Bayes Methods for Combining the Results of Cancer Studies in Humans and Other Species: Comment

Author(s): Robert E. Kass

Source: *Journal of the American Statistical Association*, Vol. 78, No. 382 (Jun., 1983), pp. 312-313
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2288634
Accessed: 28-02-2018 18:08 UTC

Comment

ROBERT E. KASS*

Laboratory experiments often provide strong evidence about carcinogenesis or mutagenesis at high doses of a toxic agent in certain species. With specific assumptions about the relevance of this evidence, could some of its strength be borrowed, so to Tukey-speak, in order to better estimate the effect of exposure to the agent on the probability of cancer in humans? DuMouchel and Harris suggest that it can and they propose Bayesian analysis of covariance components models as an appropriate technology for accomplishing the task.

The great benefit of the Bayesian approach to this problem is that it makes precise the assessments of relevance and uncertainty in related results. Just as the purpose of a Bayesian theory of scientific inference is, as Jeffreys (1967, p. 8) put it, to "tidy up" the process, so too might Bayesian methods improve what advisors to policy makers do informally. In principle, the paradigm is simple and, in its ability to accommodate a variety of beliefs, it is comprehensive. Yet the Bayesian approach has its difficulties, for while it is surely desirable to express beliefs explicitly, in particular through models, it is often difficult to do so accurately. Lurking beside each analysis are the interrelated dangers of oversimplification, overstated precision, and neglect of beliefs other than the analyst's.

In the authors' model (3.2), the normal exchangeable prior on δ is probably the most important cause for concern. Interspecies comparisons begin with the assumption that results among different species and toxic agents are somehow related, and that the data carry information about the relationships. The normal exchangeable prior, however, represents a firm belief in a very simple situation; its use entails a strong statement about the behavior of the interactions. It is plausible that some groups of species are sufficiently similar with respect to the toxic action of some groups of agents, that, within those groups, model (3.2) would not be grossly inaccurate. In principle, the model could be modified to allow for grouping and, in Section 5, the authors provide an illustration of this sort of modification. However, it is not clear how Table 1 should be refined so as to improve the accuracy of the model. Thus, there is a serious complication associated with missing information as to group membership, and I would hesitate to apply the model without additional theoretical or empirical knowledge.

Overstated precision is a potential problem in the authors' model. Not only does each estimate "shrink" toward others in its row and column, but its variance is reduced, as well. If there is doubt about the accuracy of the model, then the resulting variance is inappropriately small. Consider, for example, the value $c^* = 1.02$ for roofing tar in Table 2. My own reaction to that number is that it is too small and, to interpret it, I crudely but quickly modify model (3.2) by mixing it with the model that assumes no relationship among the entries in Table 1. In addition, since the original data come from an observational study, the standard deviation of 1.41 might be considered too small, and, before mixing, it could be replaced with a value considered more realistic. Thus, my modified posterior, based on viewing Table 2, would be roughly equal to a mixture of the reported posterior and the variance-inflated original (normally distributed) data.

Even greater overstated precision results from the model selection described in Section 6. Evaluation of precision is worrisome whenever the same data are used both to select the model and to make inferences. Here, the problem is compounded. Not only is each estimate affected by the selection of entries in its row and column, but both the estimate and its variance are affected by increased homogeneity of the interactions. In particular, I would not agree with the conclusion that knowledge of the human roofing tar log slope may be summarized by a Normal$(1.53, (.74)^2)$ distribution.

The concern expressed here about difficulties in applying the authors' methodology raises a general question of policy analysis: how much inferential and decision-theoretic formalism should be used? A second issue that arises in this and many other policy problems is, at what stage in the process should enter the desire for adequate caution with regard to human risk? In arriving at a policy decision, scientific evidence must be assessed and social values must be evaluated and considered. Some argue that it is important to distinguish, as much as possible, the assessment phase from the evaluation and decision-making phase of policy formulation. A thorough assessment would then include descriptions of the consequences of proposed actions, the likelihoods of their occurrence, and the actions that should be taken under each possible set of assignments of value to the consequences. The use of Bayesian decision theory in this context is described in detail by Keeney and Raiffa (1976). A complication is that probabilities under various assumptions may differ widely (as they usually do for dose-response models). While, in principle, the probability of an outcome may be expressed as a mixture over the assumptions, in practice the probabilities of the validity of the assumptions may be extremely difficult to obtain. One solution to this problem would further complicate an already elaborate scheme by providing assessments according to each of the many possible combinations of

* Robert E. Kass is Assistant Professor, Department of Statistics, Carnegie-Mellon University, Pittsburgh, PA 15213.

assumptions that could reasonably be entertained. This seems to me to be the best way of facing the dangers mentioned above when sufficient time and resources are available. (Other, non-Bayesian, methodologies for expressing belief attempt to deal with some of these problems in the inferential context; see, e.g., Shafer 1982.)

