% of all these deaths have been amongst those between the ages of 20 to 89. The data used in this report are based on deaths of males and females between the years 1968 to 2006. Figure 1 shows the observed deaths amongst males and females aged 20 to 89 between the years 1968 to 2006.

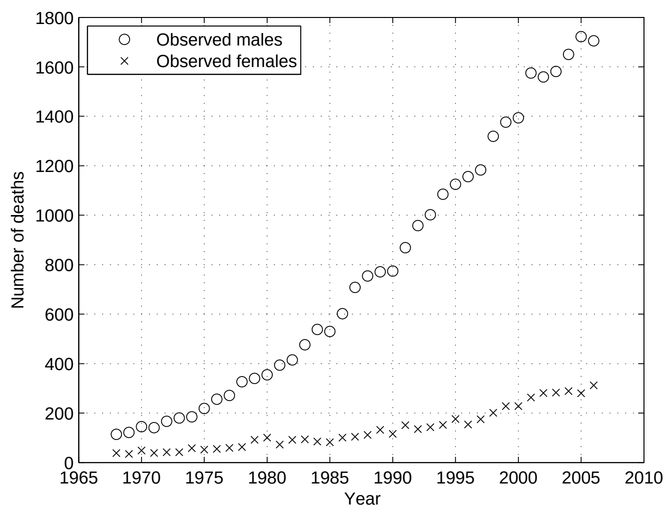

Figure 1 Male and female mesothelioma deaths (aged 20 to 89) from 1968 to 2006

1.2 MODEL

The current model developed by Hodgson *et al.* (2005) is based on the dose-response model for mesothelioma where an individual's risk is proportional to the cumulative exposure multiplied by the second or third power of time since exposure lagged by 10 years. The dose-response model is expressed as

$$R \propto D(t-10)^k \tag{1}$$

where R is the risk, D is the increase in cumulative exposure, t is the time since exposure and k is the power of time. Since the predicted risk after a given time varies linearly with cumulative exposure, the model can be applied to the collective exposure for groups of individuals.

In the current model, the following additional assumptions were also incorporated.

(1) The average asbestos exposure for males in Great Britain in each year can be summarised by a single estimate and that their exposure in any given year depends on their age.

(2) A parameter to model completeness of mesothelioma diagnosis over time was included.

(3) A parameter for the half-life for the proportion of asbestos fibres remaining in the lungs was included.

The mesothelioma death rate for men of a given age in a given year was then assumed to be proportional to the sum of the risks due to exposure in all previous years of their lifetime (excluding the last L years).

The current model can be represented as follows:

$$F_{A,T} = \frac{[\sum_{l=1}^{A-1} W_{A-l} D_{T-l} I \left(l+1-L\right)^{k} 0.5^{l/H}] D_{x_T} P_{A,T} M}{\sum_{A=20}^{89} \sum_{T=1968}^{2006} [\sum_{l=1}^{A-1} W_{A-l} D_{T-l} \left(l+1-L\right)^{k} 0.5^{l/H}] D_{x_T} P_{A,T}} \tag{2}$$

where $F_{A,T}$ is the number of deaths at age A in year T, W_A is the overall age-specific exposure potential at age A, D_T is the overall population exposure in year T, D_{x_T} is the proportion of mesothelioma deaths in year T that are recorded, L is the lag period in years between exposure and disease occurrence, H is the half-life in years for asbestos clearance from the lungs, k is the exponent of time representing the increase of risk with increase of time since exposure, $P_{A,T}$ is the person-years at risk for age A in year T, M is the total observed mesothelioma deaths from 1968 to 2006, I is an indicator variable where $I = 0$ if $l < L - 1$ and $I = 1$ otherwise and l indexes years lagged from the risk year. A consequence of the formation of the model is that the sum of the estimated deaths over all ages over the period for which observations of deaths were available is equal to the total observed number M. The number of deaths is assumed to follow a Poisson distribution.

An allowance for background rate was not included in the models of Hodgson *et al.* (2005) despite a widely assumed rate of 1 to 2 per million per year. If the percentage of deaths due to background cases is assumed to be 1.5%, this equates to around 32 deaths in 2006. In the HSE (2003) paper 'Mesothelioma Mortality in Great Britain: Estimating the Future Burden', the background rate was estimated by carrying out a linear regression analysis of the annual number of female mesothelioma deaths against male annual deaths and a background rate of 26.5 cases per year was calculated. As the number of deaths in certain years and amongst the most recent birth cohorts is small, the proportion of deaths due to background cases may be high. It is thus important not only to include the possibility of background cases, but to correctly model the number of background cases. The model with background cases taken into account can be represented as follows:

$$F_{A,T} = \frac{[\sum_{l=1}^{A-1} W_{A-l} D_{T-l} I \left(l+1-L\right)^{k} 0.5^{l/H}] D_{x_T} P_{A,T} (M - \sum_{A=20}^{89} \sum_{T=1968}^{2006} B_{A,T})}{\sum_{A=20}^{89} \sum_{T=1968}^{2006} [\sum_{l=1}^{A-1} W_{A-l} D_{T-l} \left(l+1-L\right)^{k} 0.5^{l/H}] D_{x_T} P_{A,T}} + B_{A,T} \tag{3}$$

where $B_{A,T}$ is the number of background cases for age A at year T.

2 STATISTICAL METHODOLOGY

The model was originally fitted by Hodgson *et al.* (2005) using a manual approach to minimising the model deviance, a measure of how well the model fits the observed data. The Poisson deviance can be expressed as

$$D = 2 \sum_{A,T} \left[Y_{A,T} \log \left(\frac{Y_{A,T}}{\hat{F}_{A,T}} \right) - (Y_{A,T} - \hat{F}_{A,T}) \right] \tag{4}$$

where $Y_{A,T}$ are the observations and $\hat{F}_{A,T}$ are the fitted values. Obtaining parameter estimates by maximum likelihood is equivalent to obtaining estimates by minimising the model deviance. Due to the iterative fitting approach used by Hodgson *et al.* (2005), confidence intervals for both the parameter estimates and predictions of mortality could only be obtained using an informal numerical approach rather than analytically. Instead, approximate 95% confidence intervals for the level and the timing of the predicted peak in mesothelioma deaths were calculated by adjusting the model parameters to produce a lower/earlier peak and a higher/later peak, corresponding to a change in deviance from the optimal model to the 5% critical value of the χ^2 distribution on the number of degrees of freedom in the model.

In this report, statistical models have been fitted to the data using both the *fminsearch* function in Matlab (The MathWorks, Inc., 2008) and the Metropolis-Hastings algorithm (Hastings, 1970), a Markov Chain Monte Carlo (MCMC) technique. The former allowed the data to be fitted quickly and easily by minimising the model deviance, although the disadvantage was that confidence intervals could not be provided. The latter allowed not only model parameters to be estimated, but also allowed credible intervals to be easily obtained using formal statistical methods.

2.1 Model parameters

- W_A: The age-specific exposure potential, W_A, allowed the exposure of a male to differ by age. Nine parameters were assigned to W_A, representing the exposure weighting for the age groups (in years) 0 to 4 (W_1), 5 to 15 (W_2), 16 to 19 (W_3), 20 to 29 (baseline), 30 to 39 (W_4), 40 to 49 (W_5), 50 to 59 (W_6), 60 to 64 (W_7) and $65+$ (W_8), with the age group 20 to 29 years chosen as the baseline category.

- D_T: The overall population exposure, D_T, represents the average 'effective carcinogenic dose' in the breathing zone of men aged 20 to 89 years and is included as a unit-free parameter vector in the model. The shape of the exposure curve and the change in exposure levels over time is the main interest in the inclusion of D_T. D_T was defined by growth and decline rates for years in multiples of 10 before and after the maximum exposure year, '$Peakyear$' (at which the gradient of the exposure curve is zero). The growth rates for intermediate years were determined by linear interpolation. The set of growth rates at $Peakyear - 65$ (D_1), $Peakyear - 55$ (D_2), $Peakyear - 45$ (D_3), $Peakyear - 35$ (D_4), $Peakyear - 25$ (D_5), $Peakyear - 15$ (D_6), $Peakyear - 5$ (D_7), $Peakyear + 5$ (D_8)and $Peakyear + 15$ (D_9) was included as a parameter in the model. From the year 2000 onwards, the exposure distribution assumed in Hodgson *et al.* (2005) was used. Between the last year for which the growth rate was estimated and 2000, the value of the exposure was determined by linear interpolation.

- The diagnostic trend D_{x_T} was defined by a parameter α, representing the annual percentage decrease in the number of missed cases working backwards in time from the year 1997, in which diagnosis was assumed to be essentially complete (98%).

- The background rate ($Rate$) is represented by the number of cases per million in the male population. The age distribution of the background cases in each year is assumed to be $(A-L)^k$. The proportion of background cases at age A in each year is therefore assumed to be $\frac{(A-L)^k}{\sum_A (A-L)^k}$.

2.2 MATLAB'S FMINSEARCH FUNCTION

Matlab's *fminsearch* function can be used to minimise the model deviance D, a function of several variables, starting at initial estimates. The starting values which were used are the parameter estimates obtained by Hodgson *et al.* (2005) for the non-clearance model. The advantage of using *fminsearch* is that the function is easy to implement and can quickly provide parameter estimates however confidence intervals are not provided.

2.3 MARKOV CHAIN MONTE CARLO

2.3.1 Metropolis-Hastings

From a Bayesian perspective, the parameters of a statistical model are considered random quantities. Bayesian inference can usually be summarised by random draws from the posterior distributions of the model parameters. Let $L(Y|\theta)$ be the likelihood function of the data Y, θ be the vector of model parameters and $\phi(\theta)$ be the prior distribution of the parameters. Assuming that the observations follow a Poisson distribution, the likelihood function is

$$L(Y|\theta) = \prod_{A,T} \left(\frac{e^{-\hat{F}_{A,T}} \hat{F}_{A,T}^{Y_{A,T}}}{Y_{A,T}} \right)$$

which is the product of the individual likelihood contributions for each observation over all ages and years of death. The posterior distribution $\pi(\theta)$ of θ is $\pi(\theta) \propto L(Y|\theta)\phi(\theta)$. Unfortunately, evaluation of the posterior distribution is normally extremely difficult and numerical techniques, particularly MCMC, are required. MCMC techniques require simulation to generate random samples from a complex posterior distribution. A large number of random draws from the posterior distribution is generated. After a burn-in period (where an initial portion of samples are discarded to minimise the effect of initial values on posterior inference), the empirical distribution should eventually closely approximate the true shape of the posterior distribution. The MCMC chain is thinned in order to reduce autocorrelation. The process of thinning records samples periodically (e.g. at every 20th iteration) and discards the remaining samples. Point estimates and credible intervals are then calculated.

In the Metropolis-Hastings algorithm, given θ_t at time point t, the next state θ_{t+1} in the chain is chosen by sampling a candidate point θ^* from a proposal distribution $q(\cdot|\theta_t)$. The candidate point θ^* is then accepted with probability p where

$$p = \min\left[1, \frac{\pi(\theta^*)q(\theta_t|\theta^*)}{\pi(\theta_t)q(\theta^*|\theta_t)}\right]. \tag{5}$$

If the candidate point is accepted, the next state $\theta_{t+1} = \theta^*$. If the point is rejected, the chain does not move, i.e. $\theta_{t+1} = \theta_t$. The process is then repeated for state θ_t at every time point t to obtain a sequence of values $\theta_1, \theta_2, \ldots$. The approximate distributions at each step in the simulation converge to the target distribution of interest, $\pi(\theta)$. As θ is a vector of model parameters, each component will be individually updated for convenience.

2.3.2 Prior distributions

Non-informative prior distributions for each parameter were chosen by considering plausible ranges, taking into account the results in Hodgson *et al.* (2005), as follows.

The power of time since exposure is represented by k and has been estimated at between 2 and 3 in previous analysis. It was unlikely that the risk decreased with time since exposure, hence the prior for k was chosen to be $U(0, 10)$. Each of the W parameters represents age-specific exposure potential and can only take positive values. It was considered unlikely that the risk in any of the age groups was 10 times greater than that of males aged 20 to 29 (the baseline age group), hence the priors for W were chosen to be $U(0, 10)$. Each of the D parameters represents the growth rates of population exposure levels. As the overall population exposure can only take positive values the decline rate must not exceed 100%, and hence the lower bound for D must be -100. Taking into account the data on asbestos imports as well as the levels of asbestos use in Great Britain, the peak year of exposure was assumed to be between 1950 and 2000, hence the prior distribution of $Peakyear$ was chosen to be uniformly distributed on integer values between 1950 and 2000. By definition, the background rate can only take positive values. Hodgson *et al.* (2005) suggest that a background rate of 1% to 2% of total mesothelioma deaths, equating to about 25 to 50 male deaths annually, is widely assumed, A uniform $U(0, 20)$ prior was chosen for α (cases per million).

Due to problems encountered when fitting the model, various priors distributions for H were considered. However convergence was not attained after several thousand iterations (see 3.2 for further details). Table 1 shows the prior distributions that were used.

2.3.3 Proposal distributions

The proposal distributions for the model parameters are shown in Table 2. Each proposal distribution was chosen such that it was easy to sample from $q(\cdot|\theta_t)$, each step $\theta^* - \theta_{t-1}$ moves a reasonable distance in the parameter space, and the steps generated are not rejected too frequently. Apart from the proposal distribution for $Peakyear$, each distribution was chosen to be normal with a standard deviation such that the acceptance probability was approximately 0.2 to 0.5. The proposal distributions do not have an impact on the posterior parameter estimates, only on the convergence, mixing and autocorrelation of the chains generated by the Metropolis-Hastings algorithm.

