chapter 8) and the book on modelling by Burnham and Anderson (2002) that is unusual in not assuming the existence of a true model.

As well as model uncertainty, this paper looks at another important outstanding problem, namely how to handle *incomplete data*. The key equation in the paper is equation (16) and the authors show how this can facilitate handling incomplete data when model uncertainty is present. Thus, from my reading of the paper, I wonder whether a clearer title for the paper would be 'Bias arising from incomplete data in the presence of local model uncertainty'.

I am less sure whether, and if so how, equation (16) might be used to tackle the typical problem that is faced by a modeller with *complete* data but incomplete information about the model, as for example the time series analyst trying to decide whether an appropriate model is an AR(l), or AR(2), or MA(1) or.... If equation (16) can help, could we have an example?

Section 8 (re)raises the fundamental question about how to combine model uncertainty with sampling variability. The authors go on to say that the 'Bayesian paradigm provides a complete solution, at least in principle'. For this to be true, the modeller would need to know the relevant equation (16), assume that it is true and have priors for all necessary quantities. This is unlikely in practice, to say the least, and I doubt that this sort of statement is helpful. The authors also comment that a general formulation for handling model uncertainty when the model is chosen to depend on the data (as it often is) is 'difficult'. I am tempted to change this to 'impossible'.

Finding some of the mathematics rather difficult, I looked with particular interest at the main example in Section 7. Although I do not understand all the details, it seems to me that the example is mainly concerned with *model sensitivity* rather than model uncertainty. It is, of course, valuable to investigate how sensitive any conclusions are to the model assumptions, but this does not solve the general model uncertainty problem. The example assumes knowledge of a log-linear model and of likely departures from it, which is more information than the modeller often has.

Robert Curnow (*University of Reading*)

I congratulate the authors on an interesting paper on an important topic.

The example that they use to illustrate their approach, a meta-analysis of the risk of lung cancer from passive smoking, is taken from Hackshaw *et al.* (1997) and, as the authors state, was used by the Scientific Committee on Tobacco and Health in their report to the Chief Medical Officer at the Department of Health (in 1998). I was a member of this Scientific Committee at the time of the report. The authors do not make clear that the advice in the report took account of much wider evidence on the likely risks from environmental tobacco smoke. This included a parallel study of the literature on passive smoking and ischaemic heart disease (Law *et al.*, 1997) and evidence concerning sudden infant death syndrome and serious respiratory illness, asthmatic attacks and middle ear disease in children.

We did consider the possible biases in the epidemiological studies that were mentioned by the authors. We may have underestimated the evidence of a publication bias. Certainly possible important biases should be identified and dealt with as well as possible. However, the major problem is always the translation of the results of analyses into relevant information about the likely consequences of alternative policies. We estimated that the relative risk of lung cancer from passive smoking of 1.24 translated into several hundred extra lung cancer deaths a year. We concluded that the similar estimated relative risk for ischaemic heart disease translates into much larger numbers and 'represents a substantial public health hazard'. We hoped that these estimates together with measures of their uncertainties would be used by policy advisors and decision makers in comparing the costs and benefits of alternative policies. Confidence intervals as in Fig. 6 of the paper are interesting but the cost–benefit comparisons will surely influence our attitude to uncertainties in the estimates and in the relation of these uncertainties to the unknown level of correlation of exposure and the confounder. In this context, the importance of an interval at an arbitrary confidence level not including 1 is far from clear. Any rule that relative risks of less than 2 (quoted in Section 7) should be 'regarded with considerable caution' takes no account of the costs and benefits of alternative policies.

David Draper (*University of California, Santa Cruz*)

I can only add two small notes to this interesting and important paper.

(a) In Section 8 the authors
  (i) mention data-driven model search as normal practice,
  (ii) note that after such a search 'it is misleading to use (unconditional) sampling distributions as if the model were fixed' and