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

WILLIAM H. DuMOUCHEL and JEFFREY E. HARRIS*

We propose a class of Bayesian statistical methods for interspecies extrapolation of dose-response functions. The methods distinguish formally between the conventional sampling error within each dose-response experiment and a novel error of uncertain relevance between experiments. Through a system of hierarchical prior distributions similar to that of Lindley and Smith (1972), the dose-response data from many substances and species are used to estimate the interexperimental error. The data, the estimated error of interspecies extrapolation, and prior biological information on the relations between species or between substances each contribute to the posterior densities of human dose-response. We apply our methods to an illustrative problem in the estimation of human lung cancer risk from various environmental emissions.

KEY WORDS: Interspecies extrapolation; Carcinogenesis; Mutagenesis; Lung cancer; Empirical Bayes; Exchangeability.

1. INTRODUCTION

The assessment of cancer risks from environmental agents has been plagued by an abundance of precise data in nonhuman species, but little accurate information on direct effects in humans. While it is desirable to use all the available information to reduce the uncertainty about human health risks, the relevance of the nonhuman data to man is itself uncertain.

In this article we propose a class of Bayesian statistical methods for interspecies extrapolation of dose-response functions. The approach provides a general theory within which previous studies of interspecies extrapolation can be incorporated.

Our statistical methods are founded on the standard Bayesian paradigm for the analysis of covariance components models. We distinguish formally between the conventional measurement error within each dose-response experiment and a novel error of uncertain relevance between experiments. Through a system of hierarchical prior distributions similar to that of Lindley and Smith (1972), we use dose-response data from many substances and species to estimate the interexperimental error. From the data, the estimated error of interspecies extrapolation, and prior biological information on the relations between species or between substances, we calculate posterior densities of human dose-response.

We assume the availability of a set of independently performed studies on various substances in humans and other species. Each "experiment," we further assume, has already been analyzed according to a specific dose-response model. We do not address the problem of selecting dose-response models to analyze individual experiments. Nor do we belabor the well-known fact that the choice of dose-response model critically influences the predicted risk of cancer at very low exposures. However, as we discuss in Section 7, there is a formal connection between errors of model misspecification within individual experiments and errors of extrapolation between experiments.

The main message of this article is that convincing estimates of human cancer risk from nonhuman studies require evidence from both multiple substances and multiple species, including at least some informative data in humans. Human risk extrapolations based solely on the results of one compound in one nonhuman species will be highly sensitive to prior beliefs about the relevance of the nonhuman species to man.

Our strategy of presentation is to analyze a real problem in the assessment of human lung cancer risks from various environmental emissions. While we regard our specific quantitative results as a meaningful contribution, our main purpose is to use the data to point out generally the statistical and biological questions surrounding interspecies extrapolation.

The next section presents the data. In Section 3 we formally develop our approach; in Section 4 we apply our model to the data, drawing distinctions between a fully Bayesian analysis and various computational shortcuts. In Section 5 we show how prior biological information can be incorporated in our Bayesian framework, and in

* William H. DuMouchel is Associate Professor, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139. Jeffrey E. Harris is Associate Professor, Department of Economics, Massachusetts Institute of Technology, and Clinical Associate, Massachusetts General Hospital, Boston. DuMouchel's research was supported by National Science Foundation Grant MCS-80-05483. Harris's research was supported by Public Health Service Research Grant DA-02620 and by Research Career Development Award DA-00072. A preliminary version of this paper was presented at the 21st meeting of the NBER-NSF Seminar on Bayesian Inference in Econometrics, Chicago, October 31, 1980, and appeared as Technical Report No. 24, Department of Mathematics, Massachusetts Institute of Technology, February 1981. The authors gratefully acknowledge the comments and suggestions of H. Chernoff, D. Freedman, R. Olshen, J. Koziol, D. Krantz, R. Solow, J. Yellin, and six referees.

June 1983, Volume 78, Number 382
Invited Papers Section

Section 6 we explore the problem of deciding which experiments are most relevant to human cancer risks.