Table 1 Metropolis-Hastings Algorithm: Prior distributions for model parameters

Parameter	Prior
H	various
k	$U(0, 10)$
$W_k \forall k$	$U(0, 10)$
$D_k \forall k$	$U(-100, 200)$
α	$U(-0.07, 0.09)$
$Peakyear$	$U(1950, 2000)$
$Rate$	$U(0, 20)$

Table 2 Metropolis-Hastings Algorithm: Proposal distributions for model parameters

Parameter	Proposal
H	$N(H_{t-1}, 100^2)$
k	$N(k_{t-1}, 0.05^2)$
W_1	$N(w_{1,t-1}, 0.004^2)$
W_2	$N(w_{2,t-1}, 0.01^2)$
W_3	$N(w_{3,t-1}, 0.04^2)$
W_4	$N(w_{4,t-1}, 0.12^2)$
W_5	$N(w_{5,t-1}, 0.12^2)$
W_6	$N(w_{6,t-1}, 0.1^2)$
W_7	$N(w_{7,t-1}, 0.5^2)$
W_8	$N(w_{8,t-1}, 0.8^2)$
D_1, D_2, D_3	$N(d_{1,t-1}$, various)
D_4	$N(d_{2,t-1}, 3^2)$
D_5	$N(d_{3,t-1}, 2.8^2)$
D_6	$N(d_{4,t-1}, 1.1^2)$
D_7	$N(d_{5,t-1}, 1^2)$
D_8	$N(d_{6,t-1}, 2^2)$
D_9	$N(d_{7,t-1}, 4.2^2)$
α	$N(\alpha_{t-1}, 0.045^2)$
$Peakyear$	P$(Peakyear_t = Peakyear_{t-1} + 1) = 0.5$
	P$(Peakyear_t = Peakyear_{t-1} - 1) = 0.5$
$Rate$	$N(Rate_{t-1}, 0.6^2)$

2.4 MODELS FITTED

Several different models were fitted to the dataset using Matlab's *fminsearch* function and the Metropolis-Hastings algorithm. The parameters k, W, D and $Peakyear$ were present in all the models. L was fixed at 10 as in Hodgson *et al.* (2005). Non-clearance models with H fixed at 1,000,000 were fitted. The parameter α was removed from some of the models after considering the results in Hodgson *et al.* (2005) and the results of preliminary analyses. The background rate was estimated in some models and fixed at 1.4 cases per million in others, as derived from preliminary analysis using *fminsearch*. Table 3 shows the different models that have been fitted, indicating the state of H, α and $Rate$ in the model.

Table 3 Models fitted

Model	Presence of parameters in model		
	Clearance (H)	Diagnostic trend (α)	Background rate ($Rate$)
A	Fixed at 1,000,000	Absent	Estimated
B	Fixed at 1,000,000	Absent	Fixed at 1.4 per million
C	Fixed at 1,000,000	Estimated	Estimated
D	Fixed at 1,000,000	Estimated	Fixed at 1.4 per million
E	Estimated	Estimated	Fixed at 1.4 per million

In the MCMC analysis during the Metropolis-Hastings update steps, the parameters were updated one at a time in the following order:

- H (where estimated)
- k
- $W(1)$,...,$W(8)$
- $D(1)$,..., $D(9)$
- α (where estimated)
- $Peakyear$
- $Rate$ (where estimated)

The population data used in the analyses were the ONS mid-year population estimates for 1968 to 2006 and GADpopulation projections for 2007 to 2050.

3 RESULTS

3.1 MATLAB'S *FMINSEARCH*: MALES

Models A, B, C, D and E were all fitted to the dataset using the *fminsearch* function in Matlab. The results from fitting Model A are displayed in Table 4. When H was estimated in the model (Model E), the deviance decreased as the value of H increased. An optimal value of H which minimised the deviance was unattained after running *fminsearch* for several thousand iterations. This suggested that the optimal value of H is infinitely large, equivalent to removing the H term from the model and, in effect, resulting in a non-clearance model. An inspection of the change in deviance with a change in H (keeping the other parameters fixed) indicated that although the deviance did decrease as H increased, the change in deviance for very large values of H was very small. In light of this, H was retained and fixed at 1,000,000 in the non-clearance models A, B, C and D.

There was strong negative correlation between H and k. When H was fixed at small values in the preliminary analysis, the value of k minimising the deviance was larger than the corresponding value obtained when H was fixed at 1,000,000. The estimates of k in Models A to D were in the range 2.47 to 2.55 when H was fixed at 1,000,000, which are close to the value 2.60 obtained in the non-clearance model in Hodgson *et al.* (2005).

Convergence of the exposure change parameters could not be achieved when all of these were included as parameters to be estimated. In particular, successive iterations in initial attempts to fit the model using *fminsearch* led to ever increasing values particularly for the growth rate at $Peakyear - 45$ ($D(3)$) and $Peakyear - 55$ ($D(2)$). $D(1)$, $D(2)$ and $D(3)$ were therefore assigned fixed values of 1, 1000 and 100000 respectively. Fitting the model with these constraints led to an exposure profile with a sharp local peak at $Peakyear - 35$ and $Peakyear - 15$. Attempts were then made to smooth the population exposure profile prior to the peak year by altering the starting values as well as changing the assumptions of the population exposure prior to $Peakyear - 45$. However, several of the attempts resulted in $D(3)$ increasing and failing to converge, as well as $D(4)$ eventually taking up negative values. One approach that was used to smooth the exposure profile was to constrain $D(4)$ ($Peakyear - 35$) to be positive and replace $D(1)$, $D(2)$ and $D(3)$ with a single parameter so that the growth rate was the same for all years prior to $Peakyear - 45$. This resulted in a smooth population exposure profile, however the deviance statistic obtained from fitting this model was statistically significantly higher than the deviance statistic obtained from fitting a model with fixed initial growth rate parameters. In light of this, $D(3)$ was fixed at 100000. The growth rate at $Peakyear - 55$ ($D(2)$) and $Peakyear - 65$ ($D(1)$) were fixed at 1000 and 0 respectively. The estimated exposure curve indicated a high level of exposure around the year 1930 followed by a sharp decrease in exposure in the following years. A rapid increase in population exposure then followed from the 1940s to the mid-1960s, reaching a maximum in 1963 and decreasing thereafter.

The estimates of the age-specific exposure potential parameters suggested that this was highest for males aged 30 to 49 years. Males aged below 15 years and above 50 years were least likely to be exposed.

When the background rate was included in the model as a parameter, it was estimated at 1.22 cases per million in Model A. This corresponds to approximately 26 background cases in males aged between 20 and 89 in 2006.

The diagnostic trends estimated for Models C and D were negative, suggesting that fewer cases of mesothelioma are missed moving backwards in time. The small positive estimates in Hodgson *et al.* (2005) previously suggested that the number of missed cases increased moving backwards in time. An inspection of the change in deviance with a change in the diagnostic trend (keeping the other parameters fixed) indicated that changes in the diagnostic trend from small absolute positive values to small absolute negative values resulted in very small changes in deviance, indicating that diagnostic trend does not play a large role in the fit of the model compared to some of the other parameters. A plot of the change in deviance with diagnostic trend can be found in Appendix 2.

The fit of Model A is illustrated in Figures 2A to 2D which show plots of fitted and observed deaths by year of birth, age and year of death. For Model A, the year at which mortality reaches a peak was estimated at 2016, with a peak level of 2,020 cases. Hodgson *et al.* (2005) predicted a peak at around 1,650 to 2,100 deaths between 2011 and 2015. The results for Models B to D can be found in Tables 8 to 10 in Appendix 1.

Table 4 *fminsearch:* Parameter estimates for Model A

Parameter estimates			
k	2.47	Background rate	1.22
Maximum exposure year	1963	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1898 ($D(1)$)	0 (fixed)	0 to 4	0.00
1908 ($D(2)$)	1000 (fixed)	5 to 15	0.00
1918 ($D(3)$)	100000 (fixed)	16 to 19	0.30
1928 ($D(4)$)	-92.4	20 to 29	1.00 (baseline)
1938 ($D(5)$)	104.9	30 to 39	1.79
1948 ($D(6)$)	-26.0	40 to 49	1.54
1958 ($D(7)$)	38.0	50 to 59	0.07
1963	0 (by definition)	60 to 64	0.33
1968 ($D(8)$)	-7.7	65+	0.00
1978 ($D(9)$)	-16.3		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	2020	Peak year	2016
Deviance	213	Diagnostic trend	-

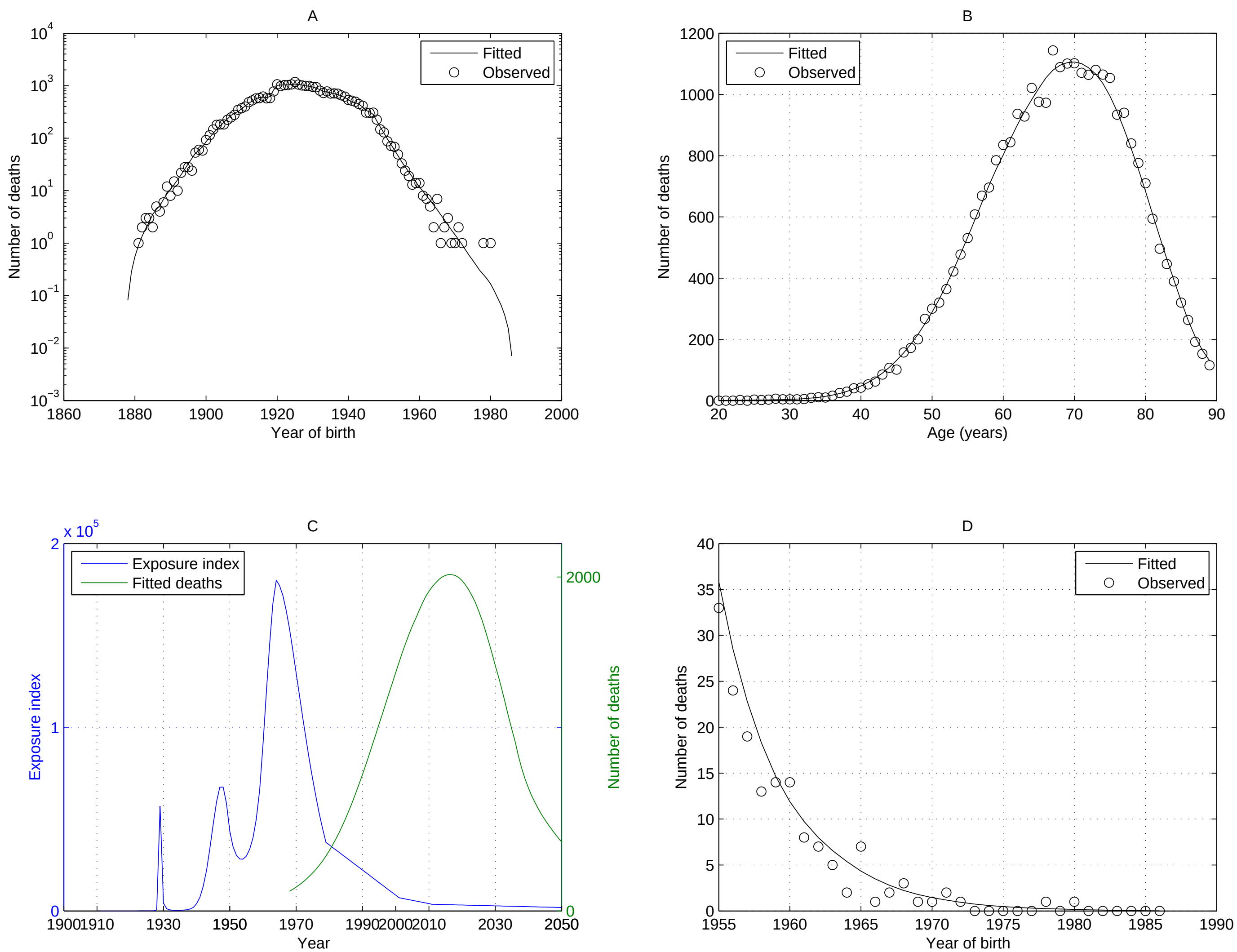

Figure 2 *fminsearch*: (A) Observed and fitted deaths by year of birth. (B) Observed and fitted deaths by age. (C) Observed and fitted deaths by year of death, with derived exposure index. (D) Observed and fitted deaths for 1955-1985 birth cohorts.

3.2 METROPOLIS-HASTINGS: MALES

Models A, B, C and E were fitted to the dataset using the Metropolis-Hastings algorithm for 35,000 iterations after a burn-in of 20,000 iterations. The starting values used were $k = 2.4$, $W = [0.001, 0.001, 0.01, 1.3, 1.3, 0.00, 0.001, 0.001]$, $D = [10000, 10000, 10000, -60, 40, -8, 20, -16, -10]$, $Peak = 1966$ and $Rate = 1.4$. The results from fitting Model A are displayed in Table 5. When H was present in the model in Model E, the H values in the chain produced by the algorithm generally increased and convergence in the distribution of H was unattained even after several hundreds of thousands of iterations. This suggested that there is no finite optimal value of H, just as the results of fitting Model E using *fminsearch* indicated.

When α was present in the model, the α values in the chain spanned both positive and negative values even after several hundreds of thousands of iterations. The median value of α was -0.006 with a 90% credible interval of (-0.065,0.078) for Model C. As the credible interval covers zero, the results suggest that the diagnostic trend component can be removed from the model (as was seen in the *fminsearch* results), resulting in a more parsimonious model.

Convergence of the exposure change parameters could not be achieved when all of these were included as parameters to be estimated. It was also noted that $D(4)$ eventually took up negative values in the MCMC chain. These observations were both seen in Section 3.1 when *fminsearch* was used. In light of this, $D(1)$, $D(2)$ and $D(3)$ were fixed at 0, 1000 and 100000 respectively. The estimated exposure curve in all of the fitted models indicated a high level of exposure around 1930, soon followed by a sharp decrease in exposure. A rapid increase in population exposure followed from the 1940s to the mid-1960s, reaching a maximum in 1963 and decreasing thereafter.

The estimates of the age-specific exposure potential parameters suggested that this was highest for males ages 30 to 49 years. Males aged below 15 years and above 50 years were least likely to be at risk. Due to the lag period before the effects of exposure starts, there was high uncertainty in the estimates of relative exposure potential for males aged 50 and above.

