



Institute of Actuaries of Australia

# Actuarial projections for mesothelioma: an epidemiological perspective

Prepared by Mark Clements, Geoffrey Berry  
and Jill Shi

Presented to the Institute of Actuaries of Australia  
XIth Accident Compensation Seminar 1-4 April 2007  
Grand Hyatt Melbourne, Australia

*This paper has been prepared for the Institute of Actuaries of Australia's (Institute) XIth Accident Compensation Seminar 2007.*

*The Institute Council wishes it to be understood that opinions put forward herein are not necessarily those of the Institute and the Council is not responsible for those opinions.*

© Mark Clements, Geoffrey Berry and Jill Shi

The Institute will ensure that all reproductions of the paper acknowledge the Author/s as the author/s, and include the above copyright statement:

The Institute of Actuaries of Australia  
Level 7 Challis House 4 Martin Place  
Sydney NSW Australia 2000  
Telephone: +61 2 9233 3466 Facsimile: +61 2 9233 3446  
Email: [actuaries@actuaries.asn.au](mailto:actuaries@actuaries.asn.au) Website: [www.actuaries.asn.au](http://www.actuaries.asn.au)

## Introduction

There is substantial workers' compensation and product and public liability due to asbestos exposure and subsequent disease outcomes. Such diseases include mesothelioma, lung cancer, asbestosis and pleural plaques. Prediction of the future burden for asbestos-related diseases is both important and challenging. The challenges arise due to uncertainty with: a) describing the historical asbestos exposure; b) predicting the future propensity to claim; and c) predicting the pattern of future costs. Moreover, there is some complexity in the epidemiological relationship between asbestos exposure and the related diseases, where there is a substantial lag between exposure and disease incidence. In Australia, the numbers of cases of asbestos-related disease are continuing to rise, although population-level asbestos exposure declined markedly during the 1970s and 1980s. To take account of these patterns, Australian actuaries have taken a variety of approaches to predicting asbestos-related diseases. Huszczo et al (2004) provided a useful review of the Australian actuarial setting for asbestos-related diseases. There is particular interest in mesothelioma, which is both a substantial fraction of the total disease burden due to asbestos and a useful marker for population-level asbestos exposure.

In the following paper, we focus on the prediction of population-level mesothelioma incidence and asbestos exposure. Any such predictions may be useful for actuarial predictions for broad portfolios, such as former James Hardie entities or the Dust Diseases Board, or portfolios that have a risk profile that is broadly similar to the population-wide asbestos exposure. Importantly, these predictions may not be useful for specific industries that do not follow the population-wide risk profile; as an example, the predictions are inappropriate for the mining industry, where mining of asbestos ceased 10-20 years earlier than the consumption of asbestos products.

Note, however, that the pattern of incidence is *not* directly applicable to the pattern of claims: further effort is required by the actuary to calculate the propensity to claim, which is likely to be a complex and dynamic process. However, we argue that predictions of mesothelioma incidence are important for characterising asbestos exposure and mesothelioma claims. The main advantage of mesothelioma incidence is that it is *predictable*: given that there is a smooth biological process between asbestos exposure and mesothelioma incidence, we can theoretically reconstruct asbestos exposure from historical mesothelioma incidence and then use the reconstructed exposure distribution to predict future mesothelioma incidence.

To outline the paper: first, we develop a theoretical framework for population-level modelling of mesothelioma incidence; then we review several earlier models for mesothelioma incidence predictions, both theoretically and empirically, using male incidence of mesothelioma for Australia and New South Wales; finally, we review the “best” approaches for mesothelioma incidence predictions.

## Methods

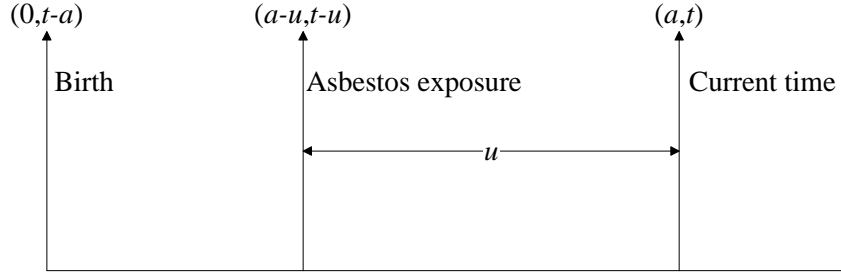
We first describe an individual-level rate model for mesothelioma incidence which depends on asbestos exposure; we then apply this model to the population context to develop a general framework for predictions of mesothelioma incidence.

### Individual-level rate model

From aetiological studies, we can make the following general observations about the relationship between asbestos exposure and mesothelioma incidence (Peto et al 1982; Berry 1991, 1999). First, the rate of mesothelioma is approximately proportional to linear dose of asbestos. Second, mesothelioma incidence depends on time from initial exposure to asbestos; given time from initial exposure to asbestos, mesothelioma incidence is independent of age. Third, mesothelioma incidence rises by a power of time from initial exposure to asbestos.

Fourth, there is evidence for clearance of asbestos from the lungs, which dampens the power relationship with time from initial exposure to asbestos. Fifth, there is a period of some years between the time of initial malignancy for the first cell through to clinical diagnosis of mesothelioma. Sixth, survival from mesothelioma is poor and death is rapid, typically occurring within the first year from diagnosis (Leigh et al 1991).

**Figure 1: Notation for time, with the tuple  $(a,t)$  representing age  $(a)$  and calendar period  $(t)$**



For an individual  $i$  aged  $a$  years at time  $t$ , we can consider a time  $u$  since exposure to asbestos (see Figure 1). Note that  $t-a$  is the year of birth, representing the birth cohort. Given the above observations, the general form of the rate function for mesothelioma incidence given a level of exposure to asbestos can be described by

$$rate_i(a, t) = \int_0^a dose_i(a-u, t-u)g(u)du$$

where  $dose_i(a-u, t-u)$  represents the linear effect of asbestos dose for an individual  $i$  exposed to asbestos at time  $t-u$  at age  $a-u$ , and  $g(u)$  represents a function of time since exposure. Under the Armitage-Doll model of carcinogenesis with an extension for clearance of asbestos from the lungs (Berry 1999), the full form for  $g(u)$  is

$$g(u) = \beta(u-\tau)^k e^{-\lambda(u-\tau)}$$

where  $\beta$  is the normalising constant,  $\tau$  is the lag time for time from initial malignancy to clinical diagnosis,  $k$  is a power term and  $\lambda$  is the rate of asbestos clearance from the lungs.

For completeness, we could also include a background rate of mesothelioma, however the background rate is typically considered to be negligible compared with specific asbestos exposures. We could also include exposure for different types of asbestos and for different periods and levels of exposure. It is difficult to tease apart the specific composition of an asbestos exposure; there is also some contention as to the risk for exposures with different compositions. Aetiological studies have typically used time since initial exposure as the primary time measure, with a univariate measure of dose. Such a time measure is appropriate for a short period of exposure. For a longer period of exposure, we can make the observation that the Armitage-Doll model of carcinogenesis is additive; this allows for disaggregating the rate by year of exposure, giving:

$$rate_i(a, t, u) = dose_i(a-u, t-u)g(u) \tag{1}$$

We can now consider Equation (1) in the context of populations.