2. AN ILLUSTRATIVE PROBLEM

Table 1 displays the results of five types of studies on nine related environmental agents, arranged in a two-way table. Experimental data are available in 36 of the 45 cells in the table. The first row of experiments represents the results of epidemiological studies of human lung cancer incidence in relation to occupational exposure to roofing tar emissions (Hammond et al. 1976); occupational exposure to coke oven emissions (Lloyd 1971; Mazumdar et al. 1975; U.S. Environmental Protection Agency 1979); and cigarette smoking (Kahn 1966). The remaining rows represent the results of various laboratory experiments on the dichloromethane extracts of roofing tar emissions, coke oven emissions, particulate emissions from four different diesel engines and one gasoline-powered engine; on the polyaromatic hydrocarbon benzo(a)pyrene; and on whole cigarette smoke condensate. The laboratory experiments include skin tumor initiation in Sencar mice (Nesnow et al. 1981); enhancement of viral oncogenic transformation in Syrian hamster embryo (SHE) cells (Casto et al. 1980); and mutagenesis experiments in L5178Y mouse lymphoma cells performed with and without metabolic activator (Mitchell et al. 1980). The laboratory experiments in each row were conducted under identical conditions, as part of the U.S. Environmental Protection Agency's diesel emission research program (Huisingh et al. 1980).

The choice of experiments in Table 1 reflected the limited availability of experimental data comparing diesel emissions to other related environmental emissions in the same bioassay. An assay was included in the table only if (a) it was considered to have been performed reliably and reproducibly; (b) sufficient data were provided to estimate a dose-response relation; and (c) the experiment was regarded as a valuable, quantitative measure of carcinogenicity or mutagenicity in mammalian systems (Harris 1981,1983).

The entries in each nonempty cell represent the results of fitting a linear dose-response model to the observed experimental data. For each experiment, three numbers are given: the estimated slope of the dose-response relation; its coefficient of variation (the ratio of the standard error of the estimated slope to the estimate itself); and the natural logarithm of the estimated slope. (The role of these log slopes will become apparent shortly.) In our analysis of these data, we shall interpret the dose-response slope to be a measure of the potency of a given agent in a given species. The slopes and their standard errors were estimated by maximum likelihood methods, as described in Harris (1981,1983). Appendix A briefly summarizes the dose-response models and their estimation.

Table 1 provides relatively precise data on the human effects of cigarette smoke and coke oven emissions. But the effect of roofing tar is not precisely estimated, and for the remaining agents we have no human data. The main question is how to use all of the evidence in Table 1 to obtain more precise estimates of the human lung cancer risks.

To answer this question we need to posit an underlying common mechanism or hypothesis that generates all the data in the table. Specifically, we shall examine the hy-

Table 1. Two-Way Table of Epidemiological Studies and Laboratory Experiments on Nine Environmental Mixtures[a]

	Roofing Tar Emissions	Coke Oven Emissions	Diesel Engine Emissions				Gasoline Engine Emissions	Benzo(a) pyrene	Cigarette Smoke[f]
			A	B	C	D			
Lung Cancer	1.64	4.40							.03
(Humans)[b]	1.41	.34							.15
	.50	1.48							−3.46
Skin Tumor	.54	2.10	.53	.16		.01	.03	85.28	.00
Initiation (Sencar	.04	.04	.04	.22		.82	.26	.03	1.30
Mice)[c]	−.63	.74	−.64	−1.86		−4.51	−3.61	4.45	−5.88
Enhancement of	2.07	.86	.65	.07	.13	.04	.20	540.00	.58
Viral Transform.	.18	.10	.15	.33	.18	.59	.12	.04	.08
(SHE cells)[d]	.73	−.15	−.44	−2.70	−2.06	−3.24	−1.59	6.29	−.54
Mutagenesis −MA	.31	.73	1.66	.27	2.55	.16	.35		.59
(Mouse 5178Y	.39	.21	.31	.43	.16	.24	.11		.23
Lymphoma Cells)[e]	−1.17	−.32	.51	−1.31	.93	−1.86	−1.06		−.53
Mutagenesis +MA	9.56	9.96	1.87	.76	1.01	.05	.99		.45
(Mouse 5178Y	.16	.07	.26	.14	.20	.43	.10		.13
Lymphoma Cells)[e]	2.26	2.30	.63	−.27	.01	−3.02	−.01		−.79

[a] Each nonempty cell contains the observed dose-response slope, its coefficient of variation, and the natural logarithm of the dose-response slope.
[b] Units of measurement (for all agents except cigarette smoke) are the increment in the relative risk per 10^8 $\mu g/m^3$ of organic extractables per year.
[c] Units of measurement are papillomas per mouse per mg of organic extractables at 27 weeks.
[d] Units of measurement are transformations per 2×10^6 surviving cells per $\mu g/ml$ of organic extractables; "SHE" = Syrian hamster embryo.
[e] Units of measurement are average mutant colonies per 10^6 survivors per $\mu g/ml$ of organic extractables; "−MA" = without metabolic activator; "+MA" = with metabolic activator.
[f] Units of measurement for cigarette smoke refer to whole smoke condensate rather than organic extractables.

pothesis that the ratio of dose-response slopes (that is, the relative potency) of any two environmental emissions is preserved across species. This hypothesis arises naturally from the data of Table 1 because the dose-response slopes in each species are measured in the same units, and therefore their ratios are comparable unitless quantities. (Although the results in the last column are measured per unit of cigarette smoke condensate rather than per unit of extractable organics, the ratios of the slopes for cigarette smoke to the other slopes are similarly in identical units.) We therefore avoid the need to specify potentially complex or implausible conversions of dosage units between species.