The background rate was estimated at 1.08 and 1.15 with 90% credible intervals of (0.71,1.51) and (0.72,1.61) in Models A and C respectively. The credible intervals include 1.4, the background rate assumed in Models B, D and E, suggesting that this assumption of background rate was a reasonable one to make.

For Model A, the peak year was estimated at 2016. The peak level was estimated to be 1990 deaths with a 90% credible interval $(1915, 2072)$ for males aged 20 to 89. Figure 3 shows a plot of fitted and observed deaths by year of death along with a 90% credible interval. Figures 4A to 4D show plots of fitted and observed deaths by year of birth, age and year of death. Projections of the peak number for all males can be found in Section 4. The results for Models B to D can be found in Tables 11 to 13 in Appendix 1.

Table 5 Metropolis-Hastings: Posterior median and 90% credible intervals for Model A

Parameter estimates and 90% credible interval			
k	2.42 (2.28,2.56)	Background rate	1.08 (0.71,1.51)
Maximum exposure year	1963	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1898 ($D(1)$)	0 (fixed)	0 to 4	0.0019 (0.0001,0.0074)
1908 ($D(2)$)	1000 (fixed)	5 to 15	0.0023 (0.0002,0.0091)
1918 ($D(3)$)	100000 (fixed)	16 to 19	0.25 (0.048,0.393)
1928 ($D(4)$)	-91.3 (-98.2,-50.1)	20 to 29	1.00 (baseline)
1938 ($D(5)$)	104.6 (44.8,135.5)	30 to 39	1.79 (1.51,2.03)
1948 ($D(6)$)	-25.5 (-34.9,-8.28)	40 to 49	1.59 (1.25,1.94)
1958 ($D(7)$)	36.6 (23.2,47.8)	50 to 59	0.13 (0.01,0.41)
1963	0 (by definition)	60 to 64	0.56 (0.06,1.54)
1968 ($D(8)$)	-7.5 (-14.1,-1.4)	65+	0.42 (0.03,1.56)
1978 ($D(9)$)	-18.6 (-27.5,-8.8)		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	1990 (1915,2072)	Peak year	2016 (2015, 2017)
Deviance	230 (215,238)	Diagnostic trend	-

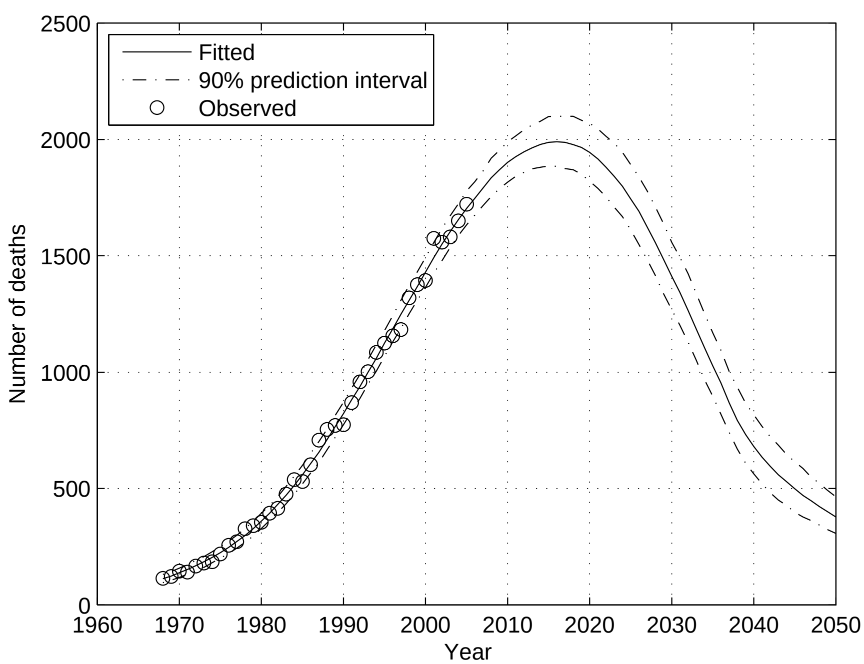

Figure 3 Observed deaths with 50th percentile curve and 90% prediction interval for males

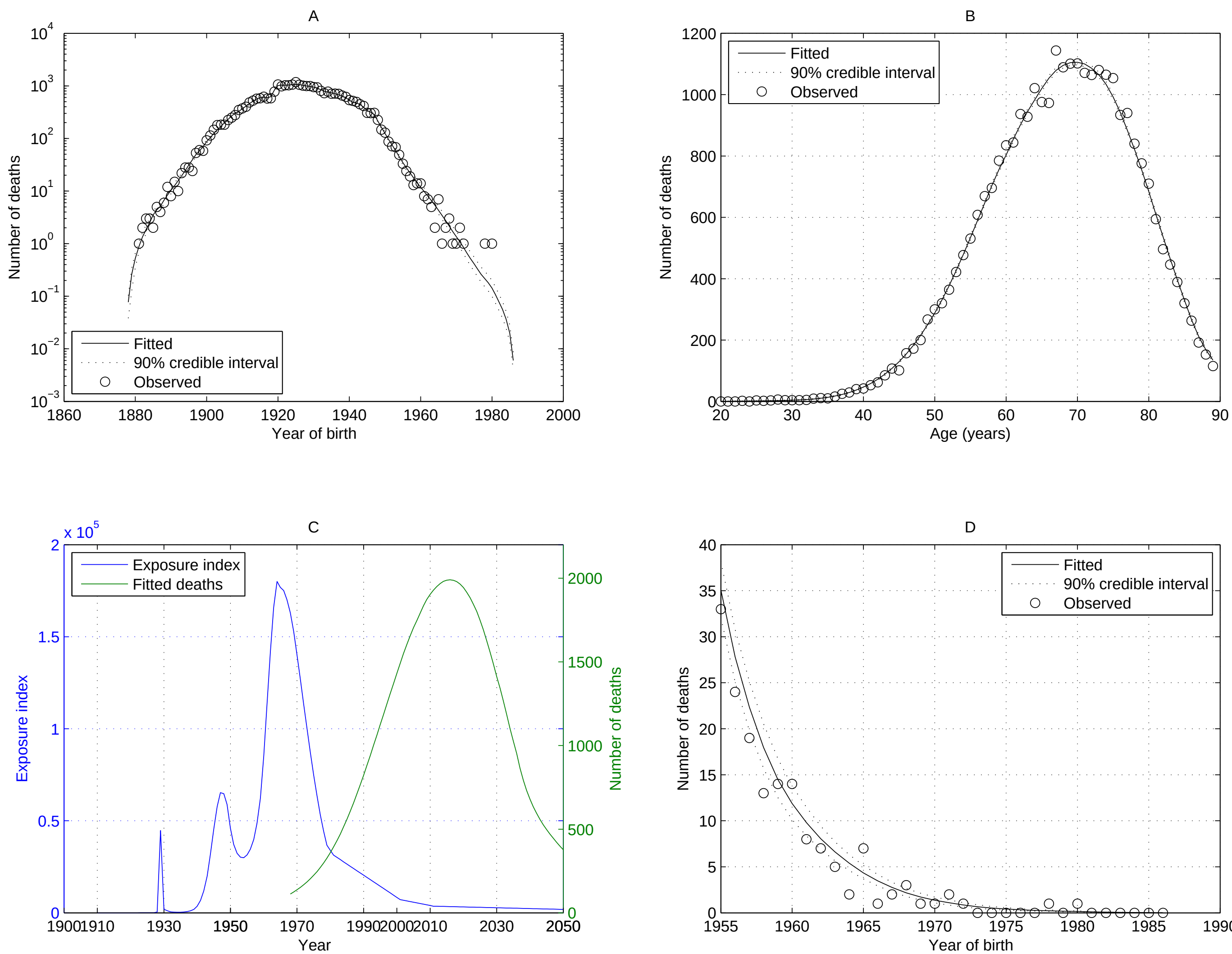

Figure 4 Metropolis-Hastings (males): (A) Observed and fitted deaths by year of birth. (B) Observed and fitted deaths by age. (C) Observed and fitted deaths by year of death, with derived exposure index. (D) Observed and fitted deaths for 1955-1985 birth cohorts.

3.3 FEMALES

This report has so far concentrated on modelling deaths amongst males. The small number of observed female deaths in comparison to male deaths leads to greater uncertainty in modelling female deaths alone using equation (3) where all model parameters are estimated. In exploratory analysis carried out on data for females, a simple substitution of the estimated parameters obtained in Section 3.1 and 3.2 for males did not result in a satisfactory estimation of female deaths. This suggested that some of the parameter values may not be common to both males and females and that a set of separate parameter estimates are required in order to make reliable inferences about female mortality and make inference on model parameters for females.

When Model A was fitted to data on females using the Metropolis-Hastings algorithm, there was very high uncertainty in the growth rate parameters as well as high estimates of some of the age specific exposure parameters. This suggested that the approach used to model female data may have been inadequate and that alternative methods of approach must be sought.

As the number of deaths amongst females is much lower than amongst males, a higher proportion of female deaths are due to background cases. The data on females are thus important in their own right as they potentially allow more reliable estimation of background rates to be made.

4 PROJECTIONS

So far, the models have been fitted to data on males aged between 20 and 89. Very few deaths have occurred in males outside this range. The estimate of the peak number of deaths in Section 3.2 also only included males aged between 20 and 89. To estimate the number of deaths amongst all males, the estimate of the peak number was rescaled as follows.

At year T, the ratio $R_T = Y_T^{All}/Y_T$ is calculated, where Y_T is the number of observed deaths in males aged between 20 and 89, and Y_T^{All} is the number of deaths in males of all ages. An inspection of the change in R_T over time suggests that R_T remains close to one until the year 1985, after which R_T begins to increase. In light of this, a linear regression model is fitted to R_T as follows:

$$R_T = \alpha + \beta(T - 1967) + \epsilon_T \tag{6}$$

where R_T is the ratio at year T, α is the intercept, β is the fixed effect term corresponding to T and ϵ_T is a normally distributed random error with mean zero and variance σ^2. The estimates of α and β were found to be 0.9872 (95% C.I. [0.9802,0.9941]) and 0.0008 (95% C.I. [0.0005,0.0010]) respectively. The estimate $\hat{\sigma}$ of σ is 0.0033. A projection of the ratio at the peak year (predicted to be 2016) is

$$\begin{aligned} R_{2015} &= 0.9872 + 0.000763(2016 - 1967) \\ &= 1.025. \end{aligned}$$

This projection, however, assumes that (7) is valid for every year up to the peak year. Prediction intervals and credible intervals for the peak number were obtained via simulation, based on data for males 20 to 89 using Model A. The simulation routine involves obtaining a chain of values for each parameter using the Metropolis-Hastings algorithm as described in Section 2.3.1. After a burn-in of 20,000 iterations, the simulation routine is as follows:

- generate a residual error ϵ_T for R_T using the normal distribution $N \sim (0, \hat{\sigma}^2)$, for every year T between 1968 and 2050;
- using (7) and the simulated value ϵ_T, calculate R_T for every year T between 1968 and 2050;
- obtain a new value in the MCMC chain for the model parameters H (where included), k, W, D, α (where included) and $Rate$ (where included) as described in 2.3;
- using the new parameter values in the chain, calculate the estimated number of deaths F_T in males aged 20 to 89 in year T for each year between 1968 and 2050;
- calculate the estimated total number of deaths $F_T^{All} = R_T F_T$ in all males for each year between 1968 and 2050; and
- generate a random number $\tilde{F}_T^{All}$ with distribution Poisson$(R_T F_T)$.

Repeating the above procedure for 35,000 iterations and thinning to retain every 20th iteration, percentiles of $\tilde{F}_T^{All}$, and thus prediction intervals, at each year between 1968 and 2050 can be obtained. These prediction intervals include stochastic variability. In particular, the 50th percentiles can be used to obtain median estimates and projections of the total number of male deaths. After carrying out the above routine, the estimate of the peak number of deaths in all males is 2038 (90% C.I. [1959,2123]) in the year 2016 (90% C.I. [2016,2018]) . Predicted ratios, projections and credible intervals in males aged between 20 and 89, and in all males, are given in Table 6.

