A simple benchmark for mesothelioma projection for Great Britain

MD Martínez-Miranda, B Nielsen, JP Nielsen

ABSTRACT

Background: It is of considerable interest to forecast the future burden of mesothelioma mortality. Data on deaths are available, whereas no measure of asbestos exposure is available.
Methods: We compare two Poisson models: a response-only model with an age-cohort specification and a multinomial model with epidemiologically motivated frequencies.
Results: The response-only model has 5% higher peak mortality than the dose-response model. The former performs slightly better in out-of-sample comparison.
Conclusion: Mortality is predicted to peak at about 2100 deaths around 2017 among males in cohorts until 1966 and below 90 years of age. The response-only model is a simple benchmark that forecasts just as well as more complicated models.

WHAT THIS PAPER ADDS

State of scientific knowledge: The future mesothelioma burden is traditionally projected using multinomial models with epidemiologically motivated frequencies.
Main messages: An age-cohort response-only model appears to forecast just as well, but slightly higher, than traditional methods. The model gives a simple benchmark for the evaluation of future epidemiological models.
Policy implication: Mesothelioma mortality is predicted to peak at about 2100 around 2017 for UK males.

INTRODUCTION

The number of mesothelioma deaths in Great Britain continues to increase with 2049 male deaths below 90 years of age in 2013. It is of considerable interest to forecast the future burden of mesothelioma deaths. Mesothelioma is a cancer that is mainly caused by exposure to asbestos fibres. It has a long latency period and once discovered it is rapidly fatal. The regulation of the use of asbestos was tightened substantially in 1969. This led to a substantial reduction in exposure by about 1980. Most UK mesothelioma cases are caused by occupational exposure[1]. This exposure tended to affect men. In any case, while the UK has good records of mesothelioma deaths there are currently no detailed data on the number of people who have been exposed to asbestos fibres.

When exposure is not recorded a first approach is to construct synthetic exposure measures. In [2] a dose-response model with an age-cohort parametrization is used where the exposure, or dose, is measured in terms of the total UK population. The Hodgson et al. model [3] is a multinomial model, where the probabilities are based on epidemiological knowledge with respect to time since exposure, age at exposure and lung clearance. This has been implemented in a Bayesian set-up to achieve a forecast distribution.[4] In a later multinomial model the probabilities are based on a two-stage clonal expansion model, which is a carcinogenesis model that assumes that the development of a malignant cell is the result of two critical and irreversible events.[5] Detailed forecasts using these methods are provided by the UK Health and Safety Executive (http://www.hse.gov.uk/statistics/tables/index.htm/#lung, Sep 2015) and the UK Asbestos Working Party (http://www.actuaries.org.uk/research-and-resources/documents/f10-uk-awp-mesothelioma-tsce-model, Sep 2015). A recent response-only method avoids the need for synthetic exposure.[6] This is simpler and less dependent on uncertainty in the epidemiological insight and therefore it may give more robust forecasts.

METHODS

The Health and Safety Executive reports the number of mesothelioma deaths in Great Britain by age and period for 1968-2013, see Supplement. Most deaths are male (85%) and in the age groups 25-89 (99%). The number of deaths is increasing over time with 2049 deaths in 2013, see Figure 1. Most deaths occur at ages 45-85 corresponding to cohorts 1920-1950. In recent years the number of deaths at age 90+ has increased with 74 such deaths in 2013. There are only 45 cases for the 1967-1988 cohorts, which have benefitted most from predictive measures since the 1970s. To be consistent with the literature we use ages 25-89 for men.

The response-only model is an age-cohort Poisson regression model for the number of deaths.[6] Inference is carried out by conditioning on the overall number of observed deaths. This results in a multinomial sampling scheme, where standard inference methods apply. The model can be estimated by a generalized linear model routine. We used the R package apc[7,8], see Supplement.

The age-cohort model has predictor $E(Y_{age,period})=exp(\alpha_{age}+\gamma_{cohort})$, where $Y_{age,period}$ is the number of deaths at a given age and a given period, recalling that $age+cohort=period$. The model is over-parametrised so we use the canonical parametrisation,[9,10] which ensures that standard generalized linear model techniques apply. We check the specification by considering a more general model that includes a period factor.

We only forecast cohorts in the sample since the youngest cohorts appear to have benefitted from protective measures in the workplace. Point forecasts can be constructed directly from the estimated canonical parameters. Distribution forecasts are calculated numerically using asymptotic methods.[6,8] We apply intercept correction so as to line the forecasts up with the most recent observation. This robustifies the forecasts against possible shifts in data at the forecast origin. The robustification is achieved by multiplying the point forecasts with a factor given by the number of observations in the last period, divided by the prediction for the last period.

We compare the age-cohort model with existing methods in the literature: the Hodgson et al model[4] and the two-stage clonal expansion model.[5] These models have considerably fewer parameters than the age-cohort response model, but they enter non-linearly and Bayesian methods are used to get distribution forecasts.

RESULTS

We first compare the fits and then the forecasts of the age-cohort response-only model and the epidemiologically motivated models.