Even a cursory examination of Table 1 suggests that the constant relative potency hypothesis does not fit the data all that well. Thus, coke oven emissions are apparently carcinogenic in humans, while the coke oven extract initiates skin tumors in mice. Since the diesel A extract also initiates skin tumors, with about one-quarter the potency of the coke oven extract, then diesel A emissions should likewise be human carcinogens, with about one-quarter the potency of coke oven emissions. But cigarette smoke is carcinogenic in man, while its condensate is at best a weak skin tumor initiator in mice. Moreover, in the mutagenesis experiments on mouse lymphoma cells, the addition of metabolic activator markedly increases the effect of the coke oven extract, whereas the response of the diesel extract is less consistent.

These findings are not surprising. Carcinogenic processes vary considerably across environmental agents and species. There are differences in chemical composition, bioavailability, target site of action, metabolism, and genetic repair mechanisms. Any distinctive action of a particular agent in a particular species would constitute a deviation from the hypothesis of constant relative potency. Such an hypothesis at best approximates the relation between the experiments. The question is: can we use the data of Table 1 to gauge the accuracy of the approximation?

A partial answer is given in Figure 1, which depicts on a logarithmic scale the means and standard errors of the dose-response slopes from the upper left 3 × 3 submatrix of Table 1. The error bars correspond to the standard errors of the log slopes, which have been approximated by the coefficients of variation in the table. Since the three vertical axes are drawn to the same scale, the hypothesis of constant relative potency requires that the lines connecting successive pairs of estimates be parallel across species. We have therefore inserted the weighted least squares parallel lines, the weights having been chosen according to a procedure to be described in Section 3. The deviations of the fitted lines from the observed data points in Figure 1 suggest that the hypothesis of constant relative potency, as a model of the relation between experiments, could be no more accurate than a factor of 2.

Underlying this conclusion is a critical assumption. Since the nonhuman data in Figure 1 are more precise

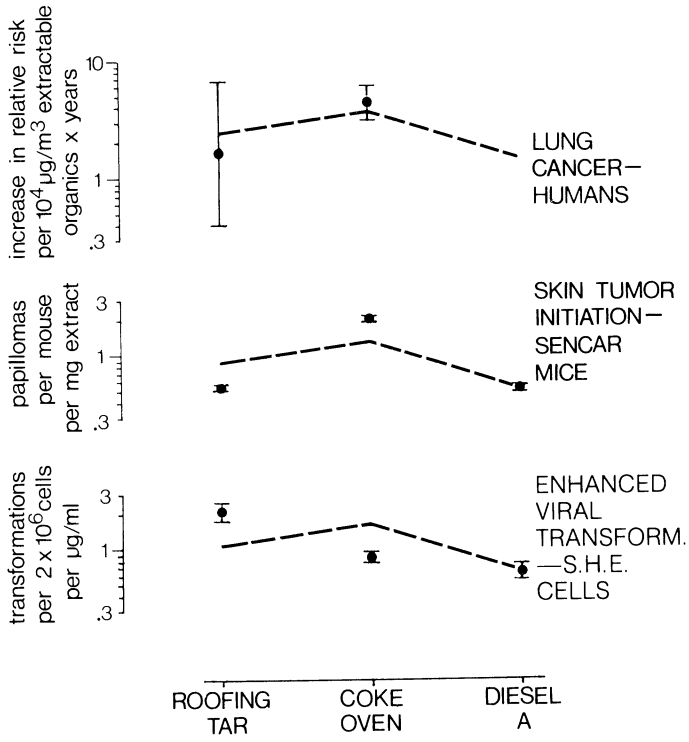


Figure 1. ACCURACY OF THE CONSTANT RELATIVE POTENCY HYPOTHESIS FOR THE UPPER LEFT 3 × 3 SUBMATRIX OF TABLE 1. The means and standard error bars for the dose-response slopes are plotted on a logarithmic scale. Superimposed are the weighted least squares parallel lines, as described in the text.

than the epidemiological results, our test of the relevance of the nonhuman experiments to the humans relies heavily on the relevance of one nonhuman experiment to another. Put differently, any observed species-emission interaction gives us some information on the probability of a distinctive effect in humans.