Table 6 Projections of male mesothelioma deaths using Model A

		Projection (90% credible interval) [90% prediction interval]	
Year	Ratio	Males 20-89	All males
2007	1.018	1791 (1759,1822) [1715,1864]	1823 (1789,1855) [1747,1898]
2008	1.019	1835 (1799,1871) [1755,1920]	1869 (1831,1907) [1788,1951]
2009	1.019	1869 (1832,1914) [1788,1953]	1910 (1866,1951) [1827,1993]
2010	1.020	1902 (1857,1948) [1817,1990]	1941 (1892,1989) [1855,2026]
2011	1.021	1926 (1876,1979) [1842,2015]	1968 (1914,2022) [1870,2059]
2012	1.022	1947 (1893,2006) [1859,2042]	1993 (1933,2051) [1897,2084]
2013	1.022	1964 (1906,2030) [1874,2062]	2012 (1947,2076) [1913,2106]
2014	1.023	1979 (1914,2049) [1881,2079]	2027 (1956,2097) [1926,2129]
2015	1.024	1988 (1915,2062) [1886,2099]	2035 (1958,2112) [1929,2141]
2016	1.025	1990 (1911,2069) [1885,2100]	2038 (1959,2121) [1928,2156]
2017	1.025	1988 (1902,2072) [1875,2100]	2037 (1950,2123) [1928,2147]
2018	1.026	1978 (1888,2070) [1870,2100]	2031 (1939,2123) [1912,2152]
2019	1.027	1966 (1869,2062) [1851,2083]	2017 (1920,2118) [1903,2141]
2020	1.028	1945 (1843,2046) [1821,2070]	1997 (1895,2103) [1871,2132]
2021	1.028	1916 (1810,2023) [1790,2045]	1969 (1862,2081) [1843,2103]
2022	1.039	1881 (1773,1995) [1753,2014]	1938 (1824,2055) [1804,2075]
2023	1.030	1841 (1730,1961) [1709,1984]	1899 (1780,2019) [1762,2041]
2024	1.031	1799 (1683,1923) [1668,1945]	1856 (1735,1984) [1711,2003]
2025	1.031	1745 (1628,1876) [1612,1893]	1803 (1678,1935) [1660,1958]
2026	1.032	1692 (1569,1821) [1549,1839]	1746 (1618,1880) [1601,1895]
2027	1.033	1625 (1503,1760) [1485,1780]	1680 (1552,1819) [1534,1838]
2028	1.034	1557 (1432,1694) [1416,1710]	1612 (1480,1751) [1460,1768]
2029	1.035	1486 (1360,1622) [1338,1639]	1538 (1407,1678) [1390,1695]
2030	1.035	1412 (1286,1549) [1268,1558]	1462 (1333,1603) [1314,1626]
2040	1.043	681 (572,804) [563,817]	708 (596,839) [588,851]
2050	1.051	378 (315,458) [307,464]	396 (330,481) [326,487]

5 MODEL ADEQUACY

5.1 DEVIANCE RESIDUALS

The deviance residual can be used as a measure of the lack of fit of a model. For the Poisson model, the deviance residual is defined as

$$r^D_{A,T} = \text{sign}(Y_{A,T} - \hat{F}_{A,T})\sqrt{2[Y_{A,T}\log\frac{Y_{A,T}}{\hat{F}_{A,T}} - (Y_{A,T} - \hat{F}_{A,T})]} \quad (7)$$

where $r^D_{A,T}$ is the contribution to the deviance of the observation at age A and year T and

$$\text{sign}(x) = \begin{cases} -1 & \text{if} x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

The distribution of the deviance residuals should be approximately normal. For a good fit, about 95% of the deviance residuals should lie in the range [-2,2]. Figure 5 shows plots of the deviance residuals by age group and birth cohort. Out of the 126 deviance residuals resulting from fitting the model using *fminsearch*, 119 (94%) lie in the range [-2,2]. This increased to 121 (96%) when using the posterior medians of the Metropolis-Hastings algorithm, suggesting a satisfactory fit.

5.2 ALTERNATIVE STARTING VALUES FOR METROPOLIS-HASTINGS ALGORITHM

To check whether convergence was reached using the Metropolis-Hastings algorithm, Model A was refitted using different starting values. The starting values that were chosen were $k = 1$, $W = [1, 1, 1, 1, 1, 1, 1, 1]$, $D = [10000, 10000, 10000, 20, 20, 20, 20, -20, -20]$, $Peak = 1950$ and $Rate = 0$. Plots of the MCMC chains of model parameters using the two sets of starting values can be found in Appendix C. For each parameter, even when the starting values for the MCMC chain differed, the two chains appeared to converge to the same target distribution, suggesting that convergence was reached.

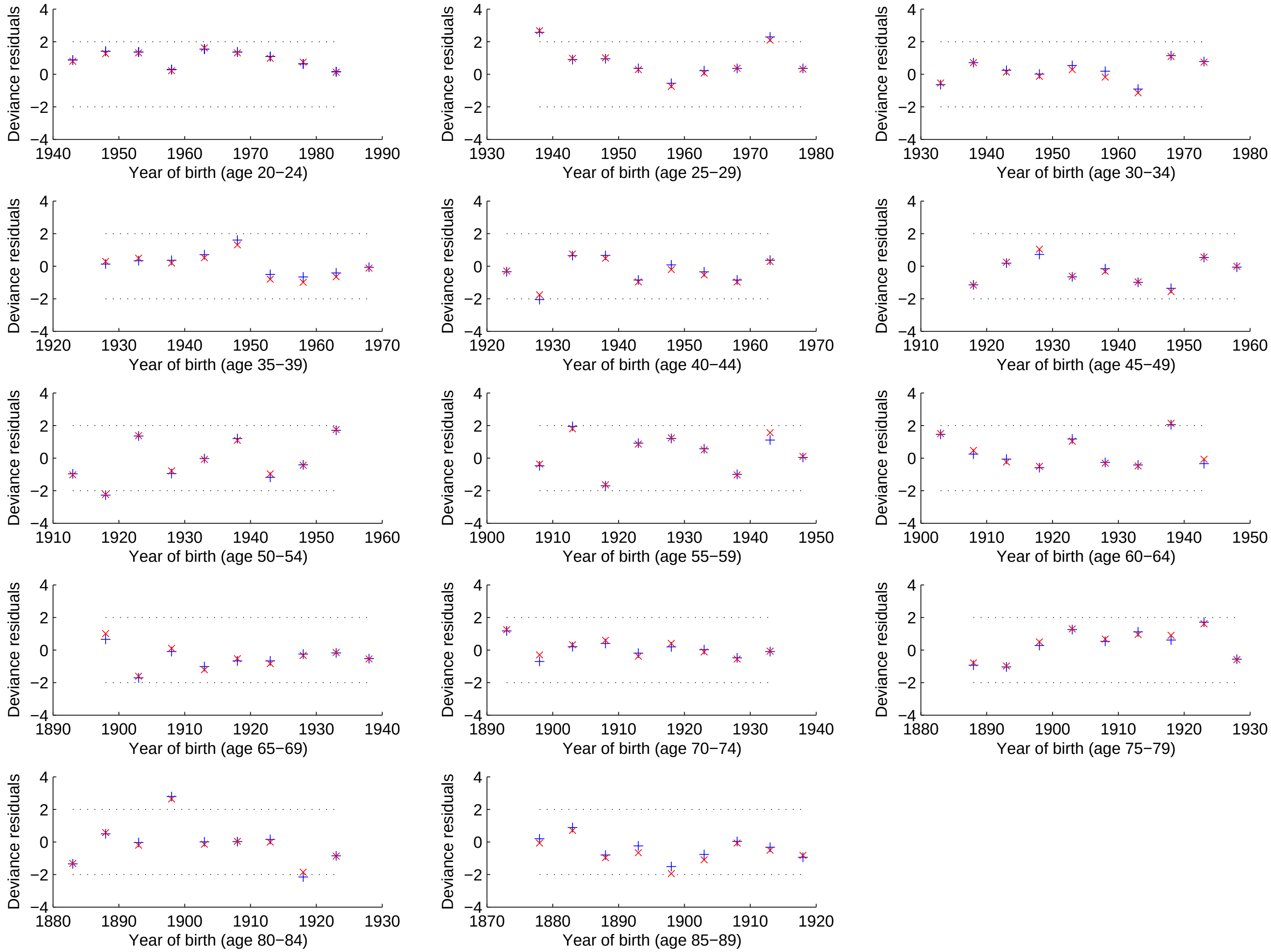

Figure 5 Deviance residuals by age group and birth cohort from fitting Model A using *fminsearch* (+) and MCMC (x)

5.3 FITTING THE MODEL TO PRE-2006 DATA

To test model adequacy, Model A was fitted to pre-2006 data using *fminsearch* and the predicted number of deaths up to 2006 were compared with the actual number of deaths. The projections resulting from fitting Model A to data up to 1987, 1992, 1997 and 2002 are shown in Table 7. Parameter estimates can be found in Tables 14 to 18 in Appendix 1.

Table 7 Projections based on pre-2006 data for males aged 20-89 years

	Projections (number of deaths)					Peak number	Year at which
For data up to year:	2002	2003	2004	2005	2006	of deaths	peak number occurs
1987	1626	1685	1741	1795	1840	2082	2015
1992	1616	1676	1731	1785	1831	2080	2016
1997	1500	1545	1586	1624	1654	1765	2013
2002	-	1617	1668	1718	1759	1976	2015
2006	-	-	-	-	-	2020	2016
Observed numbers	1559	1581	1650	1722	1705	-	-

Fitting Model A to data to 1987 and 1992 led to slightly higher predicted numbers of deaths for the years 2002 to 2006. The predicted peak was approximately 60 deaths higher than based on fitting Model A to data to 2006. In contrast, fitting to data to 1997 led to substantially lower predicted numbers of deaths than observed for the years 2002 to 2006 and a lower and earlier overall peak. Fitting to data to 2002 led to predicted values close to observed over the period 2003 to 2006. An inspection of the observed number of cases indicates that the increase in cases in 1997 on the previous year is small compared to the much larger increase seen in 1998, suggesting that the data for 1997 may be outliers with high leverage that will have influenced the fit of the model, when fitting to data to 1997. This suggests that once data for future years is made available, they should be used to update model parameters and provide updated projections as long as the data for the most recent years are not outliers.

6 DISCUSSION

This report has presented a statistical analysis of mesothelioma mortality in males based on Markov Chain Monte Carlo methods using the model formulated by Hodgson *et al.* (2005). Posterior medians and credible intervals for each of the model parameters have been calculated. Projections of mesothelioma mortality in males have also been made. Although the models used in this report are of the same form as that adopted by Hodgson *et al.* (2005), the use of Markov Chain Monte Carlo techniques has allowed credible intervals (in the Bayesian sense) for the parameters to be calculated using Bayesian methods, thus allowing more informed statistical inferences to be made. This was not possible using the optimisation approach adopted by Hodgson *et al.* (2005), which also made it more difficult to obtain prediction intervals for future mesothelioma mortality.

Refitting the model revealed an estimated population exposure curve with several local maxima, whereas the exposure curve estimated by Hodgson *et al.* (2005) increased monotonically prior to the peak year and decreased monotonically thereafter. The estimate of the global peak year of exposure was 1963 with local peaks around 1930 and 1950, after which exposure rapidly decreased. These peaks coincided with specific events that took place in Great Britain around the same time which had an impact on the use of asbestos. These events may explain the pattern observed in the population exposure profile to some extent, though actual changes in population exposure are not likely to be as extreme. The first peak coincides with the establishment of Asbestos Industry Regulations in 1931 and the Great Depression around the same period. The second peak coincides with the end of World War II after which shipyard activity - especially in naval yards - will have reduced. These features of the population exposure curve persist when refitting the model to observations of mortality to 2001, which suggests that the difference between the updated exposure curve and that of Hodgson *et al.* (2005) is because of the improved model-fitting approach rather than because of refinements to the model and additional observations of mortality. As mesothelioma is usually only diagnosed several decades after exposure to asbestos and as the peak year of mortality has yet to be reached, there is greater uncertainty in the estimates of population exposure from the mid-1960s onwards.

In the final model, the last year for which the population exposure is estimated is 1978. The extent of the population exposure beyond this point has limited impact on the predicted mesothelioma deaths within the range of years for which observations of mortality are available (up to 2006), and thus on the model fit. Furthermore, predictions of the scale and timing of the peak number of mesothelioma deaths are not highly dependent on exposure after the late 1970s.

However, the shape of the exposure curve after 1978 is required in order to use the model to make longer term predictions. Some limited investigation of different exposure curves suggests that a levelling off of the exposure in the late 1970s provides a marginally better fit than a continuing very steep decline in exposure.

However, such considerations cannot be used as grounds for preferring one exposure curve over another. Decisions about the shape of the exposure profile in this region must draw on other sources of evidence about the extent of population exposure more recently.

For the projections, the same assumptions about exposure beyond the year 2000 as in Hodgson *et al.* (2005) were used, and a linear decline in exposure between 1978 and 2000 was assumed. However, the prediction intervals of the long range projections incorporate only the uncertainty in the fitted model parameters, and not the unquantifiable but potentially considerable degree of additional uncertainty arising from the particular chosen shape of the exposure curve beyond 1978. For example, if the population exposure levelled off in 1978 and then continued indefinitely at this level (rather than continuing to decline, as has been assumed) the model predicts a much slower decline in mortality after the peak year, and consequently much larger estimates of the total mortality to year 2050 which exceed those based on our upper prediction interval.

Whilst this analysis confirms that the current model provides a good fit to the observations of mesothelioma mortality to date, and provides a reasonable basis for projections in the short term, it is much less clear whether it would provide a good basis for longer term projections, even if we could be more confident about the exposure curve beyond 1978. Male mortality to date is still dominated by the effect of substantial past occupational exposures and in these circumstances the model, in which mesothelioma risk depends on a power of time since first exposure, seems to fit the data well. However, future mortality will increasingly be a reflection of exposures in more recent times, and in this context mesothelioma risk might be better described in terms of the particular pattern of exposure rather than the time since each small component of exposure.

The background rate has been included in the models as fixed as well as an estimated parameter. The model which provided the best fit was one where the background rate was estimated at 1.08 cases per million amongst males, equivalent to 23 cases in 2006 amongst males aged 20 to 89. This is in good agreement with the value of 1% to 2% of total cases as suggested by Hodgson *et al.* (2005). Although the proportion of background cases in recent years amongst males has been small compared to the relatively large number of asbestos-related cases, the background cases will represent a larger proportion of all cases in future years when the number of asbestos-related cases will have fallen. It is thus important to take into account background cases in order to accurately make projections of mesothelioma mortality.

Hodgson *et al.* (2005) included a diagnostic trend parameter in their models which was estimated at 5% in their non-clearance model. The results of the analyses carried out in this report suggested that the inclusion of the diagnostic trend component in the model did not appear to improve the fit of the model. Although the best fitting model was one where the diagnostic trend component was excluded, this does not necessarily imply that the proportion of missed cases has remain unchanged over time; it may be due to the presence of confounders.

The peak number of mesothelioma deaths amongst all males reported in Hodgson *et al.* (2005) is around 1,857 deaths between 2011 and 2015 based on data up to 2001, which is lower than the peak of 2,038 (90% C.I. [1959,2123]) deaths amongst males in the year 2016 predicted in this report. More than half of the difference in the scale of the peak number of deaths is due to the use of updated projections of the future British population.

Different estimates of peak mesothelioma mortality have been predicted in other countries; in Australia, the peak is expected at around 700 cases per year in 2010 (Leigh and Driscoll, 2003). In

France, the peak of around 2,200 cases per year is expected some time after 2020 (Ilg *et al.*, 1998), whereas in the Netherlands (Segura *et al.*, 2003), up to 900 cases per year of pleural mesothelioma is expected around the year 2028. These projections, amongst others that have been made on mesothelioma mortality in Europe, indicate that although the number of deaths has been rapidly increasing in recent years, mortality may not reach a peak for several years.