### Population-level rate model

To develop a population-level rate model, we can take advantage of the linear form for dose: for a given age  $a$  and time  $t$  with a given time  $u$  since exposure, the mean rate across doses will be equal to the mean dose times the function  $g(u)$ :

$$rate(a, t, u) = \overline{dose}(a-u, t-u)g(u) \tag{2}$$

However, the mean dose function is a bivariate function, describing a dose “surface”, which may be overly complicated. Following a suggestion by Hodgson and colleagues (2005), we can further simplify the function by assuming that dose is proportional to a function  $W()$  for

age at exposure and proportional to a function  $D()$  for time at exposure, that is,  $dose(a - u, t - u) = W(a - u)D(t - u)$ . Then the rate function becomes

$$rate(a, t, u) = W(a - u)D(t - u)g(u)$$

By summing across the time from exposure to asbestos, we can calculate the population rate, which takes the rather elegant form:

$$rate(a, t) = \int_0^a W(a - u)D(t - u)g(u)du \quad (3)$$

Predictions for the number of cases proceeds by weighting the rate function  $rate(a, t)$  by the at-risk population aged  $a$  at time  $t$ :

$$cases(a, t) = rate(a, t) \times Pop(a, t)$$

The at-risk population could also be calculated by the size of an initial population times the probability of survival to time  $t$  for that population. Total cases for a year can be estimated by summing the cases across age groups:

$$cases(t) = \int_0^{\infty} rate(a, t) \times Pop(a, t) da$$

For fitting a model to observed mortality or incidence rates, we typically assume that the number of cases has a Poisson distribution (Brillinger, 1986).

## Uncertainty

Given the small numbers of cases available to fit these models, it is important to also carefully represent the uncertainty in any predictions. We separate out *model uncertainty*, which depends on whether we have selected a good model, from *statistical uncertainty*, which depends on the statistical imprecision for a given model. For simplicity, we restrict our attention to prediction of mesothelioma incidence; there are further sources of uncertainty when attempting to calculate claims and costs.

For assessing model uncertainty, we note that aetiological studies are often based on relative small numbers of cases. Moreover, there are many biases in aetiological studies, with substantial variation in those biases between studies. As a consequence, there is considerable uncertainty as to the “best” model, with several different models providing comparable fits within data, whilst potentially providing qualitatively different predictions. The model we have selected is based on our understanding of the available data. As a pointer for future developments, the function  $g(u)$  has historically been based on the Armitage-Doll model of carcinogenesis; meanwhile, an alternative class of models for carcinogenesis based on clonal expansion may provide other useful predictions.

For a given model, there is substantial uncertainty when fitting that model to observed data. Such uncertainty can be ascribed to a) the relatively small numbers of cases of mesothelioma observed in the population and b) the relative complexity of the models, which are attempting to reconstruct some measure of exposure. The exposure reconstruction is particularly sensitive to rates at younger ages, where there is paucity of information. The complex models may suffer from over-parameterisation and may be non-linear, requiring non-linear optimisation of a Poisson likelihood. For models that are stratified by age, the statistical analysis is further complicated by the need to summarise across age groups; the modelled age-specific estimates are then correlated and variance estimation becomes increasingly difficult. Possible approaches to deal with these complications include variance estimation using the bootstrap. Note that interval estimation can be for the mean (“confidence intervals”) or for the individual predictions (“prediction intervals”), which take account of Poisson variability from year to year.

## Software

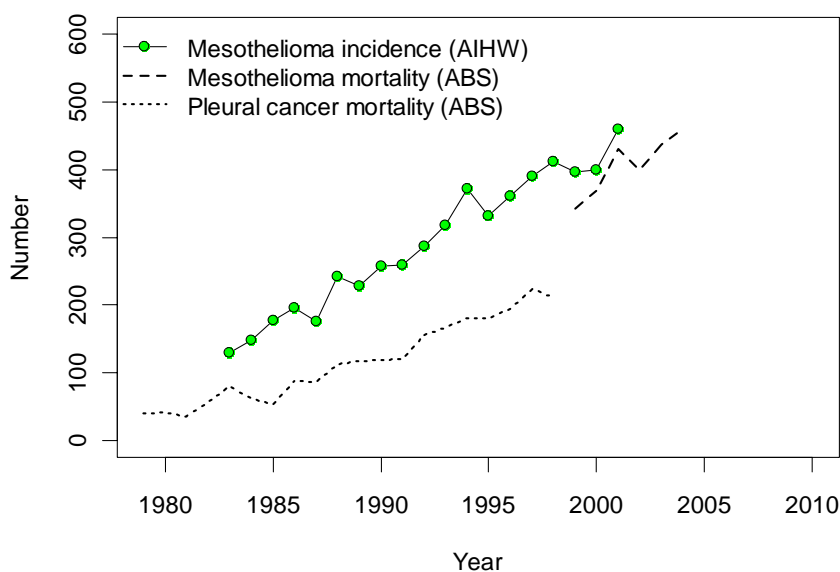
Usefully, available software can now readily deal with fitting non-linear models, combined with bootstrap interval estimation. We have found that the open source statistical software “R” readily and flexibly supports all of these analyses using the `mle()` function for maximum likelihood estimation and the “boot” package for bootstrapping. There is a range of alternative approaches. In SAS, maximum likelihood estimation for non-linear models can be undertaken in PROC NLP (SAS/OR) and PROC IML; PROC NLIN can also be used for model fitting via iteratively re-weighted least squares. In Microsoft Excel, the Solver, potentially in combination with VBA, would be sufficiently flexible to fit many of these models, although implementing the bootstrap estimation would be challenging. The econometric software Stata has a pleasant likelihood optimiser that would facilitate most of these analyses. Most matrix programs have flexible optimisers, including S-Plus, GAUSS and Matlab. Finally, there are powerful libraries available for Fortran, C, C++ and Ada for writing compiled code to fit these analyses.

## Data sources

Mesothelioma incidence is reported by state and territory cancer registries. Australia-wide estimates have not been reported by the Australian Institute of Health and Welfare since 2004 for the 2001 calendar year; interestingly, 2002 and 2003 data are expected to be reported within the next few months. More timely data are available from individual state cancer registries. The patterns for mortality and incidence are very similar as expected survival from mesothelioma is short and poor (Leigh et al 1991). The incidence pattern for males is quite different from females, where male rates are considerably higher; we restrict further attention to males only.

Deaths due to mesothelioma mortality are now regularly reported from the Australian Bureau of statistics using ICD-10; however, ICD-9 only reported pleural cancer deaths, which comprise 70-90% of mesothelioma deaths. As a consequence, we have a long time series for pleural cancer mortality, with a break in the time series in 1998 with the change in coding to ICD-10 (see Figure 2). Data for mesothelioma incidence from NSW are currently available from 1972 through to 2004.