In extrapolating nonhuman results to man, some scientists might place greater reliance on in vivo studies in mammals than on in vitro studies in mammalian cells or unicellular organisms. But the argument that whole animal experiments, such as skin tumor initiation in mice, are better models of carcinogenicity in man is not based on systematic empirical study (Purchase 1980). In fact, it is hardly obvious that the skin painting studies in Figure 1 are any more or less relevant to humans than the cell transformation experiments, or that the human epidemiologic studies are any more or less relevant to mouse skin than the hamster embryo cell experiments. Nor is it apparent a priori that the coke oven results are any more or less relevant to diesel emissions than the roofing tar results. To go beyond these expressions of ignorance requires sharp prior information on chemical composition, metabolism, and other biological characteristics.

Assumptions of minimal prior information have been implicit in previous studies of interspecies comparison of carcinogenicity. And most have implicitly used the notion of constant relative potencies. Thus, Meselson and Russell (1977) compared the log mutagenic potency in the Salmonella Ames test against the log carcinogenic po-

tency in mice for 14 compounds. Clive (1977) compared the log mutagenic potency in the L5178Y mouse lymphoma assay against the log carcinogenic potency in mice for 23 chemicals. Crouch and Wilson (1979) compared log carcinogenic potencies of various chemicals in several pairs of species, most extensively in rats and mice. Numerous investigators have studied the concordance of positive and negative test results among many test compounds in various species (McCann et al. 1975; Purchase et al. 1978; Rinkus and Legator 1979; Poirier and de Serres 1979; Purchase 1982). Freireich et al. (1966) compared the toxicity of anti-cancer agents in mouse, rat, hamster, dog, monkey, and man. These authors implicitly assumed that each cell in a two-way table of experiments constitutes an independent, equally likely draw from a population of species-chemical interactions. Crouch and Wilson (1981) explicitly suggest that data on the variability of relative potencies in rats and mice can be used to assess variations between humans and other species. Likewise, Purchase (1980) concludes that beyond detailed knowledge of the fate of a chemical in man, the only additional evidence supporting human carcinogenicity is the consistency of carcinogenic response across other species.

In what follows, therefore, we put forth the notion that the experiments in Table 1 are all related through some unifying biological hypothesis. More formally, we assume that the results of each experiment are summarized by a single number, such as the slope of the dose-response relation. Each slope is subject to conventional measurement error. The actual slopes, we posit, lie near the response surface of an underlying regression model. The unknown parameters of this model correspond to the uncertain relative potencies among emissions and the uncertain conversion factors between species. Moreover, since some agents have distinctive effects in some species, such a model necessarily entails some error. These errors gauge the uncertain relevance of one experiment to another. The critical factor linking the experiments is the scientist's prior information on the exchangeability (or otherwise) of these errors of interspecies extrapolation.

3. THE STATISTICAL MODEL

Let y_{kl} denote the logarithm of the estimated dose-response slope for the experiment in species k on environmental agent l, where $k = 1, \ldots, K$ and $l = 1, \ldots, L$. Although some y_{kl} may be missing from the two-way table of experiments, we assume that the set of observed y_{kl} is connected. (That is, any observed (k, l) can be reached from any other observed (k', l') by a series of moves from one observed experiment to another in which each move is along a single row or column. Although connectedness is not strictly required by the theory, in practice unless strong prior information is used, two unconnected experiments cannot each contribute information about any single parameter and thus may as well be analyzed separately.)

Each y_{kl}, we further assume, is normally distributed with unknown mean θ_{kl} and known standard error c_{kl}. This assumption is motivated by the fact that the observed y_{kl} in Table 1 are maximum likelihood estimates derived from relatively large experiments.

The quantities θ_{kl} are the true log slopes, the primary parameters of interest. Each θ_{kl}, we assume, has a prior distribution with mean value

$$E[\theta_{kl} \mid \mu, \alpha_k, \gamma_l] = \mu + \alpha_k + \gamma_l, \tag{3.1}$$

where the parameters $\{\mu, \alpha_k, \gamma_l\}$ represent the overall mean effect, species-specific effects, and agent-specific effects, respectively. Equation (3.1) embodies the hypothesis that the relative potency of the two emissions is on average preserved across species. Although the various θ_{kl} are measured in different units, the additive model (3.1) is meaningful so long as $(\theta_{kl} - \theta_{kl'}) - (\theta_{k'l} - \theta_{k'l'})$ is a dimensionless quantity, a condition satisfied by the data of Table 1. The units of measurement for μ, α_k, and γ_l can then be chosen so that the species-agent interaction effects $\delta_{kl} = \theta_{kl} - \mu - \alpha_k - \gamma_l$ are similarly dimensionless.