7 CONCLUSIONS

General

- An estimate of the half-life H for clearance of asbestos from the lungs was difficult to obtain, however the fit of the model improved as H increased. Convergence to a point estimate (using *fminsearch*) or to a posterior distribution (using MCMC) of H was not reached, suggesting that H is infinitely large and that there is no clearance of asbestos once inhaled.

- There is a sharp increase in population exposure around the year 1930, with a decrease in the following few years. The population exposure then increases rapidly from the 1940s to the mid-1960s and reaches a peak exposure in 1963, rapidly decreasing thereafter.

Males

- Mesothelioma mortality amongst all males is predicted to peak at around 2,040 deaths with a 90% confidence interval (1959,2123) in 2016. Around 91,000 deaths are predicted to occur by 2050, with around 61,000 of these occurring from 2007 onwards.

- The relative exposure potentials in males aged 20 to 49 are much higher than in males outside this age range. In particular, males in the 30 to 39 age group were most at risk of exposure. The risk was very small for males under 15 and males aged 50 and over.

- The background rate was 1.08 cases per million, corresponding to around 23 cases in 2006 amongst males aged 20 to 89.

REFERENCES

Health & Safety Executive (2003), 'Mesothelioma mortality in Great Britain: Estimating the future burden.', *Health & Safety Executive, Bootle, UK.*.

Health Effects Institute (1991), 'Asbestos in Public and Commercial Buildings: A Literature Review and Synthesis of Current Knowledge.', *Health Effects Institute - Asbestos Research, Cambridge, MA.*.

Hastings, W. (1970), 'Monte Carlo Sampling Methods Using Markov Chains and Their Applications.', *Biometrika* **57**, 97-109.

Hodgson, J., McElvenny, D., Darnton, A., Price, M. and Peto, J. (2005), 'The expected burden of mesothelioma mortality in Great Britain from 2002 to 2050.', *British Journal of Cancer* **4**, 587-593.

Ilg, A., Bignon, J. and Valleron, A. (1998), 'Estimation of the past and future burden of mortality from mesothelioma in France.', *Occup Environ Med* **55**, 760-765.

Kjaergaard, J. and Andersson, M. (2000), 'Incidence rates of malignant mesothelioma in Denmark and predicted future number of cases among men.', *Scand J Work Environ Health* **26**, 112-117.

Leigh, J. and Driscoll, T. (2003), 'Malignant mesothelioma in Australia, 1945-2002.', *Int J Occup Environ Health* **9**, 206-217.

The MathWorks, Inc. (2008), 'MATLAB 7.7.0 (R2008b)'.

McElvenny, D., Darnton, A., Price, M. and Hodgson, J. (2005), 'Mesothelioma mortality in Great Britain from 1968-2001.', *Occupational Medicine* **55**, 79-87.

Peto, J., Matthews, F., Hodgson, J. and Jones, J. (1995), 'Continuing increase in mesothelioma mortality in Britain.', *Lancet* **345**, 535-539.

Price, B. and Ware, A. (2004), 'Mesothelioma trends in the United States: an update based on Surveillance, Epidemiology, and End Results data for 1973 through 2003.', *Am J Epidemiology* **159**, 107-112.

Rake, C., Gilham, C., Hatch, J., Darnton, A., Hodgson, J. and Peto, J. (2009), 'Occupational, domestic and environmental mesothelioma risks in the British population: a case-control study.', *Lancet* **345**, 535-539.

Segura, O., Burdorf, A. and Looman, C. (2003), 'Update of predictions of mortality from pleural mesothelioma in the Netherlands.', *Occup Environ Med* **60**, 50-55.

APPENDIX 1 ALTERNATIVE MODEL RESULTS

Tables 8, 9 and 10 show the results of fitting Models B, C and D to data up to 2006 using *fminsearch*. Tables 11, 12 and 13 show the results of fitting Models B, C and D using the Metropolis-Hastings algorithm for 35,000 iterations after a burn-in of 20,000 iterations. Tables 14 to 18 show the results of fitting Model A to data to 1987, 1992, 1997, 2001 and 2002 using *fminsearch*. Analysis by Hodgson *et al.* (2005) produced an estimate of 0.05 for α. For both models where *fminsearch* was used, however, the estimate of α was negative, suggesting that the number of missed cases decreases as we go backwards in time. The posterior median of α was also negative when using the Metropolis-Hastings algorithm for fitting model C, however the 90% credible interval includes zero suggesting high uncertainty.

Table 8 *fminsearch:* Parameter estimates for Model B

Parameter estimates			
k	2.49	Background rate	1.4 (fixed)
Maximum exposure year	1963	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1898 ($D(1)$)	0 (fixed)	0 to 4	0.00
1908 ($D(2)$)	1000 (fixed)	5 to 15	0.00
1918 ($D(3)$)	100000 (fixed)	16 to 19	0.30
1928 ($D(4)$)	-92.4	20 to 29	1.00 (baseline)
1938 ($D(5)$)	105.3	30 to 39	1.78
1948 ($D(6)$)	-26.1	40 to 49	1.51
1958 ($D(7)$)	38.1	50 to 59	0.07
1963	0 (by definition)	60 to 64	0.36
1968 ($D(8)$)	-7.8	65+	0.00
1978 ($D(9)$)	-16.8		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	2026	Peak year	2016
Deviance	214	Diagnostic trend	-

Table 9 *fminsearch:* Parameter estimates for Model C

Parameter estimates			
k	2.47	Background rate	1.21
Maximum exposure year	1963	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1898 ($D(1)$)	0 (fixed)	0 to 4	0.00
1908 ($D(2)$)	1000 (fixed)	5 to 15	0.00
1918 ($D(3)$)	100000 (fixed)	16 to 19	0.30
1928 ($D(4)$)	-92.8	20 to 29	1.00 (baseline)
1938 ($D(5)$)	106.1	30 to 39	1.80
1948 ($D(6)$)	-26.3	40 to 49	1.54
1958 ($D(7)$)	38.3	50 to 59	0.07
1963	0 (by definition)	60 to 64	0.48
1968 ($D(8)$)	-7.9	65+	0.00
1978 ($D(9)$)	-16.3		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	2012	Peak year	2016
Deviance	213	Diagnostic trend	-0.0017

Table 10 *fminsearch:* Parameter estimates for Model D

Parameter estimates and 95% CI			
k	2.55	Background rate	1.4 (fixed)
Maximum exposure year	1965	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1898 ($D(1)$)	0 (fixed)	0 to 4	0.00
1908 ($D(2)$)	1000 (fixed)	5 to 15	0.00
1918 ($D(3)$)	100000 (fixed)	16 to 19	0.20
1928 ($D(4)$)	-67.8	20 to 29	1.00 (baseline)
1938 ($D(5)$)	60.40	30 to 39	1.52
1948 ($D(6)$)	-13.0	40 to 49	1.44
1958 ($D(7)$)	27.0	50 to 59	0.03
1963	0 (by definition)	60 to 64	0.00
1968 ($D(8)$)	-13.2	65+	0.00
1978 ($D(9)$)	-10.3		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	2026	Peak year	2016
Deviance	214	Diagnostic trend	-0.12

Table 11 *Metropolis-Hastings:* Parameter estimates for Model B

Parameter estimates and 90% CI			
k	2.41 (2.27,2.56)	Background rate	1.40 (fixed)
Maximum exposure year	1965	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1900 ($D(1)$)	0 (fixed)	0 to 4	0.0014 (0.0001,0.0055)
1910 ($D(2)$)	1000 (fixed)	5 to 15	0.0025 (0.0002,0.0081)
1920 ($D(3)$)	100000 (fixed)	16 to 19	0.11 (0.016,0.246)
1930 ($D(4)$)	-70.6 (-91.8,-44.4)	20 to 29	1.00 (baseline)
1940 ($D(5)$)	65.1 (41.9,99.9)	30 to 39	1.66 (1.43,1.91)
1950 ($D(6)$)	-14.2 (-21.8,-8.1)	40 to 49	1.57 (1.21,1.98)
1960 ($D(7)$)	27.8 (22.8,34.3)	50 to 59	0.16 (0.014,0.44)
1965	0 (by definition)	60 to 64	0.41 (0.04,1.24)
1970 ($D(8)$)	-9.37 (-16.7,-4.0)	65+	0.41 (0.03,1.53)
1980 ($D(9)$)	-17.6 (-26.9,-5.9)		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	2006 (1923,2089)	Peak year	2016 (2015,2018)
Deviance	230 (218,239)	Diagnostic trend	-

Table 12 *Metropolis-Hastings:* Parameter estimates for Model C

Parameter estimates and 90% CI			
k	2.42 (2.28,2.58)	Background rate	1.15 (0.72,1.61)
Maximum exposure year	1964	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1899 ($D(1)$)	0 (fixed)	0 to 4	0.0017 (0.0001,0.0069)
1909 ($D(2)$)	1000 (fixed)	5 to 15	0.0028 (0.0002,0.0097)
1919 ($D(3)$)	100000 (fixed)	16 to 19	0.16 (0.028,0.31)
1929 ($D(4)$)	-75.7 (-99.3,-37.3)	20 to 29	1.00 (baseline)
1939 ($D(5)$)	72.6 (33.9,145.9)	30 to 39	1.66 (1.38,1.94)
1949 ($D(6)$)	-17.6 (-32.0,-3.1)	40 to 49	1.52 (1.17,1.88)
1959 ($D(7)$)	30.5 (18.2,42.3)	50 to 59	0.12 (0.01,0.39)
1964	0 (by definition)	60 to 64	0.42 (0.03,1.31)
1969 ($D(8)$)	-10.2 (-16.9,-4.3)	65+	0.42 (0.04,1.48)
1979 ($D(9)$)	-16.3 (-26.2,-5.4)		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	1983 (1900,2070)	Peak year	2016 (2014,1017)
Deviance	229 (214,238)	Diagnostic trend	-0.006 (-0.065,0.078)

Table 13 *Metropolis-Hastings:* Parameter estimates for Model D

Parameter estimates and 90% CI			
k	2.45 (2.30,2.60)	Background rate	1.4 (fixed)
Maximum exposure year	1964	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1899 ($D(1)$)	0 (fixed)	0 to 4	0.0015 (0.0001,0.0057)
1909 ($D(2)$)	1000 (fixed)	5 to 15	0.0019 (0.0001,0.0074)
1919 ($D(3)$)	100000 (fixed)	16 to 19	0.21 (0.069,0.34)
1929 ($D(4)$)	-87.2 (-97.4,-66.1)	20 to 29	1.00 (baseline)
1939 ($D(5)$)	91.8 (63.3,126.2)	30 to 39	1.74 (1.53,1.96)
1949 ($D(6)$)	-22.9 (-29.7,-15.3)	40 to 49	1.51 (1.19,1.89)
1959 ($D(7)$)	35.2 (27.8,40.7)	50 to 59	0.12 (0.011,0.37)
1964	0 (by definition)	60 to 64	0.46 (0.04,1.35)
1969 ($D(8)$)	-8.9 (-14.7,-2.3)	65+	0.39 (0.04,1.47)
1979 ($D(9)$)	-17.7 (-26.9,-8.2)		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	1988 (1904,2073)	Peak year	2016 (2014,2017)
Deviance	228 (214,236)	Diagnostic trend	-0.017 (-0.066,0.076)

Table 14 *fminsearch:* Parameter estimates for Model A (fitted to data to 1987)

Parameter estimates and 90% CI			
k	2.61	Background rate	1.80
Maximum exposure year	1965	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1900 ($D(1)$)	0 (fixed)	0 to 4	0.000
1910 ($D(2)$)	1000 (fixed)	5 to 15	0.0007
1920 ($D(3)$)	100000 (fixed)	16 to 19	0.23
1930 ($D(4)$)	-87.2	20 to 29	1.00 (baseline)
1940 ($D(5)$)	91.8	30 to 39	1.42
1950 ($D(6)$)	-22.9	40 to 49	1.76
1960 ($D(7)$)	35.2	50 to 59	0.04
1965	0 (by definition)	60 to 64	0.06
1970 ($D(8)$)	-8.9	65+	0.00
1980 ($D(9)$)	-17.7		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	2082	Peak year	2015
Deviance	107	Diagnostic trend	-

Table 15 *fminsearch:* Parameter estimates for Model A (fitted to data to 1992)

Parameter estimates and 90% CI			
k	2.51	Background rate	1.53
Maximum exposure year	1965	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1900 ($D(1)$)	0 (fixed)	0 to 4	0.000
1910 ($D(2)$)	1000 (fixed)	5 to 15	0.0007
1920 ($D(3)$)	100000 (fixed)	16 to 19	0.24
1930 ($D(4)$)	-67.5	20 to 29	1.00 (baseline)
1940 ($D(5)$)	60.4	30 to 39	1.71
1950 ($D(6)$)	-12.1	40 to 49	1.79
1960 ($D(7)$)	26.4	50 to 59	0.03
1965	0 (by definition)	60 to 64	0.06
1970 ($D(8)$)	-9.4	65+	0.00
1980 ($D(9)$)	-23.0		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	2080	Peak year	2016
Deviance	140	Diagnostic trend	-

Table 16 *fminsearch:* Parameter estimates for Model A (fitted to data to 1997)

Parameter estimates and 90% CI			
k	2.35	Background rate	1.11
Maximum exposure year	1965	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1900 ($D(1)$)	0 (fixed)	0 to 4	0.000
1910 ($D(2)$)	1000 (fixed)	5 to 15	0.0044
1920 ($D(3)$)	100000 (fixed)	16 to 19	0.20
1930 ($D(4)$)	-70.1	20 to 29	1.00 (baseline)
1940 ($D(5)$)	63.1	30 to 39	1.82
1950 ($D(6)$)	-14.0	40 to 49	1.88
1960 ($D(7)$)	28.4	50 to 59	0.18
1965	0 (by definition)	60 to 64	0.33
1970 ($D(8)$)	-10.2	65+	0.00
1980 ($D(9)$)	-36.1		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	1765	Peak year	2013
Deviance	155	Diagnostic trend	-