On the other hand, varying degrees of formal structure may be thought suitable: where precise expressions of knowledge are dubious, informal assessment may be preferred. Furthermore, by anticipating value judgments, assessment may be simplified. A conservative strategy is often adopted: at each stage in modeling and analysis the route that leads to the greater estimate of human risk is taken. This approach is more restrictive than the Bayesian decision-theoretic procedure referenced above, in which appropriate prudence would be applied during the evaluation phase of the process.

The authors seem to have mixed sympathies with regard to conservativeness in analysis. Their methodology is developed as an alternative to the conservative procedure that uses only the worst case among species for a particular agent (see Crouch and Wilson 1981). Meanwhile, their analysis begins with the conservative assumption of linear dose-response curves and they consider some conservative features of their model appropriate (see Section 7).

The paper by DuMouchel and Harris gives much insight into the problem of combining information about human cancer from different but related sources, and it offers an advance from the method recommended by Crouch and Wilson. Here, as elsewhere, the Bayesian approach is especially helpful in forming a conceptual foundation for inference and decision making. As data on interspecies comparisons accumulate, formal methods that utilize the available information will become increasingly useful. The scheme laid out and discussed by the authors will then provide a Bayesian path from cognitive framework to policy analysis.

REFERENCES

CROUCH, EDMUND, and WILSON, RICHARD (1981), "Regulation of Carcinogens (with discussion)," *Risk Analysis*, 1, 47–66.

JEFFREYS, HAROLD (1967), *Theory of Probability* (3rd ed.), London: Oxford University Press.

KEENEY, RALPH, and RAIFFA, HOWARD (1976), *Decision Making With Multiple Objectives: Preferences and Value Tradeoffs*, New York: John Wiley.

SHAFER, GLENN (1982), "Lindley's Paradox (with discussion)," *Journal of the American Statistical Association*, 77, 325–351.

Rejoinder

WILLIAM H. DuMOUCHEL and JEFFREY E. HARRIS

We are grateful to Messrs. Krewski, Smith, and Kass for an insightful and stimulating discussion. Krewski's synopsis of the literature on carcinogenic risk assessment is especially welcome.

Smith and Kass inquire whether our exchangeable prior for the species-agent interactions might be too strong an assumption. Smith suggests more general forms of exchangeability. Kass finds exchangeability itself to be an oversimplification and urges prior grouping of agents or species along the lines of our Section 5. Kass cautions, however, that prior information about group membership may be sparse.

The exchangeability assumption treats distinctive effects symmetrically. Thus, in an assessment of the relevance of laboratory experiments to humans, inconsistent responses among a battery of animal tests are just as informative as inconsistencies between human and animal responses. It may be desirable to refine the exchangeability assumption by reference to current knowledge of species differences in bioavailability, metabolism, genetic repair mechanisms, and the like. In terms of strategy of presentation, however, we regarded exchangeability as an appropriate starting point.

In our illustrative example of Table 1, we could have assumed that whole animal experiments in mice were more relevant to humans than were mammalian cell experiments. But we recognized that the tumor initiation endpoint gauged in the Sencar mice studies might be no more relevant to the origin of human lung cancer from combustion mixtures than cell transformation or mutagenesis. Likewise, we had a vague inkling that cigarette smoke was less relevant to the other combustion mixtures because of its distinct chemical profile. But we were not sure if the presence of tobacco-specific nitrosamines or the relative paucity of polyaromatics in cigarette smoke were really critical factors. Our adoption of the naive exchangeability assumption, as well as our choice of a diffuse prior distribution for the relative potencies, reflected our desire not to impose vague and controversial beliefs on the reader at the outset.

Once the consequences of exchangeability were detailed, we were in a better position to study the effects of prior information in Section 5. In that section, we specifically assumed that the diesel emissions had biologically similar effects—a belief that we regarded as straightforward and uncontroversial. We showed how