Conditional on the value of the parameter σ, we further assume that the δ_{kl} are a priori independently and identically normally distributed with zero mean and variance σ^2. Accordingly, the interaction effects δ_{kl} are exchangeable (de Finetti 1937). Although certain agents may have distinctive actions in particular species, we cannot identify a priori which entry in our two-way table of experiments is more likely to exhibit a distinctive action. Nor do we have any prior information that any pair of interaction effects are in the same direction.

The parameter σ gauges the accuracy of the equal relative potency model. Since the quantities $\exp(\delta_{kl})$ represent dimensionless, multiplicative deviations from the model, 95 percent of such deviations would be within $\exp(\pm 2\sigma)$. For example, a value of $\sigma = 2.3$ would imply that with probability .95, the model would be accurate to a factor of $\exp(2\sigma) \cong 100$ in either direction.

Conditional on σ, the observed y_{kl} are thus generated by

$$y_{kl} = \mu + \alpha_k + \gamma_l + \delta_{kl} + \epsilon_{kl}, \tag{3.2}$$

where the three sets of variables $\{\mu, \alpha_k, \gamma_l\}$, $\{\delta_{kl}\}$ and $\{\epsilon_{kl}\}$ are independent a priori, and where the ϵ_{kl} are independent $N(0, c_{kl}^2)$. Following the notation for the general linear model, we replace the expressions $\mu + \alpha_k + \gamma_l$ in (3.2) by $X\beta$, where β is a column vector of parameters and X is an appropriately chosen design matrix. Of the $K + L + 1$ parameters in $\{\mu, \alpha_k, \gamma_l\}$, at most $K + L - 1$ are independently estimable in the classical sense. So long as we use an informative full rank prior distribution on all $K + L + 1$ parameters, no restrictions on the parameters are necessary. However, in certain cases to be considered below (Proposition 2 and the empirical Bayes estimates), we shall assume that β corresponds only to the independently estimable components of

$\{\mu, \alpha_k, \gamma_l\}$ and that X is the corresponding full rank design matrix.

Finally, we assume that β is a priori multivariate normal with mean vector b and nonsingular covariance matrix V, which are independent of σ, and that σ has a prior distribution Π with density $\pi(\sigma)$. Now let $i = 1, \ldots, n$ index the experiments, replacing the paired indices (k, l). Let m be the rank of X. Let Y, θ, δ, and ϵ be $n \times 1$ column vectors replacing $\{y_{kl}\}$, $\{\theta_{kl}\}$, $\{\delta_{kl}\}$, and $\{\epsilon_{kl}\}$, respectively. Let I be the $n \times n$ identity matrix, and let $C = \text{diag}(c_1^2, \ldots, c_n^2)$. Our model can be formulated generally as $Y = X\beta + \delta + \epsilon$, where $\theta = X\beta + \delta$, and

$$\sigma \sim \Pi, \tag{3.3a}$$

$$(\beta \mid \sigma) \sim N(b, V), \tag{3.3b}$$

$$(\theta \mid \beta, \sigma) \sim N(X\beta, \sigma^2 I), \tag{3.3c}$$

$$(Y \mid \theta, \beta, \sigma) \sim N(\theta, C). \tag{3.3d}$$

The experimental data Y and C, as well as b, V, and the distribution Π, are assumed to be known. The choices of a prior distribution Π and of b and V are left unspecified for now.

This model possesses a hierarchical structure similar to that formulated by Lindley and Smith (1972). Since estimation of the parameters of (3.3) proceeds along familiar Bayesian lines, we state the following proposition without proof.

Proposition 1. For the model (3.3), the posterior density of σ is

$$\pi(\sigma \mid Y) \propto \pi(\sigma) \mid C + \sigma^2 I + XVX' \mid^{-1/2} \times \exp\{-\tfrac{1}{2}(Y - Xb)'[C + \sigma^2 I + XVX']^{-1}(Y - Xb)\} \tag{3.4}$$

where $\mid A \mid$ is the determinant of A. Moreover, the posterior density of β is $f(\beta \mid Y) = \int_0^\infty f(\beta \mid Y, \sigma)\pi(\sigma \mid Y)d\sigma$, where $f(\beta \mid Y, \sigma)$ is the multivariate normal density $N(\tilde{\beta}, \tilde{V})$, and

$$\tilde{\beta} = \tilde{V}[X'WY + V^{-1}b], \qquad \tilde{V} = [X'WX + V^{-1}]^{-1}, \tag{3.5}$$

where $W = (C + \sigma^2 I)^{-1}$. Finally, the posterior density of θ is $g(\theta \mid Y) = \int_0^\infty g(\theta \mid Y, \sigma)\pi(\sigma \mid Y)d\sigma$, where $g(\theta \mid Y, \sigma)$ is the multivariate normal density $N(\tilde{\theta}, \tilde{C})$, and

$$\tilde{\theta} = \tilde{C}[C^{-1}Y + (XVX' + \sigma^2 I)^{-1}Xb], \qquad \tilde{C} = [C^{-1} + (XVX' + \sigma^2 I)^{-1}]^{-1}. \tag{3.6}$$