**Figure 2: Mesothelioma incidence and mortality and pleural cancer mortality, Australian males**



As a summary of the available data on population-wide incidence and mortality, mesothelioma incidence data from New South Wales provide a longer and more timely series,

albeit with smaller numbers. Analysis of Australian data is restricted by the length of the relevant time series, as incidence data are only available from 1983 through to 2001 and the time series for mortality has been split by changes in coding, preventing analysis using the most recent data. The following analyses will use mesothelioma incidence for males for Australia and New South Wales.

**Results: Model comparisons**

In the excellent monograph by Stallard and colleagues (2005) on the Manville Asbestos Case, the authors use a taxonomy based on whether asbestos exposure is a) directly measured or b) indirectly modelled. Population-wide measures of mesothelioma incidence using both of these approaches were applied to the United States in the early 1980s (for a review, see Stallard et al 2005). Huszczo and colleagues (2004) proposed a similar taxonomy that included exposure models, associated with directly measured exposure, and population regression modelling, associated with indirectly modelled asbestos exposure.

As noted by several authors (Atkins et al 1996; Huszczo et al 2004), population-level predictions have the inherent problem of not having exposure attached to each incident case, so that the exposure-incidence relationship can not be directly analysed or modelled. Importantly, indirect modelling of the exposure-incidence relationship has proved to be very fruitful; given the paucity of good exposure data, we argue that indirect modelling now constitutes state-of-the-art for mesothelioma prediction of population-wide patterns (e.g. Stallard et al 2005; Hodgson et al 2005). Indirect estimation has also been used successfully for large portfolios (Stallard et al 2005).

As a further dimension for any taxonomy, we propose a distinction should be drawn between i) simple calibration of a curve to a data set, which is implicitly fitting one parameter, and ii) model fitting of two or more parameters for prediction. Advantages of model fitting include a more systematic representation of statistical uncertainty and a more general investigation of the available models.

**Table 1: Summary model comparison**

<i>Model</i>	<i>Model description</i>	<i>Data inputs</i>	<i>Parameters estimated</i>
Andrews and Atkins (1993)	$cases(t)$ from an aetiological study	$cases(t)$ from the Wittenoom study (Berry, 1991)	Calibrated to the portfolio of interest
Peto et al (1995)	Age-cohort model (sub-model of Equation (3))	Mesothelioma mortality rates	Estimated the age and cohort effects
Stallard et al (2005)	Equation (2)	Mesothelioma incidence <i>or</i> mesothelioma claims	Indirect estimation of dose by age of incidence and age of initial exposure
KPMG (2006)	Exposure model with delay distribution (sub-model of Equation (3))	Australian asbestos consumption	Calibrated to the portfolio of interest
Re-implementation of KPMG (2006)	Exposure model with delay distribution (sub-model of Equation (3))	Australian asbestos consumption, mesothelioma incidence	Estimated parameters for the delay distribution
Clements et al (2007)	Equation (3)	Mesothelioma incidence rates	Estimated parameters for the dose effects and the intercept

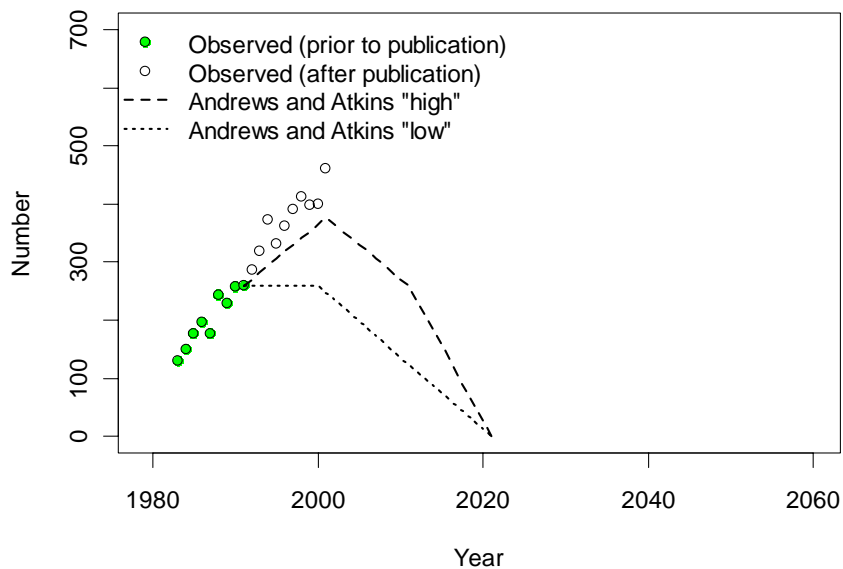
A summary comparison of the different models is given in Table 1 (see previous page). Further details of the models are given in the following sections. Note that the models proposed by Andrews and Atkins (1993) and Peto et al (1995) have been included for historical reasons; we do not propose the routine use of these models. The hypothetical model proposed by Finnis (1996) is also closely related to Equation (2).

### Andrews and Atkins (1993)

In the absence of good exposure data and sufficient a case series, Andrews and Atkins (1991, 1993) proposed using the case function  $cases(t)$  from a study of the Wittenoom mine in Western Australia (Berry 1991) and applied that pattern to the Australian population. This was a clever approach. Their model assumed that the dose pattern and population distribution over ages and over time in the Wittenoom study would be representative for Australia.

Historical experience has been unkind to the Andrews and Atkins projections: both the “low” projection, predicting a flat pattern during the 1990s, and the “high” projection, predicting a peak in 2001, were inconsistent with the observed pattern. A comparison between observed counts and predictions from Andrews and Atkins, calibrated to the observed frequency in 1991, is presented in Figure 3. The time axis has been taken out to 2060 for comparison with the other models. For Australian males, the Andrews and Atkins model predict 1485 cases for the “low” scenario and 2905 for the “high” scenario. For males in New South Wales, the Andrews and Atkins model predict 530 cases for the “low” scenario and 1040 for the “high” scenario.

**Figure 3: Mesothelioma incidence, observed numbers together with predictions from Andrews and Atkins (1993), Australian males**



These projections underestimated numbers and gave an early predicted peak because of the different time pattern of asbestos exposure at the Wittenoom mine and Australia as a whole. There was a flattening off of cases during the 1990s at Wittenoom with the peak probably reached by 2000, but given the continued use of asbestos in Australia beyond the mine's closure in 1966 for another 15 years or so, the peak in Australia would not be expected to occur until about 15 years later. Usefully, the Wittenoom study has been pivotal in the international literature in exploring the functional form for  $g(u)$ .

Alternative approaches could include using the rate function from Wittenoom and then applying the population distribution for Australia. This would still be limited by the assumption that the exposure experience between the two populations is similar.

