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 grouping of the diesel emissions was translatable into informative priors on the parameters β without reference to the species-agent interactions δ. Still, Kass expresses hesitation about applying the model until additional prior information is marshalled. We would certainly agree that better information is always better, but feel compelled to note that a paucity of such prior information hardly precludes the use of our statistical methods.

There are ways to test exchangeability. As suggested in Section 6, we could examine other data sets, particularly those with epidemiological evidence. Among the agents that have been labeled as carcinogenic on the basis of epidemiologic evidence, many have not been subject to extensive whole animal and short-term testing. This is an important point. If the current abundance of toxicological testing is to be informative about human cancer risks, then such testing needs to be performed on agents for which human evidence already exists.

An alternative method, suggested by Smith, is to examine the consequences of relaxing our exchangeability assumption. In doing so, however, we need to be careful that the substitute assumption is truly less restrictive than that under scrutiny. A sharp prior belief that nonhuman experiments are completely irrelevant to humans (which Kass wishes to mix with our exchangeable prior) would indeed yield different results. But sensitivity to such an extreme prior restriction hardly constitutes a test of our model. Surely the mammalian experiments have some relevance to human cancer risks, and the biological effects of coke oven and other emissions have some relevance to the effects of diesel engine emissions. The question is exactly how much relevance.

We performed additional calculations within the following framework. Suppose that a subset of agents and species adheres less closely to the underlying regression model, but we remain uncertain which, if any, species or agents belong to the suspicious subset. In contrast to Smith's specific suggestion, and in accordance with Section 6, we allow only species or agents, rather than individual experiments, to be the exceptions. Suppose further that each species and agent has an independent, identical probability p of belonging to the subset of exceptions. (The species-agent interactions, in that case, still have identical univariate marginal prior distributions, but are no longer fully exchangeable.) Suppose finally that $\delta_{kl} \sim N(0, \tau^2)$ when species k or agent l is an exception, and $\delta_{kl} \sim N(0, \sigma^2)$ otherwise. We studied the case where $p = .1$ and $\tau - \sigma = \frac{1}{2} \log 10$, so that $\exp(2\tau) = 10 \exp(2\sigma)$. (In effect, each species or agent has a one-tenth chance of being ten times farther from the underlying regression model.) When we applied these specific assumptions to the 5×9 array of Table 1, retaining a diffuse prior on β, we obtained posterior densities of human carcinogenic potency quite close to those reported in Table 2, with increases in posterior standard deviation ranging from 0 to 10 percent. Some subsets of species and agents indeed exhibited high likelihoods of being exceptional. For example, the Bayes factor for the model in which cigarette smoke (CIG), mutagenesis without activator (M-), and skin tumor initiation (SKIN) constituted the suspicious subset was 33 times as large as that for the model with no exceptions. But because each species and agent had an independent, equal prior probability of being exceptional, the posterior estimates from the high likelihood models contributed little to the overall posterior estimates.

From the strict Bayesian viewpoint, therefore, the posterior standard deviations for the 3×8 array in Table 5 (where we dropped CIG, M-, and SKIN through a stepwise search) are overly optimistic. Such optimism is a difficulty with any stepwise search procedure, as Kass properly notes. Still, our diagnostic procedure can help decide which animal tests are most relevant to human cancer risks. It can also help identify which of many short-term tests are most relevant to carcinogenicity in laboratory animals. Such information would be valuable not only for "risk estimation," but also for "risk identification" (as Krewski uses the terms).

Krewski and Kass raise concerns about our statistical analysis of the individual cancer studies. As noted from the outset, we did not purport to solve the problem of selecting dose-response models for individual experiments. But we did note how errors of model misspecification enter into our analysis, and we specifically examined the effects of increasing the standard deviations of the nonhuman data in Table 1. Moreover, we explained how a multivariate generalization of our model would accommodate nonlinear dose-response curves, in which the results of an experiment would be summarized by a vector of numbers. Krewski warns that tests in the same species need to be performed under similar conditions. In fact, the data in Table 1 satisfy this requirement. (For example, the skin tumor counts in Sencar mice were all performed at 27 weeks.)

Nor did we purport to offer an extensive analysis of the uncertainties in the epidemiologic method. Although not noted in our article, Harris (1981,1983) had already increased the standard deviations for the epidemiologic studies in Table 1 in an attempt to account for uncertainties in occupational exposure and other possible confounding factors.

We apologize to Smith for labeling certain computational shortcuts as "empirical Bayes." The terminology was convenient and consistent with the literature. Our point was that such computational shortcuts may conceal the genuine uncertainty that arises when there are relatively few experiments. Perhaps Smith's suggested refinement, involving a second-order approximation for the density of $(\theta \mid Y, \sigma)$ about $\hat{\sigma}$, would avoid this difficulty. But our message should still be clear. Human cancer risk assessment requires data on many agents in many species. In the absence of strong prior information on cancer mechanisms, one good rat study is just not enough.

Our goal here was to impose a formal theoretical structure on the previously ill-defined problem of interspecies extrapolation. In this effort, three ideas were central.

First, there is a crucial distinction between the error of measurement within each dose-response study and the error of imperfect relevance among the studies. Second, the uncertain relevance between experiments can be formalized by reference to a hypothetical common mechanism that generates all the experimental data. Relevance is then quantifiable, roughly speaking, by the fit of the data to the underlying model. Third, the available biological information on species differences, chemical characteristics, disease mechanisms, and the like enters the analysis in the form of prior assumptions about the parameters of the underlying model relating the experiments. If we have made any contribution, we hope that these ideas persist in future research on interspecies comparisons. To be somewhat more presumptuous, we hope that our viewpoint finds its way to other problems where combining diverse forms of scientific evidence is at issue.