The posterior distributions of both β and θ are mixtures of multivariate normal distributions, with mixing probabilities given by (3.4), the posterior density of σ. From (3.5), $\tilde{\beta}$ is the precision-weighted average of its least squares estimate and its prior mean b. From (3.6), $\tilde{\theta}$ is the precision-weighted average of the original data Y and the corresponding prior mean Xb.

The results in Proposition 1 depend upon our specifying a proper prior distribution for β. Specifying a prior distribution on β requires knowledge not only about the relative potencies of agents $\{\gamma_l\}$, but also about the conversion factors of dosage units among species $\{\alpha_k\}$. In many situations, such information could be extremely vague, and a more uninformative prior that avoids complex or implausible conversions of interspecies dosage units would be desirable. To investigate such cases in detail, we offer the following proposition, which is proved in Appendix B.

Proposition 2. If V is replaced by tV, where t is a scalar, then as $t \to \infty$ the posterior density of σ approaches

$$\pi(\sigma \mid Y) \propto \pi(\sigma) \mid W \mid^{1/2} \mid X'WX \mid^{-1/2} \exp\{-\tfrac{1}{2}Y'SY\} \tag{3.7}$$

where $S = W - WX(X'WX)^{-1}X'W$, and where X is the full rank design matrix. Moreover, the posterior density of β approaches $f(\beta \mid Y) = \int_0^\infty f(\beta \mid Y, \sigma)\pi(\sigma \mid Y)d\sigma$, where $f(\beta \mid Y, \sigma)$ is multivariate normal $N(\tilde{\beta}, \tilde{V})$ and

$$\tilde{\beta} = \tilde{V}X'WY, \qquad \tilde{V} = (X'WX)^{-1}. \tag{3.8}$$

Finally, the posterior density of θ approaches $g(\theta \mid Y) = \int_0^\infty g(\theta \mid Y, \sigma)\pi(\sigma \mid Y)d\sigma$, where $g(\theta \mid Y, \sigma)$ is multivariate normal $N(\tilde{\theta}, \tilde{C})$ and

$$\tilde{\theta} = \tilde{C}C^{-1}Y = (I + CR)^{-1}Y, \qquad \tilde{C} = (C^{-1} + R)^{-1}, \tag{3.9}$$

where $R = \sigma^{-2}[I - X(X'X)^{-1}X']$.

We also consider empirical Bayes approaches to estimating θ. For these estimates, we retain (3.3c) and (3.3d), but use the data Y to construct the prior distributions on σ and β. Several options are available. First, we could estimate both σ and β from the data Y, by maximum likelihood or other methods, and then assume that the prior density for σ is concentrated at the estimate $\hat{\sigma}$ and the prior density of β is concentrated at the estimate $\hat{\beta}$. For a given σ, the maximum likelihood estimate of β is the least squares estimate

$$\hat{\beta}_{\text{MLE}}(\sigma) = (X'WX)^{-1}X'WY,$$

where W depends on σ. The concentrated likelihood of σ, evaluated at $\beta = \hat{\beta}_{\text{MLE}}(\sigma)$, is proportional to

$$\hat{L}(\sigma) = \mid W \mid^{1/2} \exp\{-\tfrac{1}{2}Y'SY\}$$

where W and S depend on σ. Let $\hat{\sigma}_{\text{MLE}}$ be the value of σ that maximizes $\hat{L}(\sigma)$. Then the resulting empirical Bayes posterior distribution of θ, which treats both $\sigma = \hat{\sigma}_{\text{MLE}}$ and $\beta = \hat{\beta}_{\text{MLE}}(\hat{\sigma}_{\text{MLE}})$ as if they were known with certainty, is $N(\hat{\theta}_{\text{MLE}}, \hat{C}_{\text{MLE}})$, where

$$\hat{\theta}_{\text{MLE}} = \hat{C}_{\text{MLE}}[C^{-1}Y + \hat{\sigma}_{\text{MLE}}^{-2}X\hat{\beta}_{\text{MLE}}(\hat{\sigma}_{\text{MLE}})], \qquad \hat{C}_{\text{MLE}} = [C^{-1} + \hat{\sigma}_{\text{MLE}}^{-2}I]^{-1}. \tag{3.10}$$