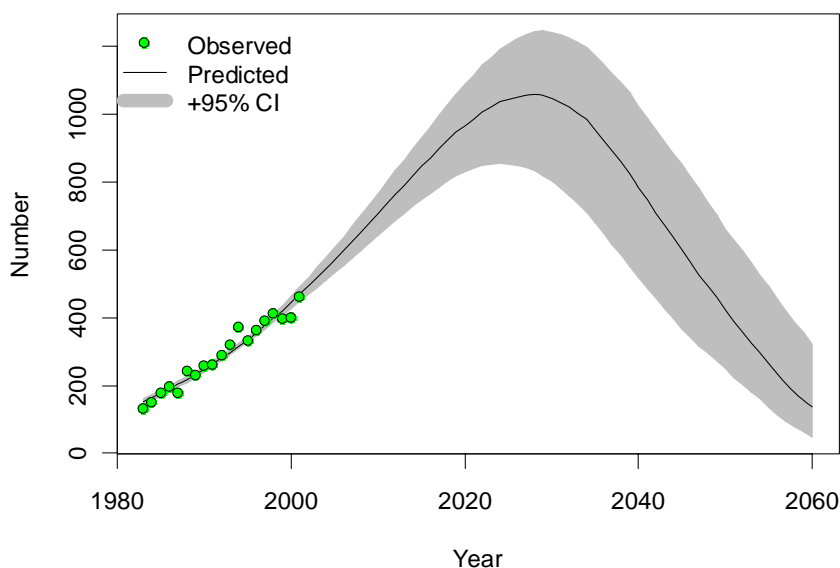
### Peto et al (1995)

Peto and colleagues used age-cohort models to predict mesothelioma mortality in European countries (Peto et al 1995, 1999). Their model can be considered as a simplification to Equation (3). If we assume a fixed age at initial exposure, at age  $a_0$ , with one year of exposure, then  $u = a - a_0$  and  $W$  is non-zero between  $a_0$  and  $a_0+1$  and zero elsewhere.

Equation (3) then simplifies to  $rate(a,t) = D(t - a + a_0)g(a - a_0)$ . As  $t - a$  represents the year of birth for a cohort, the revised model can be recognised as an age-cohort model. The age-cohort model assumes that each cohort will follow the same pattern across ages with a cohort-specific multiplier (a *proportional hazards* assumption). This is a consequence of the fixed age at initial exposure. However, there is good evidence for asbestos exposure across the working life-course, so that the fixed age at initial exposure is a simplifying assumption. Moreover, Hodgson et al (2005) found empirical evidence for cohorts following different patterns across ages, giving little support for the proportional hazards assumption. This assumption is likely to explain why earlier predictions from age-cohort models (Peto et al 1995, 1999) have tended to peak later than other models.

We have used a naïve approach to modelling an age-cohort model using generalised additive models, with interval estimation using the bootstrap. The methods are described in Clements et al (2005). Predictions used the assumption that there was negligible asbestos exposure for birth cohorts born from 1970 (see Figure 4). Under this poorly supported model, the annual number of cases for Australian males is predicted to peak in 2028 (95% confidence interval (CI): 2024, 2028), with 39,850 cases (95% CI: 30425, 49555) between 2006 and 2060. There is greater uncertainty in the model fit for males in New South Wales than for Australian males. The annual number of cases for NSW males is predicted to peak in 2029 (95% confidence interval (CI): 2025, 2034), with 14,855 cases (95% CI: 11205, 19495) between 2006 and 2060.

Figure 4: Observed mesothelioma incidence and age-cohort model predictions, Australian males



There is some flexibility in the development of the age-cohort models for predicting mesothelioma incidence. We have recently found that incorporating more prior knowledge



about the form of the age function can provide predictions that are more consistent with other methods, including a model based on work by the Health Safety Executive in the United Kingdom (Hodgson et al 2005): see the later section on Clements et al (2007).

### Stallard et al (2005)

In predicting mesothelioma claims for the Manville Asbestos Case in the United States, Stallard et al (2005) used an approach based on Equation (2), where they distributed observed cases by time since initial exposure, then indirectly estimated exposure. Stallard et al (2005) used two different models. Their first model, based on Walker (1982), used population-level mesothelioma incidence and then distributed cases by level of exposure (heavy, light) and by timing of exposure using additional data sources. Their second “hybrid” model used mesothelioma claims from the Manville Trust and then distributed cases by occupation and timing of exposure. For both models, Stallard and colleagues estimated the exposure by:

$$\overline{dose}(a-u, t-u) = \frac{cases(a, t, u)}{Pop(a, t)g(u)}$$

The approaches due to Stallard et al (2005) and Hodgson et al (2005) are similar, as seen from comparing Equations (2) and (3). Stallard and colleagues distributed cases by age at initial exposure and by level of exposure; in contrast, Hodgson et al (2005) assumed that the dose function was separable and integrated across age at initial exposure to fit Equation (3).

Advantages of the approach due to Stallard et al (2005) include the ready incorporation of more detailed data to further inform the model building and the availability of the authors’ monograph to guide other modelling efforts. As potential disadvantages: the first model used by Stallard et al is not immediately amenable to statistical analysis, as the observed cases are artificially distributed by age of initial exposure and by level of exposure; and the second model requires large numbers of cases that are free of reporting bias (e.g. non-differential propensity to claim by age and occupation) to reliably estimate exposure.

To our knowledge, these models have received only limited application in Australia.

### KPMG (2006)

The model used by KPMG to predict mesothelioma claims for former James Hardie entities takes a particularly simple and elegant form, being the convolution between the change in exposure over time and a delay distribution:

$$cases(t) = \int_0^{\infty} D(t-u)f(u)du \quad (4)$$

where  $D(t)$  is a dose function by time and  $f(u)$  is a probability density function for the delay from exposure to case diagnosis (or claim). The authors observed that the time from diagnosis to claim was typically less than a year.

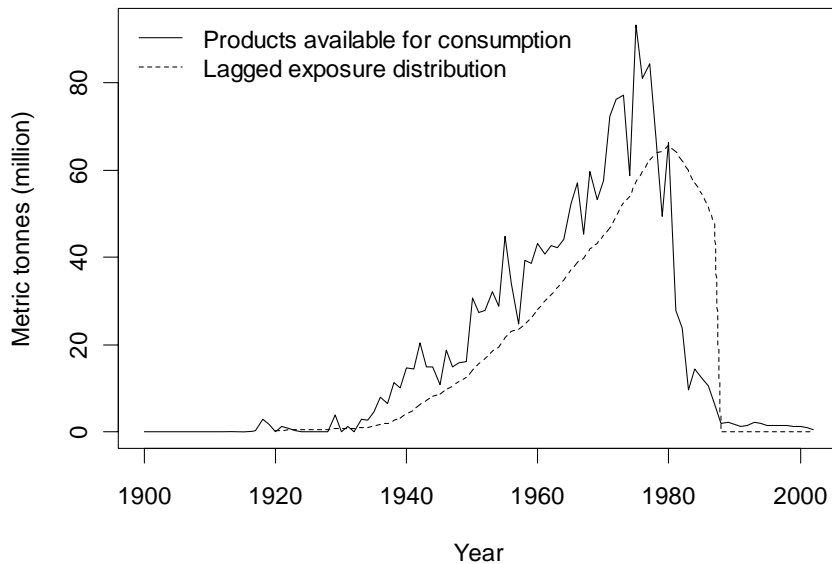
KPMG assumed that  $f(u)$  was a probability density function for a normal distribution with mean 35 and standard deviation 10. The authors used a transformation of asbestos products available for consumption (= production – exports + imports) to represent  $D(t)$ ; specifically, they used

$$D(t) = \frac{\beta}{16} \int_0^{16} Consumption(t-u) du$$

The authors also assumed that asbestos exposure was negligible from 1987 (see Figure 5). The use of cumulative consumption was motivated by the observation that the average claim to the former James Hardie entities had had exposure over 16 years. However, we argue that an Armitage-Doll model of carcinogenesis is additive for years of exposure, so that exposure for an individual over 16 years is equivalent to exposure for 16 individuals over one-year

periods, hence the length of individual exposure will not directly enter into the rate calculation. Usefully, we can re-interpret this lagged cumulative consumption as a delay distribution between products available for consumption and actual exposure, with the delay following a flat distribution between 0 and 16 years with a mean of 8 years.