Table 17 *fminsearch:* Parameter estimates for Model A (fitted to data to 2001)

Parameter estimates and 90% CI			
k	2.48	Background rate	1.52
Maximum exposure year	1965	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1900 ($D(1)$)	0 (fixed)	0 to 4	0.000
1910 ($D(2)$)	1000 (fixed)	5 to 15	0.0006
1920 ($D(3)$)	100000 (fixed)	16 to 19	0.21
1930 ($D(4)$)	-66.0	20 to 29	1.00 (baseline)
1940 ($D(5)$)	58.9	30 to 39	1.58
1950 ($D(6)$)	-12.2	40 to 49	1.68
1960 ($D(7)$)	26.5	50 to 59	0.09
1965	0 (by definition)	60 to 64	0.05
1970 ($D(8)$)	-13.1	65+	0.00
1980 ($D(9)$)	-16.7		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	1969	Peak year	2015
Deviance	182	Diagnostic trend	-

Table 18 *fminsearch:* Parameter estimates for Model A (fitted to data to 2002)

Parameter estimates and 90% CI			
k	2.51	Background rate	1.52
Maximum exposure year	1965	Half-life (years)	1000000 (fixed)
Change in exposure index (% per year) in...		Relative exposure potential by age group	
1900 ($D(1)$)	0 (fixed)	0 to 4	0.000
1910 ($D(2)$)	1000 (fixed)	5 to 15	0.0007
1920 ($D(3)$)	100000 (fixed)	16 to 19	0.21
1930 ($D(4)$)	-67.2	20 to 29	1.00 (baseline)
1940 ($D(5)$)	59.7	30 to 39	1.58
1950 ($D(6)$)	-12.5	40 to 49	1.56
1960 ($D(7)$)	26.4	50 to 59	0.04
1965	0 (by definition)	60 to 64	0.06
1970 ($D(8)$)	-11.2	65+	0.00
1980 ($D(9)$)	-24.6		
Projections of future mesothelioma deaths in males aged 20-89			
Peak level	1976	Peak year	2015
Deviance	181	Diagnostic trend	-

APPENDIX 2 DEVIANCE PLOTS

Deviance plots for each of the parameters can be found in Figures 6 to 10. The deviance values were calculated using the estimates obtained from fitting Model C using *fminsearch* on data to 2006 for males aged 20 to 89.

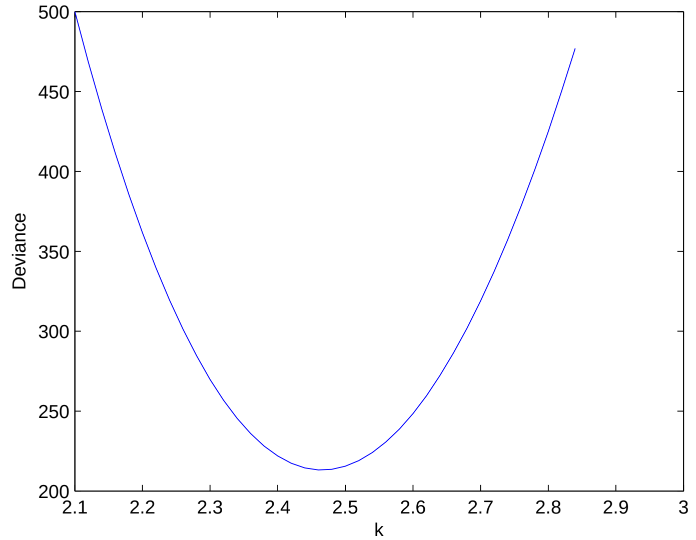

Figure 6 Deviance plot for k using *fminsearch* for Model C

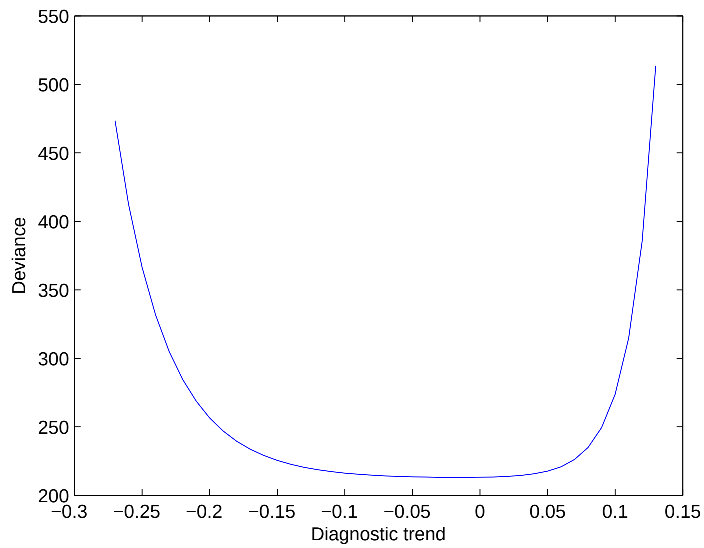

Figure 7 Deviance plot for α using *fminsearch* for Model C

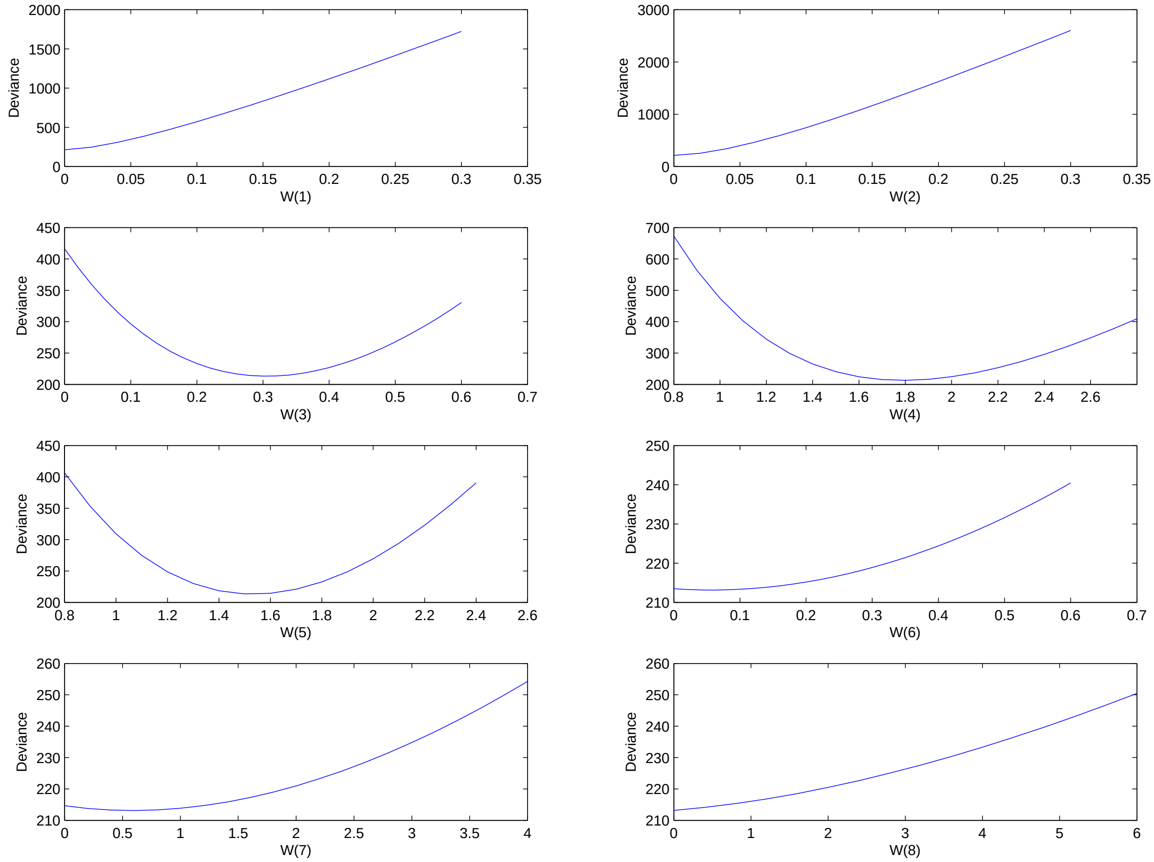

Figure 8 Deviance plot for W using *fminsearch* for Model C

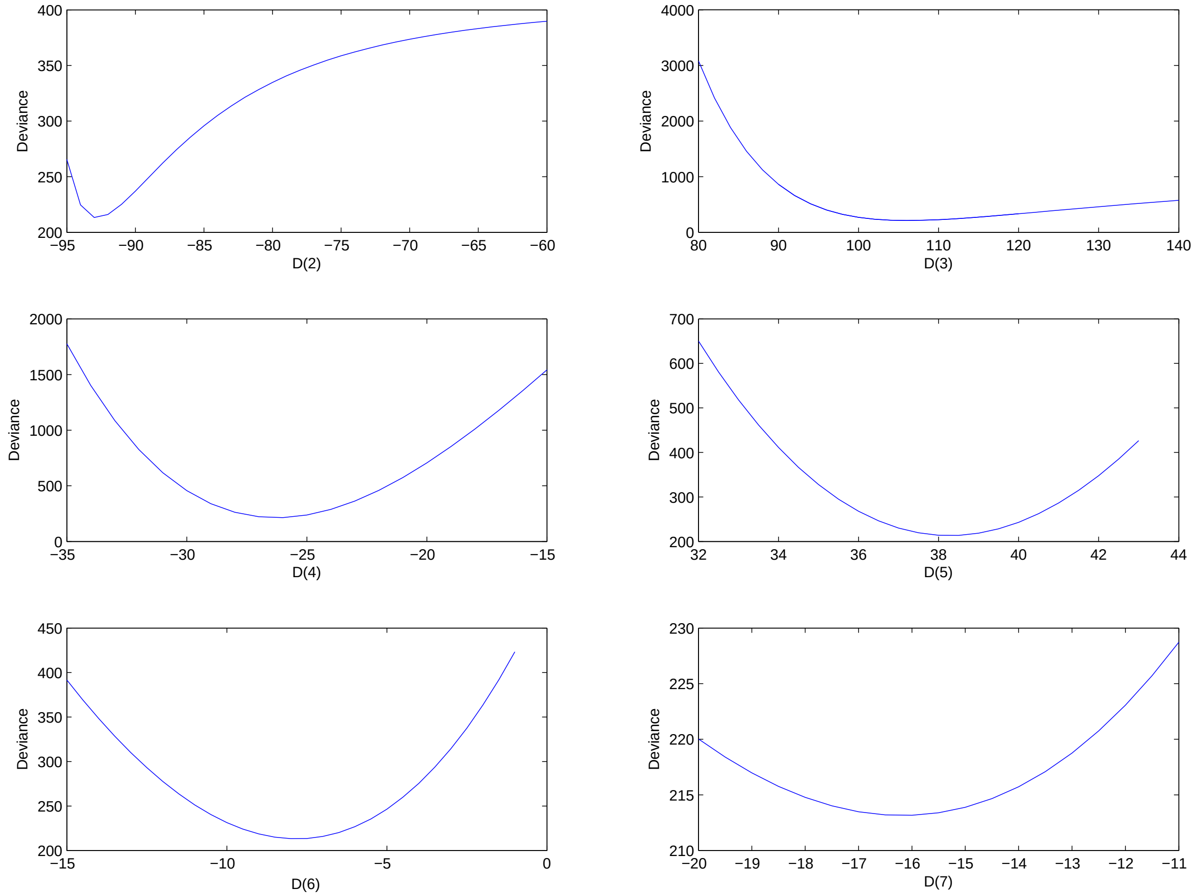

Figure 9 Deviance plot for D using *fminsearch* for Model C

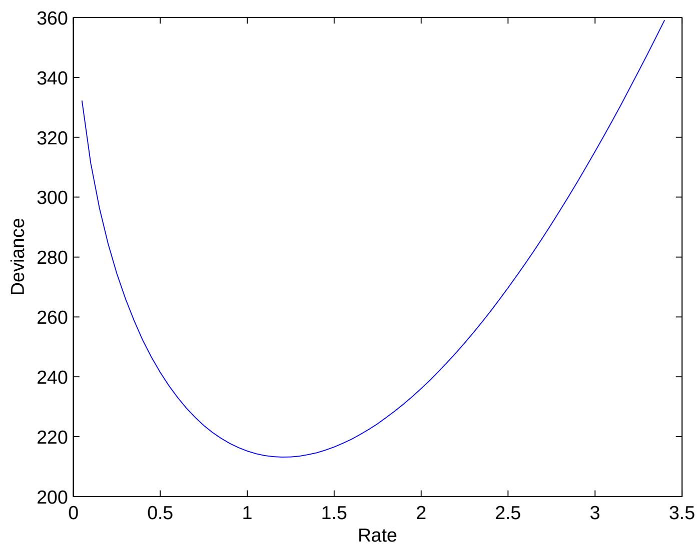

Figure 10 Deviance plot for $Rate$ using *fminsearch* for Model C

APPENDIX 3 CORRELATION PLOTS

Correlation plots for pairs of parameter values from the MCMC chain for Model A (fitted to data on males aged 20 to 89) can be found in Figures 11 to 15. There appears to be little correlation between parameters apart from the correlations between k and $W(3)$, $W(4)$, $W(5)$ and $W(6)$. As the value of k increases, $W(3)$ is seen to increase whereas $W(4)$, $W(5)$ and $W(6)$ are seen to decrease. These correlations, although interesting to note, would not have affected the posterior distribution statistics presented in this report.