The response-only model was fitted to the 1968-2013 data. The deviance of the age-period-cohort model is 2763.6 (df=2772, p=0.542) while the age-cohort model has 2818.1 (df=2816, p=0.485), so both variants appear to give a reasonable fit. The relative deviance is 54.6 (df=44, p=0.132). Based on this evidence we maintain the age-cohort model as parsimony is usually advantageous in forecasting. Likewise we do not find evidence against the model when considering plots of the residuals. Those plots are not reported here.

The quality of the fit of the two-stage clonal expansion model[5] can be judged from the in-sample predictions reported by the UK Asbestos Working Party using the data for 1968-2008.

The deviance is 2651.9 (df=2646, p=0.704). Measured in terms of fit, this model appears equally good, but that in itself is no guide to forecast performance.

Figure 1 shows sums of cases by period along with four forecasts. The sums of cases and the response-only forecasts exclude cohorts from 1967 and later. The observed cases are indicated with dots. We notice a large volatility in the most recent years, with a low 2011 count and a high 2012 count.

The forecast from the age-cohort response-only model is shown as a solid line. This uses all data until 2013 but only forecasts cohorts until 1966. The shaded region indicates pointwise 95% forecast error bands. These are computed from asymptotic methods and include both process error and estimation error. The peak forecast is 2079 in 2017 with 95% error band of (1981, 2177). The 2013 in-sample residual is very small, so intercept correction would increase the peak forecast modestly to 2083.

The dotted line is a forecast from the Hodgson et al. model[4] updated to 2010 data by the UK Health and Safety Executive. This forecast includes cohorts until 1986. Nonetheless, it comes out lower than the age-cohort forecasts. We do not show the forecast error bands as the width of the 90% error bands is nearly identical to width of the 95% bands of the age-cohort model reported in Figure 1. In particular, for 2013 the 90% error bands are (1901, 2069), as compared to the observed 2049 deaths. The peak forecast is 2003 in 2015. For comparison the two-stage clonal expansion model using 2007 data has peak 1854 in 2010.[5]

The 2013 age-cohort model and the Hodgson et al. model with data up to 2010 use different data, so they are not directly comparable. To illustrate their differences it is useful to consider forecasts from 2006: the dash-dot line is an age-cohort forecast with intercept correction,[6] while the dashed line is the Hodgson et al. model.[4] These forecasts generally have the same shape as their respective updates. The age-cohort forecast is most different, since the 2013 age-cohort forecast is affected by the recent volatility in the data. The in-sample quality of the two 2006 models are comparable, but the age-cohort response-only model forecasts slightly better.

CONCLUSIONS

Dose-response models are difficult to use in mesothelioma projections due to the lack of data on exposure. A simple response model with an age-cohort structure gives an in-sample fit and forecasts that are similar to those from existing models. It requires less epidemiological insight as it can be analysed using standard generalized linear model techniques. Since analytic expressions are available for forecast distributions it removes the need for simulation methods.

A limitation is that the age-profile is assumed to be common for all cohorts. This was a concern in the development of the Hodgson et al. model[3]. We agree with this concern although it may be less of a problem in a response-only model compared with a dose-response model using general population numbers as exposure. It could in principle be overcome by a more complex model. However, since we only have very little data the model uncertainty of such a model would be large. The improvement in forecast performance will therefore be modest as compared with the response-only age-cohort model.

In the empirical analysis we find a peak in 2017 of 2079 male deaths with 95% error band of (1998, 2161) for those of age less than 90 and cohort up to 1966. This is slightly worse and slightly later than the forecast of previous models.