**Figure 5: Asbestos products available for consumption and a hypothesised lagged distribution for exposure, Australia**



Model validity was assessed by the predicted peak (2010/2011) and moderate agreement between the predictions and observed mesothelioma incidence for Australia. Note that the exposure pattern is consistent with a moderate decline in exposure during the early 1980s and a rapid decline in the late 1980s.

By calibrating the total counts to the observed count for 2001, we obtain Figure 6. The model fit within the observed data is quite reasonable, with a suggestion that the rise in the prediction curve may be shallow. For Australian males, the model predicts 10,970 cases between 2006 and 2060. Applying the same model to NSW, we predict 3530 cases between 2006 and 2060.

We re-implemented the KPMG model by fitting the delay distribution, estimating the mean and standard deviation for the delay from the hypothesised asbestos exposure distribution through until mesothelioma diagnosis. This three parameter model was fitted using a Poisson likelihood in R to the observed total counts. The delay distribution had a mean of 39.0 (se=5.0) with a standard deviation of 10.4 (se=4.5). The imprecision in the estimates is not surprising, given that we are fitting the observed annual counts for 19 years using a three parameter model. The model predicts a peak in 2014, with 15,045 cases from 2006 through until 2060. Re-fitting the model to data from New South Wales, the estimate for the standard deviation was not significant. As an alternative approach, we fixed the standard deviation to the estimate for Australia (10.4 years) and re-fitted the model. The model predicted 4880 cases during 2006 to 2060, with a peak in 2013.

Figure 6: Calibration of KPMG predictions to mesothelioma incidence for Australian males

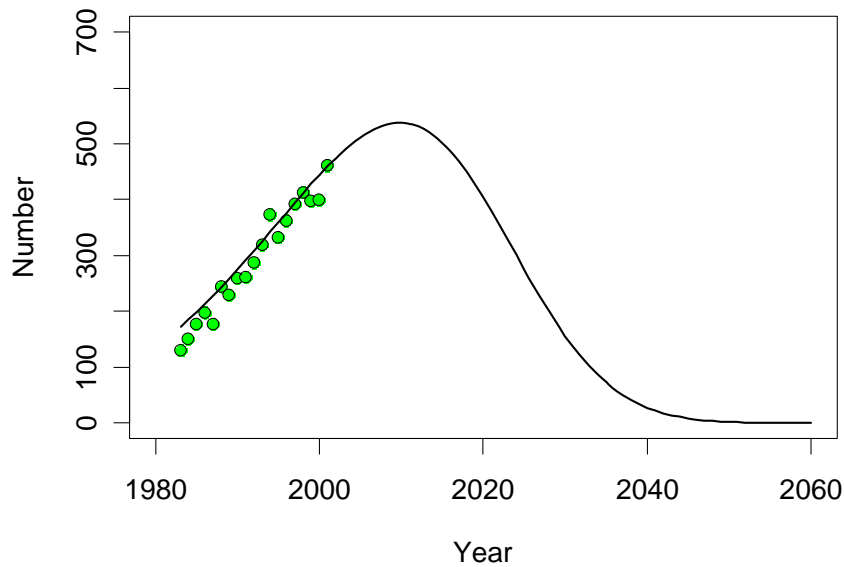
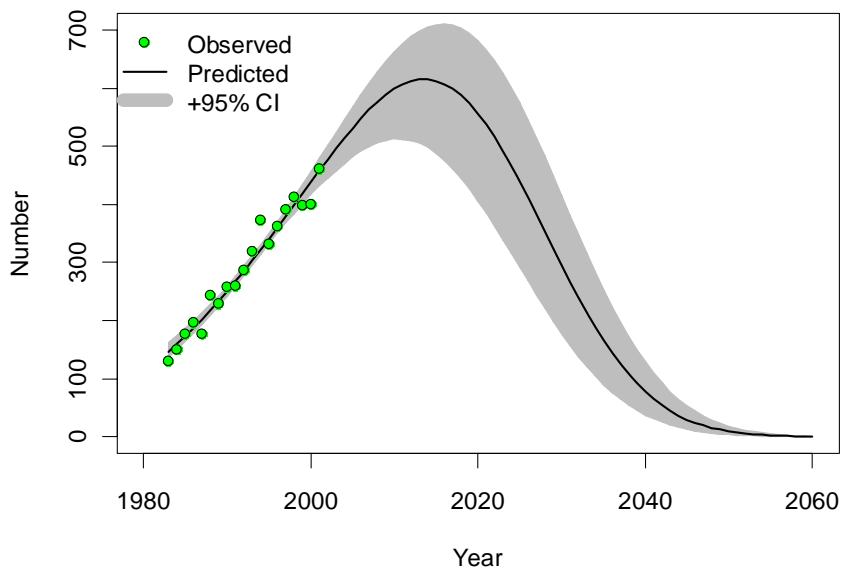


Figure 7: Predicted number of incident mesothelioma cases, re-implementation of KPMG (2006), Australian males



Note: For the confidence intervals, the standard deviation for the delay distribution was fixed at 10.4 years.

For interval estimation of the re-implementation fitted to mesothelioma cases for Australian males, we found that joint estimation of the three parameters lead to very wide and uninformative confidence intervals. If we make the moderately strong assumption that the standard deviation of the delay distribution is fixed at 10.4 years, then we find moderately close confidence intervals (Figure 7). Such an assumption would be predicated on external data supporting the standard deviation being around 10.4 years.

As a sensitivity analysis for the hypothesised delay between products available for consumption and exposure, we also re-fitted the model to the data on asbestos products available for consumption. This model accounts for the joint delay from products available, through to asbestos exposure, and then through to mesothelioma incidence. The model was not able to explicitly account for a rapid decline in exposure in 1987. The model predicted

very similar estimates to our re-implementation above both for the peak year in cases (2015 for Australia and 2013 for NSW) and for the total number of cases during 2006-2060 (15385 for Australia and 4685 for NSW). In summary, the predictions seem to be insensitive to the hypothesised delay between products available for consumption and exposure.

Using the theoretical framework, we can use Equations (3) and (4) to show that the function  $f$  is a function of both time since exposure and calendar period:

$$f(u, t) = \left[ \int_u^{\infty} W(a - u) Pop(a, t) da \right] g(u) \quad (5)$$

As the only expression in Equation (5) that depends on time is the population, we could make the moderate assumption that the age-specific population does not change over time. Equation (4) can then be viewed as having factored age out from Equation (3). Alternatively, we could include population in the model, which would require the more mild assumption that the age-specific density does not change over time. The new model would then take the form:

$$cases(t) = \left[ \int_0^{\infty} D(t - u) f(u) du \right] \times Pop(t)$$

This new model may provide a useful avenue for future modelling.