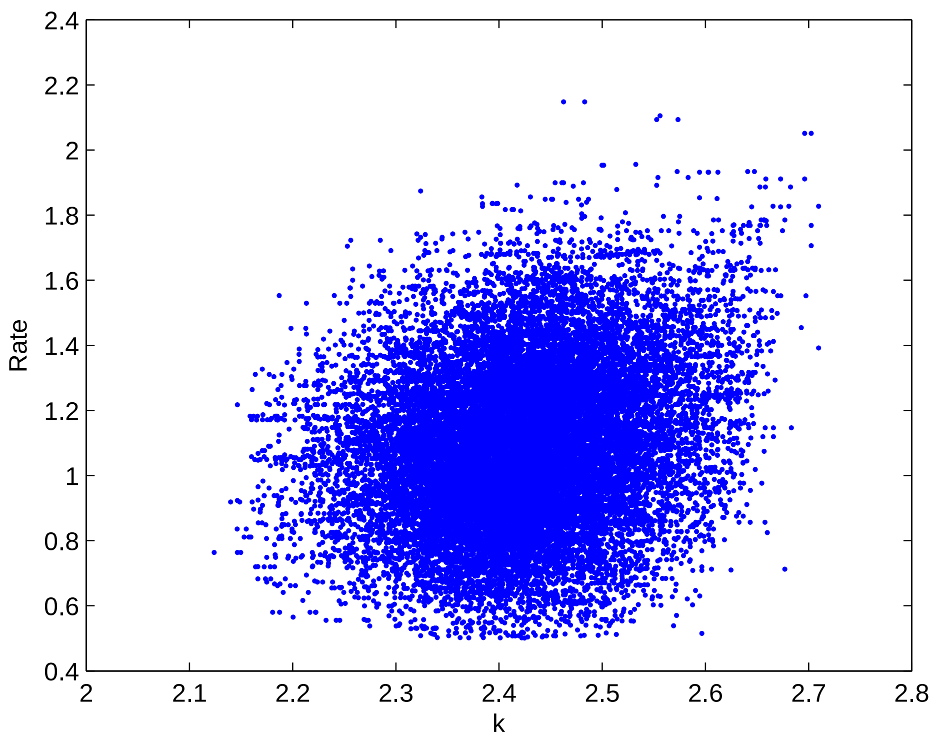

Figure 11 Correlation plot between k and $Rate$ using Metropolis-Hastings

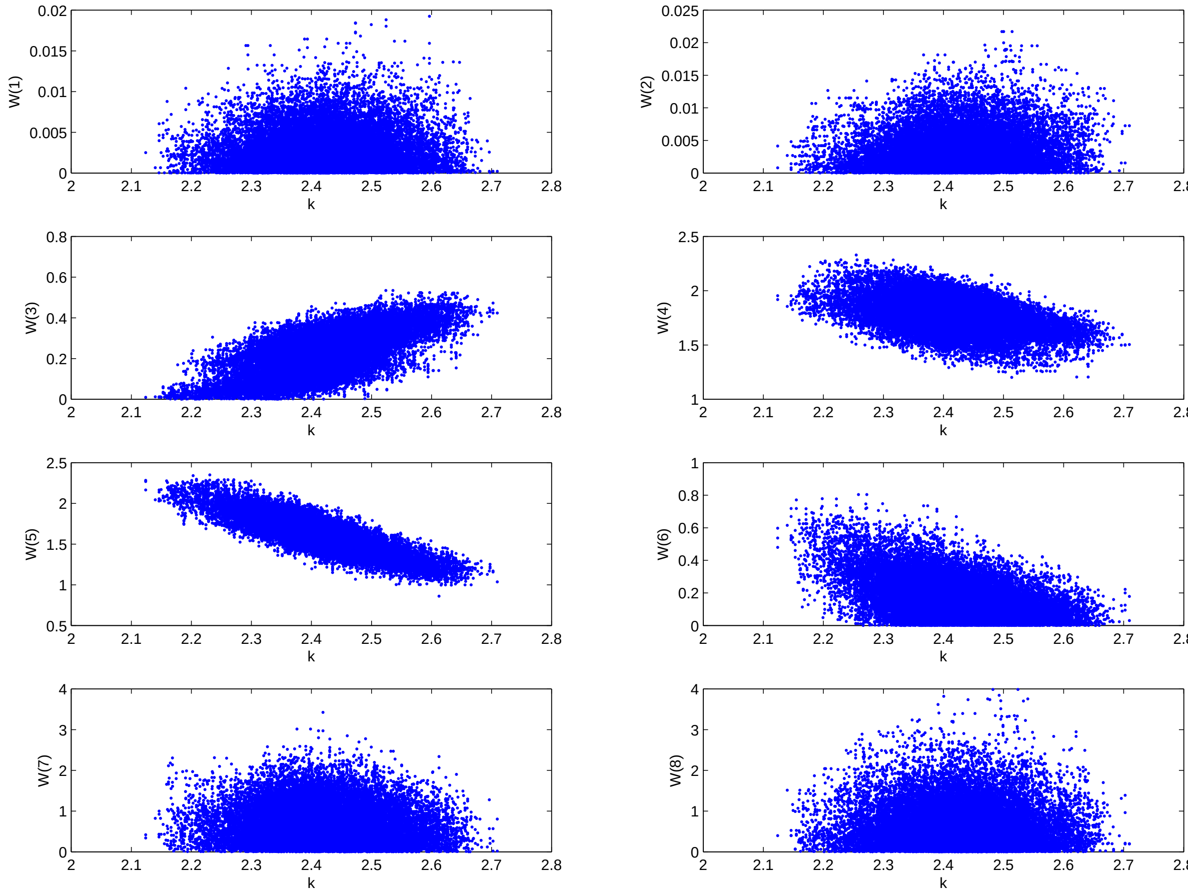

Figure 12 Correlation plot between k and W using Metropolis-Hastings

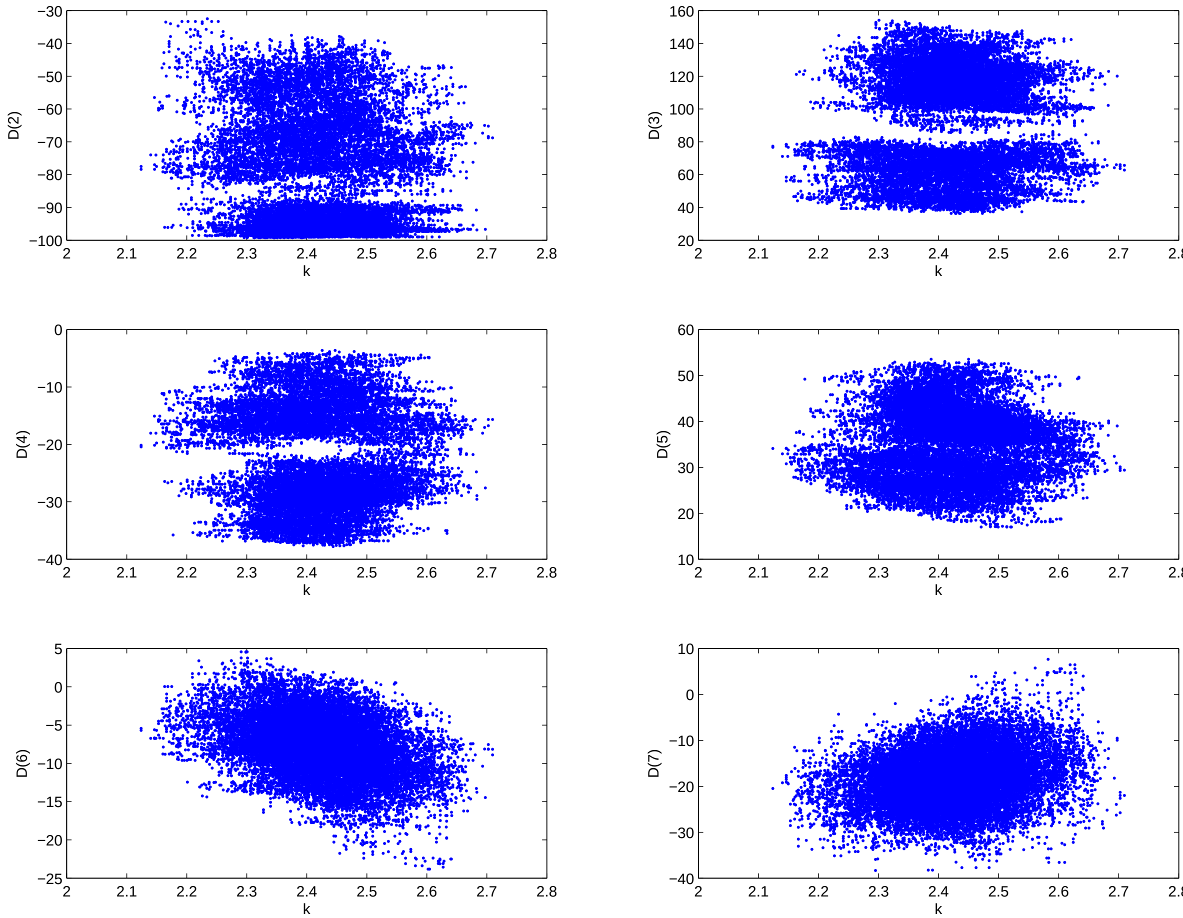

Figure 13 Correlation plot between k and D using Metropolis-Hastings

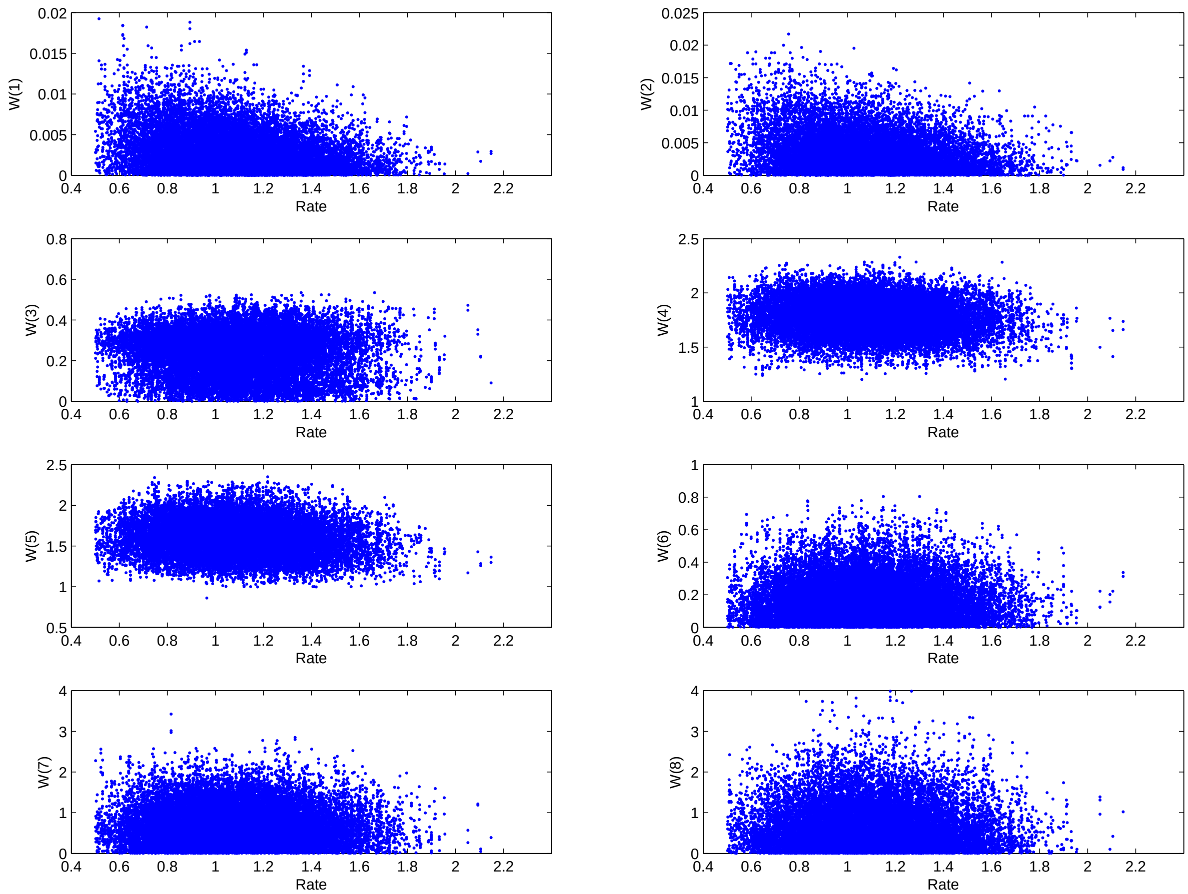

Figure 14 Correlation plot between $Rate$ and W using Metropolis-Hastings

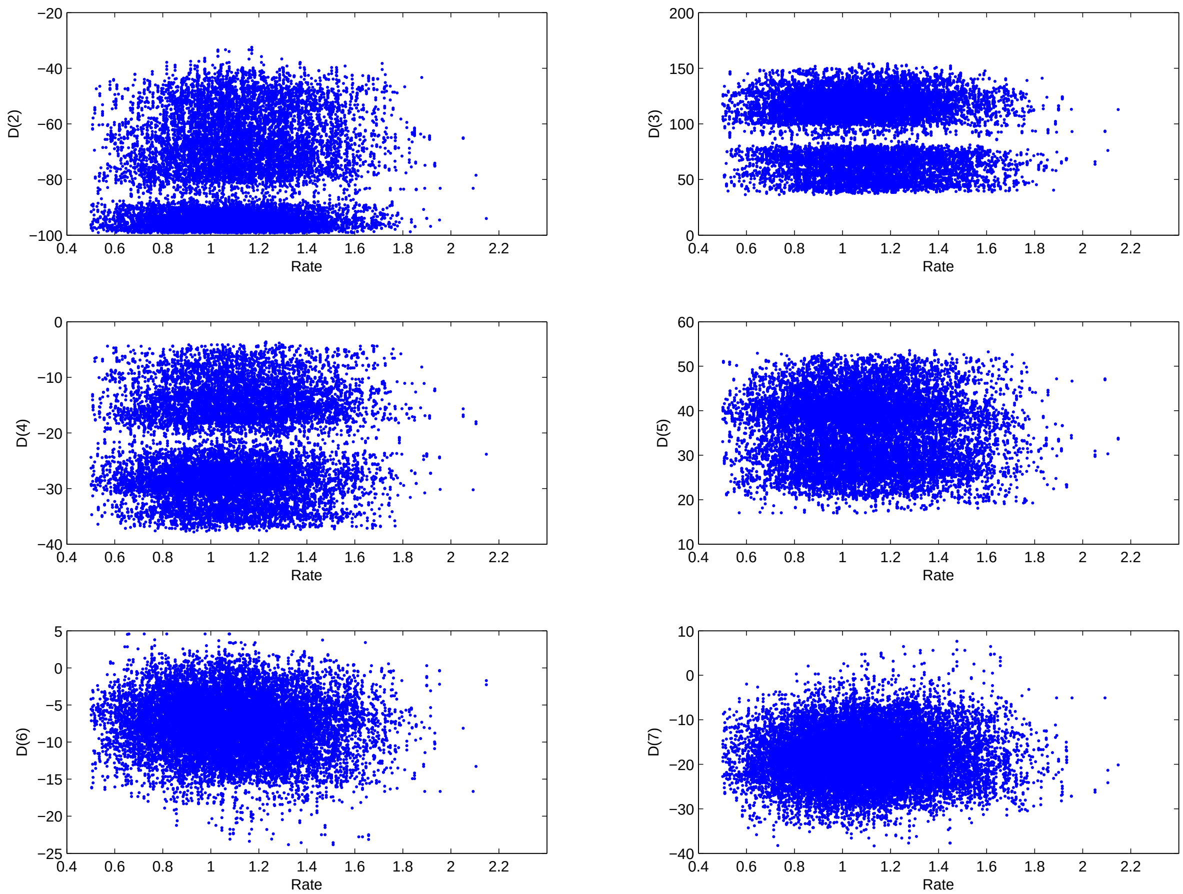

Figure 15 Correlation plot between $Rate$ and D using Metropolis-Hastings

APPENDIX 4 METROPOLIS-HASTINGS CHAINS

The MCMC chains using two sets of starting values for the model parameters in Model A (fitted to data on males aged 20 to 89) can be found in Figures 16 to 20. Even when the starting values differ between chains, the two chains eventually converge to approximately the same distribution.

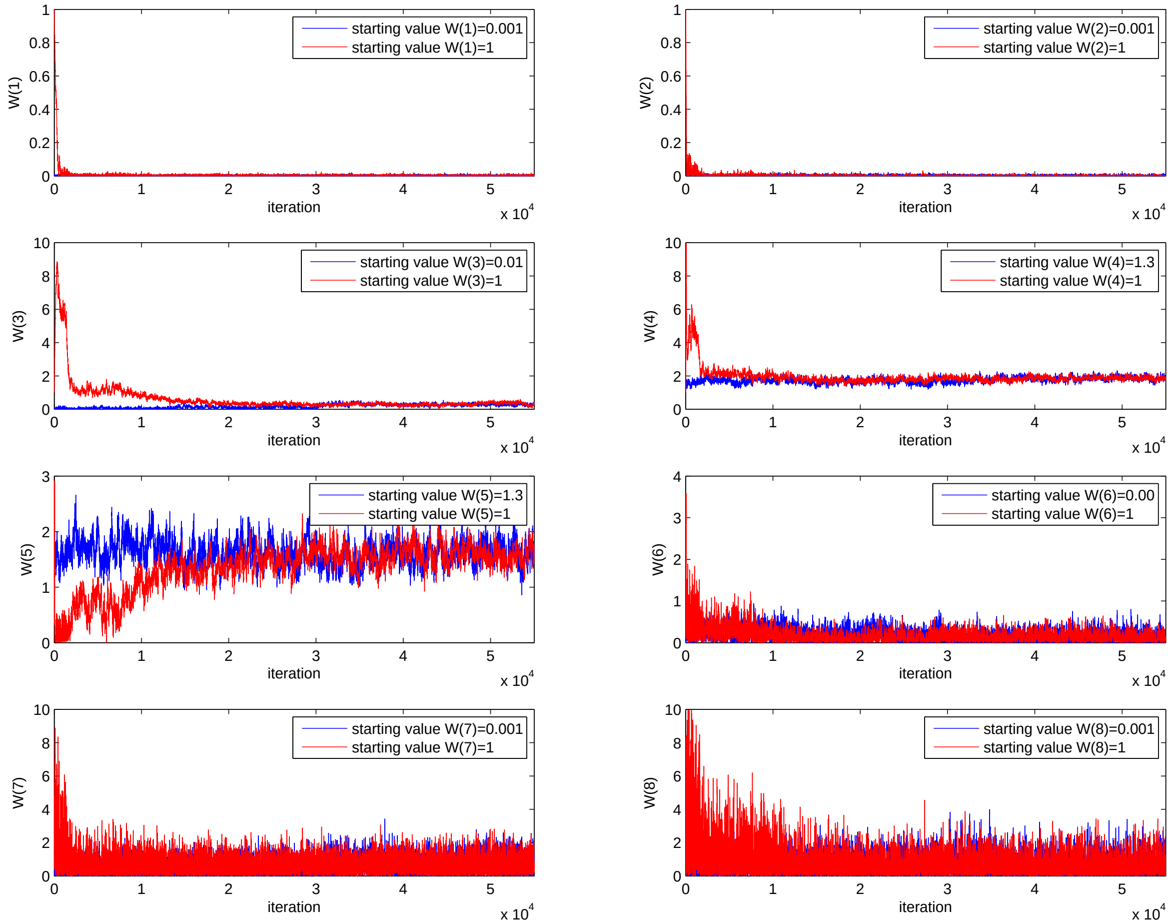

Figure 16 Metropolis-Hastings chain of W values using 2 different sets of starting values

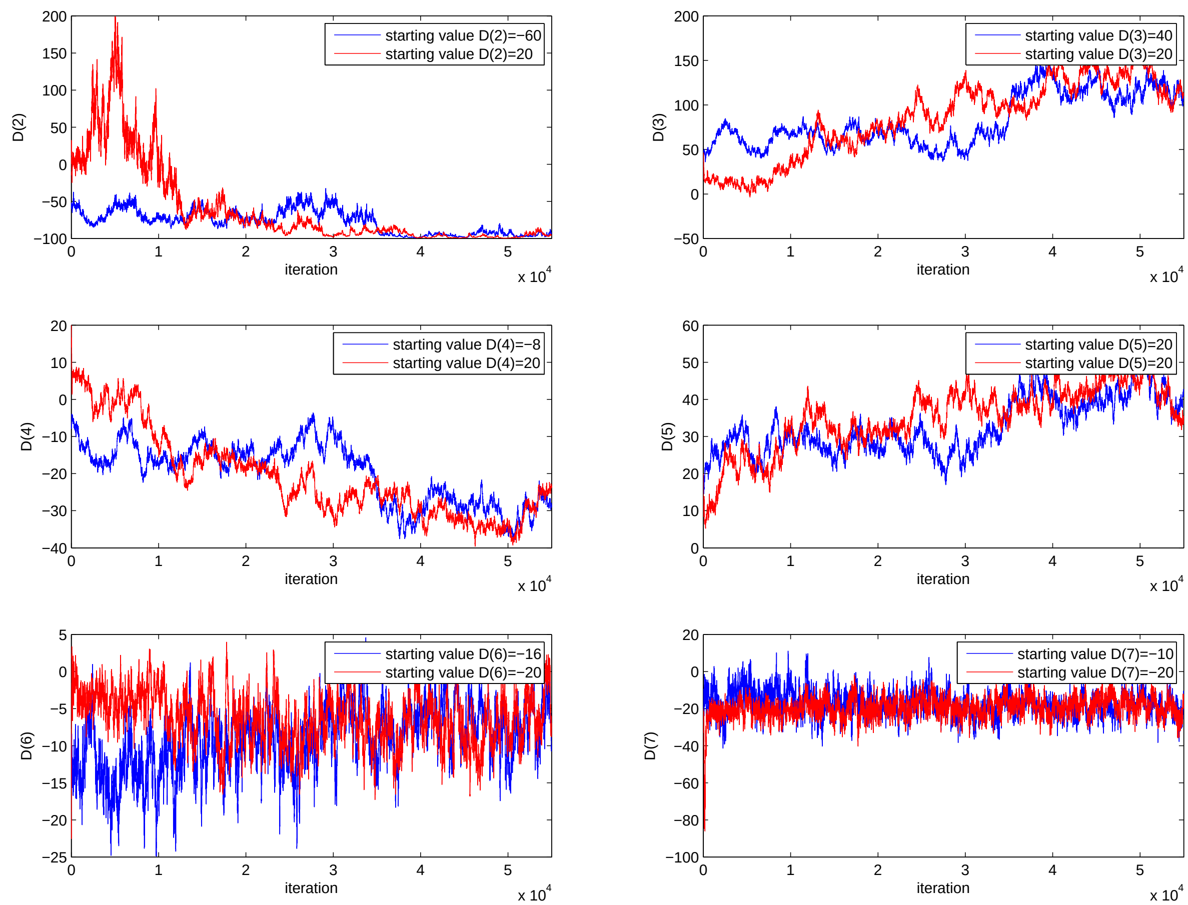

Figure 17 Metropolis-Hastings chain of D values using 2 different sets of starting values

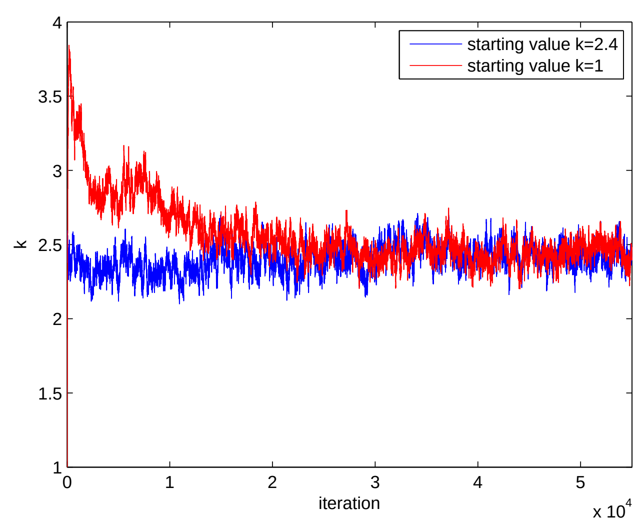

Figure 18 Metropolis-Hastings chain of k values using 2 different sets of starting values

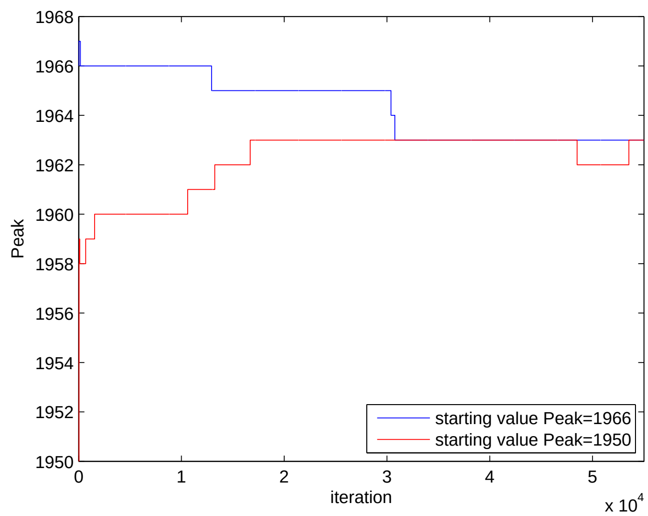

Figure 19 Metropolis-Hastings chain of $Peak$ values using 2 different sets of starting values

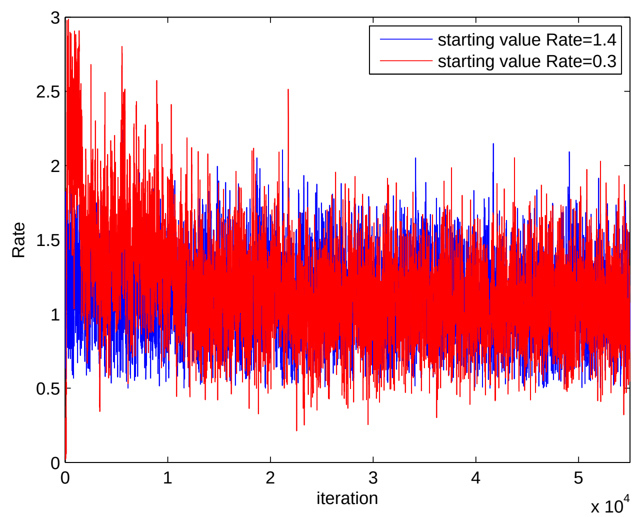

Figure 20 Metropolis-Hastings chain of $Rate$ values using 2 different sets of starting values

HSE

Health and Safety
Executive

Projection of mesothelioma mortality in Great Britain

There has been an increase in mesothelioma mortality in Great Britain, with 1705 deaths recorded in 2006. In 2005, a statistical model was developed based on a simple birth-cohort model, which assumes that the risk of mesothelioma depends on age and years of exposure and that an individual's asbestos exposure depends on the year of exposure. An optimisation technique was used to fit the model and a profile of the population exposure was estimated. Projections of the future burden of mesothelioma mortality were calculated, however statistical uncertainties in the formulation of the model could not be taken into account. In this report, the model has been refined and refitted using the MATLAB's fminsearch function and the Metropolis-Hastings algorithm, a Markov Chain Monte Carlo technique. Credible intervals for model parameters as well as prediction intervals for future cases of mortality amongst males are presented. Mortality amongst all males is expected to keep increasing, reaching a peak at around 2,040 deaths in the year 2016, with a rapid decline following the peak year. Around 91,000 deaths are predicted to occur by 2050 with around 61,000 of these occurring from 2007 onwards.

This report and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect HSE policy.

RR728

www.hse.gov.uk