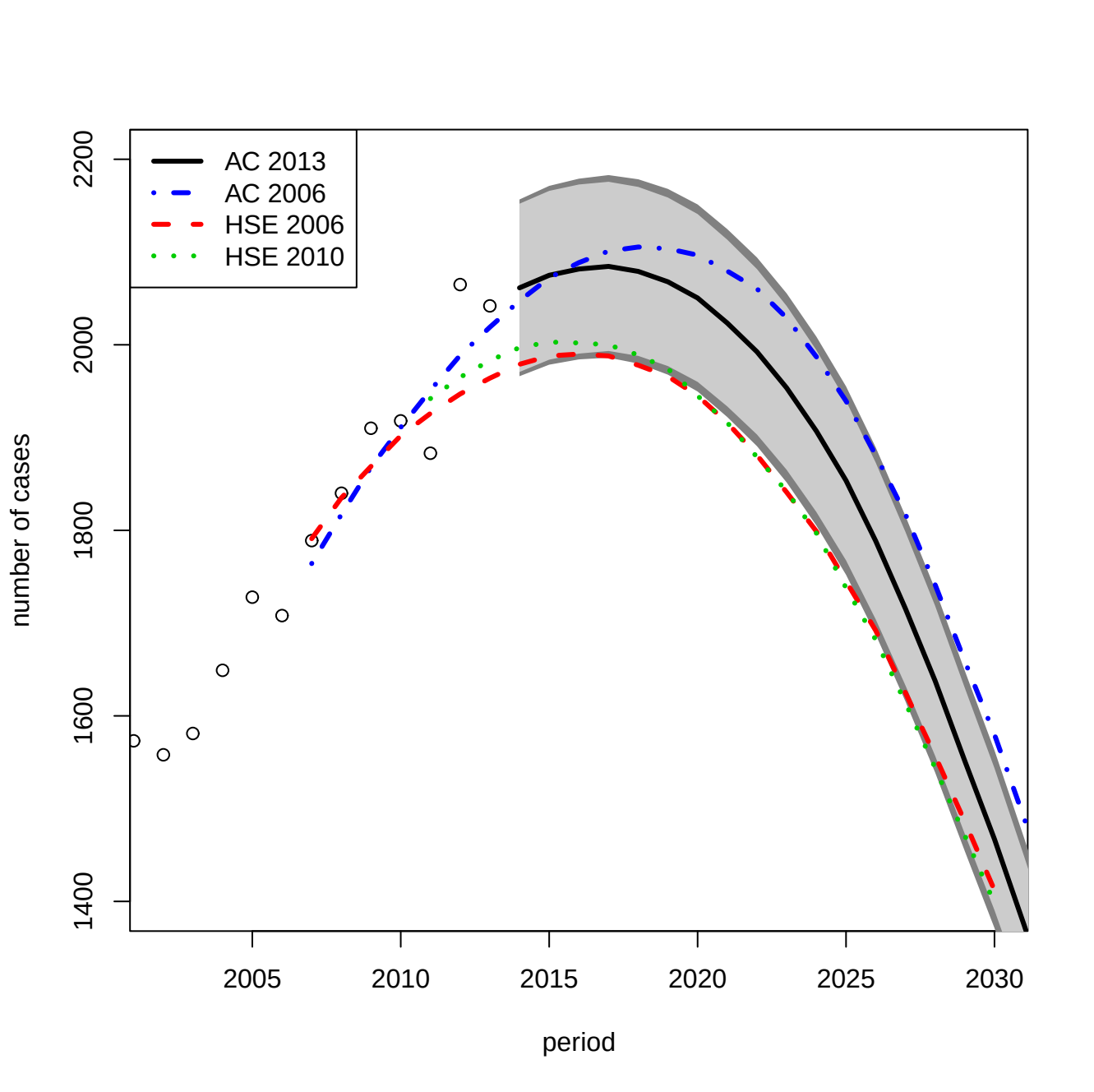
ACKNOWLEDGEMENTS
A.J. Darnton at the Health and Safety Executive provided the most recent version of the mesothelioma register. B. Nielsen is supported by the Programme for Economic Modelling at the University of Oxford.

FUNDING
M.D. Martínez-Miranda is supported by project MTM2013-41383P from the Ministry of Economy and Competiveness and the European Regional Development Fund.

REFERENCES
1 Rake, C., Gilham, C., Hatch, J. Et al. Occupational, domestic and environmental mesothelioma risks in the British population: a case—control study. British Journal of Cancer 2009; 100: 1175-1183.
2 Peto, J., Hodgson, J.T., Matthews, F.E. et al. Continuing increase in mesothelioma mortality in Britain. Lancet 1995; 345: 535-539.
3 Hodgson, J.T., McElvenny, D.M., Darnton, A.J. et al. The expected burden of mesothelioma mortality in Great Britain from 2002 to 2050. British Journal of Cancer 2005; 92: 587-593.
4 Tan, E., Warren, N., Darnton, A.J. et al. Projection of mesothelioma mortality in Britain using Bayesian methods. British Journal of Cancer 2010; 103: 430-436.
5 Tan, E., Warren, N., Darnton, A.J. et al. Modelling mesothelioma mortality in Great Britain using the two-stage clonal expansion model [abstract]. Occupational & Environmental Medicine 2011; 68: A60.
6 Martínez-Miranda, M.D., Nielsen, B. and Nielsen, J.P. Inference and forecasting in the age-period-cohort model with unknown exposure with an application to mesothelioma mortality. Journal of the Royal Statistical Society series A 2015; 178: 29-55.
7 R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2014.
8 Nielsen, B. (2014) apc: A package for age-period-cohort analysis. To appear in R Journal.
9 Kuang, D., Nielsen, B. and Nielsen, J.P. Identification of the age-period-cohort model and the extended chain ladder model. Biometrika 2008; 95: 979-986.
10 Kuang, D., Nielsen, B. and Nielsen, J.P. Forecasting with the age-period-cohort model and the extended chain-ladder model. Biometrika 2008; 95: 987-991.

FIGURE LEGEND
Forecasts of annual number of deaths.
Age-cohort models: —, 1968-2013; - · -, 1968-2006.
Hodgson et al. models from HSE: - -, 2006; ···, 2010.
Shaded area is the forecast distribution bands for the 2013 age-cohort model.

SUPPLEMENTARY MATERIAL
Data: UK mesothelioma deaths, male by age 25-89 and period 1968-2013. From HSE.
R code for reproducing results using R package apc version 1.2.


AC 2013
AC 2006
HSE 2006
HSE 2010
number of cases
2200
2000
1800
1600
1400
2005
2010
2015
2020
2025
2030
period