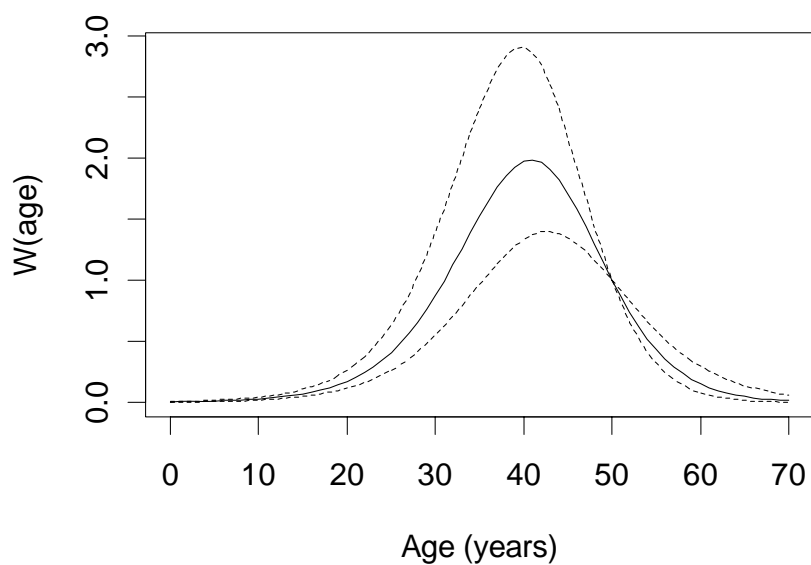
### **Clements et al (2007)**

Recently, we have re-implemented the model used by Hodgson et al (2005) for New South Wales. The rate function fitted by Hodgson et al (2005) for mesothelioma incidence predictions for the UK Health Safety Executive is very similar to Equation (3); Hodgson and colleagues also included a period term for under-diagnosis. Importantly, the model formulation was based on a strong understanding of the epidemiology. The implementation could be criticised for trying to fit too many parameters and using a univariate optimisation routine in Excel to “fit” a multivariable likelihood function. We are not able to generalise the peak from the United Kingdom to Australia because of significant differences in the pattern of asbestos use.

We have re-implemented this model, reducing the number of parameters to be estimated from 14 down to five (Clements et al 2007). We used natural splines to represent the function  $W()$  for change in dose by age, assuming that  $W()$  is 1 at age 50 years, and the function  $D()$  for change in dose by calendar period, assuming that  $D()$  is 1 in the 1970 calendar year. We fitted two parameters for each set of splines, together with an intercept term. We optimised the Poisson likelihood using a non-linear optimiser in R, with interval estimation using the bootstrap.

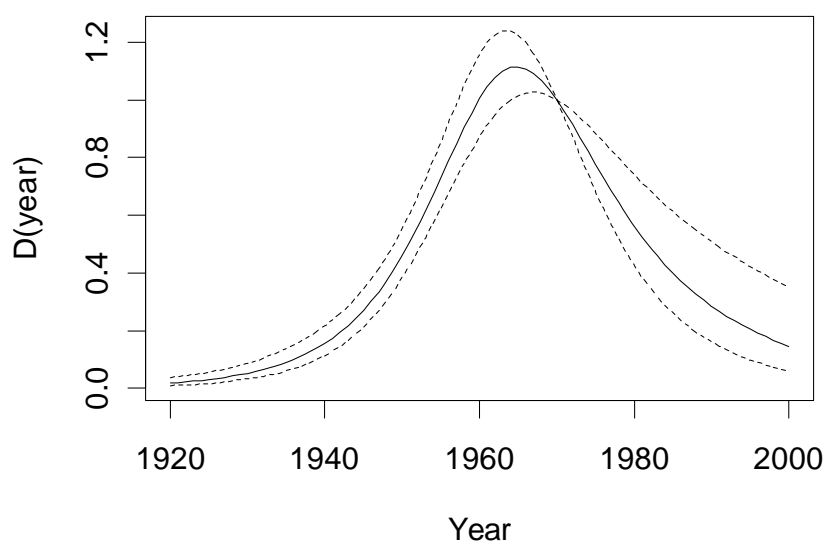
For New South Wales, the estimated pattern of exposure by age, represented by the function  $W()$ , suggests that there was a rapid rise in exposure from age 20 years, peaking around age 40 years, with a subsequent decline to negligible levels by age 60 years (see Figure 8). The estimated age pattern may be consistent with long periods of asbestos exposure, as suggested by KPMG (2006). For the related change in exposure by calendar period, represented by the function  $D()$ , there were negligible levels of exposure in 1930, with a rapid rise during 1940-1960, peaking in the mid 1960s, with a subsequent decline during the late 1970s and 1980s (see Figure 9). The decline in exposure is similar to the exposure pattern modelled by KPMG (2006). The fitted model predicts a peak in the number of cases in 2014, with 6430 (95% CI: 4920, 9060) cases during 2006-2060 (see Figure 10).

**Figure 8: Estimated dose function by age (with 95% confidence intervals), Clements et al (2007), NSW males**



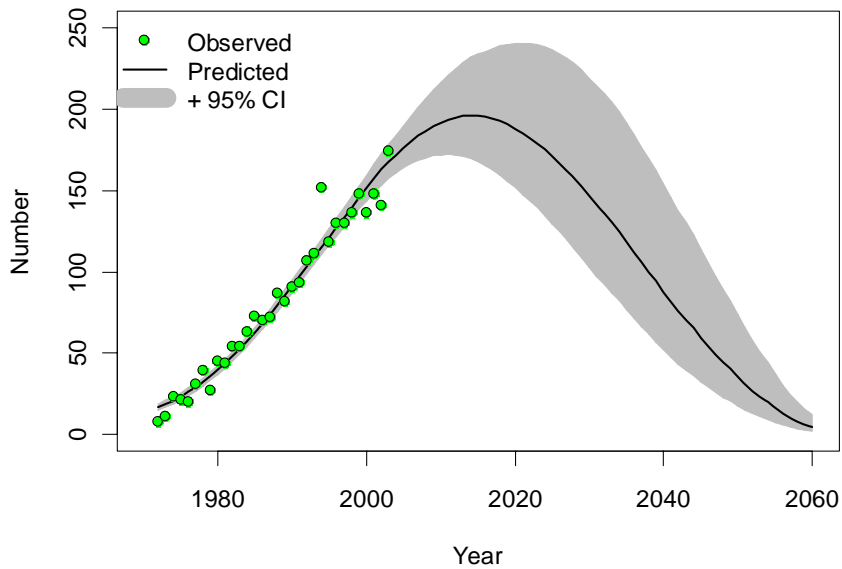
Source: Clements et al, 2007.

**Figure 9: Estimated dose function by calendar period (with 95% confidence intervals), Clements et al (2007), NSW males**



Source: Clements et al, 2007.