Alternatively, we could assume a diffuse prior on β and estimate only σ from the data Y. In that case, the appropriate likelihood function for σ is equation (3.7) with the

prior density $\pi(\sigma)$ omitted, that is,

$$L^*(\sigma) = |W|^{1/2} |X'WX|^{-1/2} \exp\{-\tfrac{1}{2} Y'SY\}.$$

Let $\hat{\sigma}_{EB}$ be the value of σ that maximizes $L^*(\sigma)$. The corresponding empirical Bayes posterior distribution of θ, which treats $\sigma = \hat{\sigma}_{EB}$ as if it were known with certainty, is $N(\hat{\theta}_{EB}, \hat{C}_{EB})$, where

$$\hat{\theta}_{EB} = [I + \hat{\sigma}_{EB}^{-2} C(I - X(X'X)^{-1}X')]^{-1} Y$$
$$\hat{C}_{EB} = [C^{-1} + \hat{\sigma}_{EB}^{-2}(I - X(X'X)^{-1}X')]^{-1}. \tag{3.11}$$

(In case $\hat{\sigma}_{EB}^2$ is near or equal to zero, the lemma of Appendix B can be used to show that the limiting values in (3.11) are $\hat{\theta}_{EB} = X\hat{\beta}_{MLE}(0)$ and $\hat{C}_{EB} = X(X'C^{-1}X)^{-1}X'$.)

Comparing (3.10) and (3.11), we note that $\hat{C}_{MLE} < \hat{C}_{EB}$, in the sense that $\hat{C}_{MLE} - \hat{C}_{EB}$ is negative definite. (To establish this result, it is enough to show that $\hat{\sigma}_{MLE} < \hat{\sigma}_{EB}$. The latter follows from the fact that $\hat{\sigma}_{MLE}$ and $\hat{\sigma}_{EB}$ maximize $\hat{L}(\sigma)$ and $L^*(\sigma)$, respectively, and that the ratio $\hat{L}(\sigma)/L^*(\sigma) = |X'(C + \sigma^2 I)^{-1} X|^{1/2}$ is a decreasing function of σ.) Thus, the assumption that $\beta = \hat{\beta}_{MLE}$ with certainty leads to a smaller estimated posterior variance for θ than does the assumption of a diffuse prior for β.

If we wish to avoid the computational burden of determining $\hat{\sigma}_{MLE}$ or $\hat{\sigma}_{EB}$, we could begin with $\hat{\beta}_{MLE}(0) = (X'C^{-1}X)^{-1}X'C^{-1}Y$. The residual sum of squares for this estimate, $RSS = \sum_{i=1}^{n} (y_i - x_i\hat{\beta})^2/c_i^2$, has expectation

$$E[RSS] = n - m + \sigma^2\left[\sum_{i=1}^{n} c_i^{-2} - \operatorname{tr} C^{-1}X(X'C^{-1}X)^{-1}X'C^{-1}\right]$$

which suggests the estimate

$$\hat{\sigma}_{RSS}^2 = [RSS - (n - m)] \Big/ \left[\sum_{i=1}^{n} c_i^{-2} - \operatorname{tr} C^{-1}X(X'C^{-1}X)^{-1}X'C^{-1}\right] \tag{3.12}$$

where we take $\hat{\sigma}_{RSS}^2 = 0$ if $RSS < n - m$. (The value of $E[RSS]$ was derived for us by H. Chernoff.) In the analysis that follows, we report the empirical Bayes estimate $\hat{\theta}_{RSS}$, derived by substituting $\hat{\sigma}_{RSS}$ for $\hat{\sigma}_{EB}$ in (3.11).

For future reference, we also define the posterior expectation of σ^2 as

$$\sigma^{*2} = \int_0^\infty \sigma^2 \pi(\sigma \mid Y) d\sigma, \tag{3.13}$$

where $\pi(\sigma \mid Y)$ is defined either by (3.4) or by (3.7). The quantity σ^{*2} is the mean squared error of prediction (on the log scale) for an experiment yet to be performed.