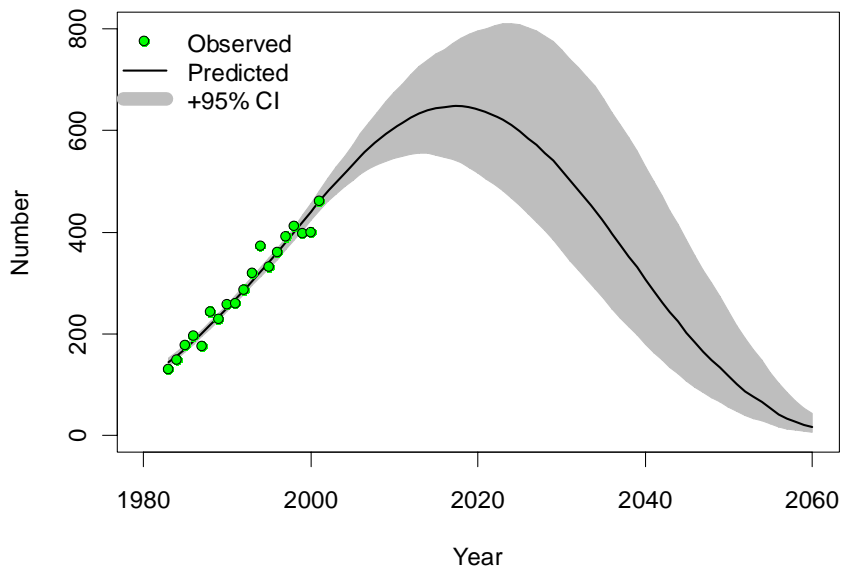
Figure 10: Predicted number of incident mesothelioma cases, Clements et al (2007), NSW males



Source: Clements et al, 2007.

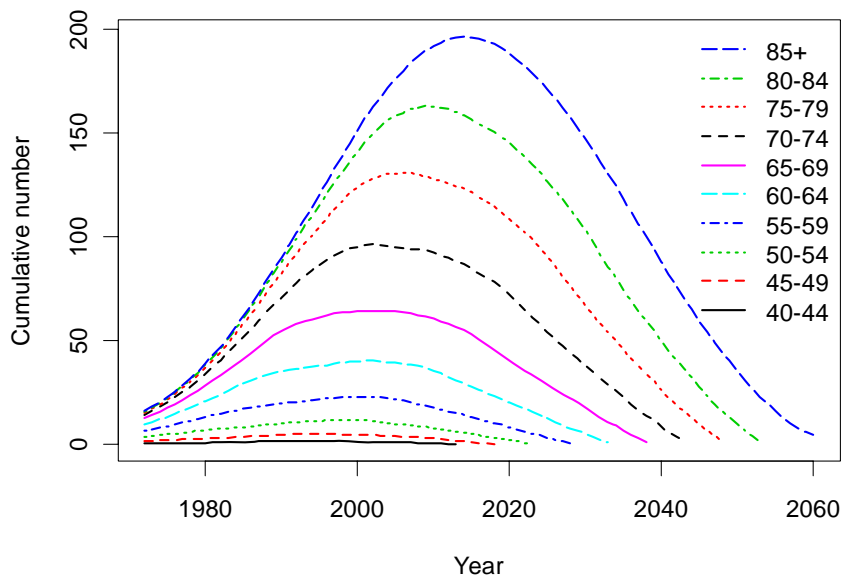
We have also applied the model to mesothelioma incidence for Australian males (Figure 11). The only model assumption that was changed between New South Wales (Clements et al 2007) and Australia was that the highest knot value for the dose function  $D()$  by calendar period was 1990 rather than 1980. The model predicts a peak in 2017 (95% CI: 2013, 2024), with the total number of cases for 2006-2060 being 21700 cases (95% CI: 16460, 30165).

Figure 11: Predicted number of incident mesothelioma cases, Clements et al (2005) model, Australian males

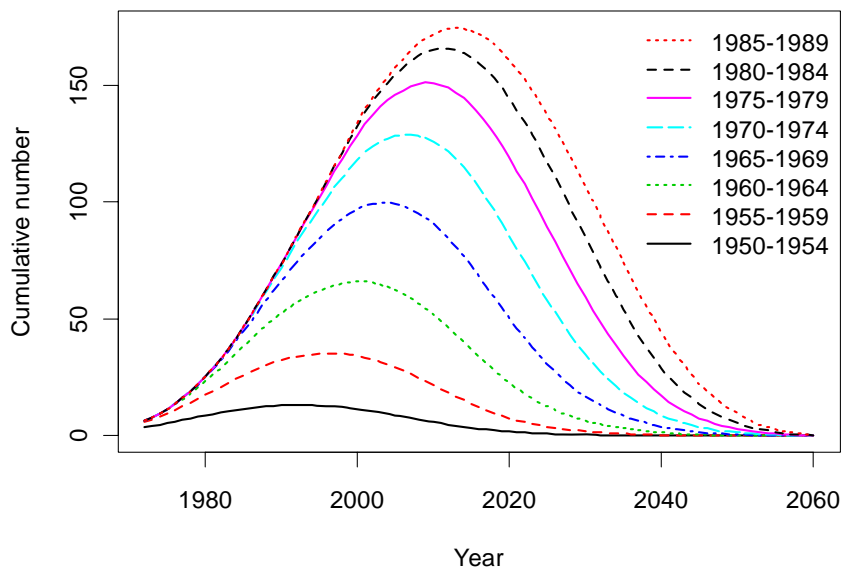


One advantage of modelling the age-specific data is that we can represent the number of predicted cases by birth cohort and by age group (Figure 12). Moreover, by modelling for age at exposure, we can predict the number of cases by age at exposure and by period of exposure (Figure 13). These plots can be used to assess the relationship between the future predictions and the modelled exposure function. The predictions may also be applied to particular subsets of the population, such as for those exposed during a particular period.

**Figure 12: Predicted number of cases, by age at cancer registration, Clements et al (2007) model, NSW males**



**Figure 13: Predicted number of cases, by period of exposure, Clements et al (2007) model, NSW males**



### Summary of results

Predictions from the different models are summarised in Table 2. As previously described, the model due to Andrews and Atkins (1993) and the age-cohort model implementation (Peto et al 1995; Clements et al 2005) are included for historical reasons; these models, as implemented, provide uninformative bounds on the number of cases and are not recommended for general use. As a counter-point, other implementations of the age-cohort model may be useful.

**Table 2: Summary of mesothelioma predictions, by model and population**

Model	Australian males		NSW males	
	Peak	Total count 2006-2060	Peak	Total count 2006-2060
Andrews and Atkins “high” Age-cohort model	2001	2905	2001	1040
re-implementation	2028	39850	2029	14855
KPMG (2006)	2010	10970	2010	3530
KPMG (2006) re-implementation	2014	15045	2013	4880
Clements et al (2007)	2017	21700 (95% CI: 16460, 30165)	2014	6430 (95% CI: 4920, 9060)

For the latter three models, there is some variation in the total predicted number of mesothelioma cases for Australia. Model predictions for the peak year from the re-implementation of the KPMG model are similar between Australia and New South Wales. In contrast, the peak years differ considerably between jurisdictions for Clements et al (2007) model. Given the limited time series for Australia and the reasonable agreement between models for the New South Wales data, the Australian estimates based on the Clements et al (2007) model may be high. However, both the re-implementation of the KPMG model and the Clements et al (2007) model predict later peaks and considerably higher total counts than the base KPMG model.

## Discussion

In summary, we have used a theoretical framework based on the Armitage-Doll model of carcinogenesis to compare models for predicting mesothelioma incidence. We also empirically compared those models. Observed data provided broad empirical support for the KPMG model, the re-implementation the KPMG model and the Clements et al (2007) model. We found that the KPMG model (2006) peaked earlier than the other two models and that the KPMG model point estimate for total counts for 2006-2060 was well outside the 95% confidence intervals for predictions from the Clements et al (2007) model.

Importantly, we consider that there is reasonable empirical evidence that the peak for mesothelioma incidence is later than 2010. This has far-reaching consequences for actuarial predictions, where the number of cases out to 2060 may be in excess of 35% higher than the number predicted by KPMG (2006).

We propose some general recommendations. First, we recommend that models be fitted to observed data, rather than undertaking simple calibrations. The mechanisms generating mesothelioma cases in a population are diverse and poorly observed, hence we recommend an approach that allows for flexible modelling of such mechanisms. Second, we recommend fully representing the statistical uncertainty in model-based predictions. The actuary must take account of other sources of uncertainty, however there is such a strong degree of uncertainty in the incidence predictions that representing predictions using a mean curve is ignoring important information. Third, we recommend that actuaries give closer attention to the epidemiological literature when predicting health-related outcomes. The two sets of literature provide complementary views on those topics that intersect, including product liability and health insurance. Moreover, there is a paucity of researchers who write in both sets of literature. There are opportunities for further cross-fertilisation between the two disciplines.

How can we assess model misspecification for any of these models? This is at the heart of the matter, as we often only observe and model aggregate-level data. The *epidemiology* suggests,



given data on mesothelioma incidence, that the models due to Stallard et al (2005), KPMG (2006) and Clements et al (2007) are well-specified. These models are closely related to each other and should, in principle, provide very similar predictions. We suggest modelling the age-specific data, which provides for richer models and more precise predictions. The hybrid model used by Stallard et al (2005) employs more information – however it depends critically on having a large number of cases, as satisfied in the Manville Asbestos Case, with no reporting bias. This approach deserves closer attention in Australia. It is more difficult to assess model misspecification for descriptive models which are predicated on strong assumptions, as has been borne out in recent history (e.g. Andrews and Atkins 1993, Peto et al 1995).

For model assessment, we can approximate the actuarial “control cycle”, either by assessing the fit on annually updated aggregate-level data or by validation using out-of-data model assessment, where we fit a model without the last five years of data and assessing the model predictions on the last five years (Hastie et al 2001, Clements et al 2005).

Stallard et al (2005) provided a taxonomy based on direct and indirect estimation of asbestos exposure. We have focussed on recent methods for population-level modelling and on indirect estimation of asbestos exposure. Some of the criticism of population-level modelling is due to the fallibility of the age-cohort model. We support population-level modelling as a general approach, whilst being aware that some such models have poor properties. Methods using direct estimates, usually based on occupational exposures, have received support in the past (see the review by Stallard et al 2005) and more recently, particularly by the IAAust working group (see Huszczo et al 2004). It remains unclear as to whether reasonable metrics of occupationally based exposure will be available. At present, we suggest that methods using indirect estimation have better support than direct modelling of asbestos exposure.

### **Acknowledgements**

We would like to thank Bruce Watson from Finity Consulting for encouragement and peer review of this paper; we would also like to thank Neil Donlevy from KPMG Actuaries for help with model details.

### **References**

- Andrews T, Atkins G. *Asbestos-Related Diseases – The Insurance Cost*. Institute of Actuaries of Australia Accident Compensation Seminar. 1991.
- Andrews T, Atkins G. *Asbestos-Related Diseases – The Insurance Cost – Part 2*. Institute of Actuaries of Australia Biennial Convention. 1993.
- Atkins G, Smith D, Watson B. *Recent Trends in Asbestos-Related Diseases*. Institute of Actuaries of Australia Accident Compensation Seminar. 1996.
- Berry G. Prediction of mesothelioma, lung cancer, and asbestosis in former Wittenoom asbestos workers. *British Journal of Industrial Medicine* 1991;48:793-802.
- Berry G. Models for mesothelioma incidence following exposure to fibres in terms of timing and duration of exposure and the biopersistence of the fibres. *Inhalation Toxicology* 1999;11:111-130.
- Brillinger DR. The natural variability of vital rates and associated statistics (with discussion). *Biometrics* 1986; 42: 693-734.
- Clements MS, Armstrong BK, Moolgavkar SH. Lung cancer rate predictions using generalized additive models. *Biostatistics* 2005; 6: 576-89.

Clements MS, Berry G, Shi J, Ware S, Yates D, Johnson A. Projected mesothelioma incidence in men in New South Wales. *Occupational and Environmental Medicine* 2007 (in press).

Finnis D. *An Employer-Based Model for Estimating Asbestos-Related Claim Liabilities*. Institute of Actuaries of Australia Accident Compensation Seminar. 1996.

Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data mining, inference, and prediction*. Springer, New York. 2001.

Hodgson JT, McElvenny DM, Darnton AJ, Price MJ, Peto J. The expected burden of mesothelioma mortality in Great Britain from 2002 to 2050. *British Journal of Cancer* 2005;92:587-93.

Huszczko A, Martin P, Parameswaran S, Price C, Smith A, Walker D, Watson B, Whitehead G. *IAAust Asbestos Working Group Discussion Paper*. Institute of Actuaries of Australia Accident Compensation Seminar. 2004.

KPMG. *Valuation of asbestos-related disease liabilities of former James Hardie entities ("The Liable Entities") to be met by the Special Purpose Fund*. Report to James Hardie Industries NV. Sydney: KPMG Actuaries Pty Limited. 2006.

Leigh J, Rogers AJ, Ferguson DA, Mulder HB, Ackad M, Thompson R. Lung asbestos fiber content and mesothelioma cell type, site, and survival. *Cancer* 1991; 68: 135-141.

Peto J, Decarli A, La Vecchia C, Levi F, Negri E. The European mesothelioma epidemic. *British Journal of Cancer* 1999; 79: 666-72.

Peto J, Hodgson JT, Matthews FE, Jones JR. Continuing increase in mesothelioma mortality in Britain. *Lancet* 1995;345:535-9.

Peto J, Seidman H, Selikoff IJ. Mesothelioma mortality in asbestos workers: implications for models of carcinogenesis and risk assessment. *British Journal of Cancer* 1982;45:124-35.

Stallard E, Manton KG, Cohen JE. *Forecasting Product Liability Claims: Epidemiology and modeling in the Manville asbestos case*. Springer Science+Business Media, Inc., New York. 2005.

Walker AM. *Projections of Asbestos-Related Disease 1980-2009*. Epidemiology Resources, Inc. 1982.