Results similar to Propositions 1 and 2 have been derived by other authors. Lindley and Smith (1972) and Smith (1973) describe a "modal Bayesian" estimation procedure, which corresponds in our model to the use of $\hat{\sigma}_{MLE}$ for $\hat{\sigma}_{EB}$ in (3.11). Our (3.9) is a special case of equation (A.1) in Smith (1973). Our likelihood function $L^*(\sigma)$ is equivalent to that derived by Harville (1974, eq. 3) for the mixed variance components model. Our empirical Bayes estimate $\hat{\sigma}_{EB}$ (the modal value of L^*) corresponds to the "restricted" maximum likelihood estimate of variance components proposed by Patterson and Thompson (1971). (See also Harville 1977.) Dempster, Rubin, and Tsutakawa (1981) analyze a covariance components model similar to ours, in which the fixed effects can assume a diffuse prior. With suitable modification, their EM-algorithm estimate of the variance of the random effects is equivalent to our $\hat{\sigma}_{EB}^2$.

Our analysis has relied upon the assumption that, conditional upon σ, the vector δ is distributed a priori as $N(0, \sigma^2 I)$. In anticipation of Section 5, however, we note that the covariance matrix $\sigma^2 I$ could be replaced by a more general matrix Σ. In place of (3.3a), the parameters in Σ would then have joint prior distribution Π. In Proposition 1, expression (3.4) would give the joint posterior density of these parameters. This density would serve as the mixing probabilities for the posterior distributions of β and θ. Proposition 2 would follow analogously, with $W = (C + \Sigma)^{-1}$ and $R = \Sigma^{-1} - \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$. Although the derivations of $\hat{\theta}_{EB}$ and $\hat{\theta}_{MLE}$ would follow a similar generalization, there is no analog for $\hat{\theta}_{RSS}$.

We have also assumed that β is a priori independent of σ. But our results are easily generalized to the case where b and V are replaced by $b(\sigma)$ and $V(\sigma)$. Finally, we assumed that each ϵ_i has known variance c_i^2. Yet our results are similarly generalizable to the case where the c_i are also uncertain and have prior distributions.

4. APPLYING THE MODEL

In the following illustrative calculations, we assume a diffuse prior distribution on the parameters β. The use of prior biological information to formulate proper priors on β is deferred to Section 5.

For the critical parameter σ, we consider two alternative prior distributions.

Prior 1: $\log \sigma$ is uniformly distributed on the interval $.05 \le \sigma \le 5.0$.

This prior distribution implies that within a range of one normal standard deviation, that is, with probability .68, the underlying constant relative potency model could be accurate within a multiplicative factor ranging from $\exp(.05) \cong 1.05$ up to $\exp(5) \cong 150$. Given our uncertainty about even the order of magnitude of error, we have assumed that $\log \sigma$, rather than σ, is uniformly distributed. Although Prior 1 assigns zero probability outside the interval [.05, 5], this restriction, as we shall see shortly, does not significantly affect our main conclusions. Prior 1 has the property that for $.05 \le \sigma \le 5$, the posterior density of $\log \sigma$ is proportional to the likelihood function L^*.

Prior 2: log σ is normally distributed with mean and standard deviation both equal to log 2.

This prior distribution has a modal value of $\sigma = 2$, with 95 percent probability in the interval $.5 \leq \sigma \leq 8$. Accordingly, with probability .68, the underlying constant relative potency model could be accurate within a multiplicative factor ranging roughly from $\exp(.5) \cong 1.7$ to $\exp(8) \cong 3000$, with an error factor of $\exp(2) \cong 7.4$ being most likely. Prior 2 is therefore more conservative than Prior 1 about the accuracy of interspecies extrapolation.

To simplify the computations, we evaluate Priors 1 and 2 and their corresponding posterior distributions only at discrete points equally spaced on the log scale. The posterior distributions of β and θ are then finite mixtures of normal distributions.

Figure 2 displays the posterior densities $\pi(\sigma \mid Y)$ based upon Priors 1 and 2. For both priors, virtually all of the posterior probability mass of σ is concentrated in the interval [.8, 1.6]. From the Bayesian estimate $\sigma^* \cong 1.1$ we calculate that, with 95 percent credibility, extrapolation will likely to be accurate only within a multiplicative factor of $\exp(2\sigma^*) \cong 9$.

Table 2 shows the posterior means and standard deviations of the log carcinogenic potencies of roofing tar emissions, coke oven emissions, and diesel A emissions in humans. Although the marginal posterior distributions of the θ_i are mixtures of normals, the distributions in this example are adequately characterized by their means θ_i^* and standard deviations c_i^*. That is, the tail probabilities $\Pr\{\theta_i \geq \theta_i^* + 2.326c_i^* \mid Y\}$ and $\Pr\{\theta_i \leq \theta_i^* - 2.326c_i^* \mid Y\}$ do not deviate substantially from the value of .01 predicted for the normal density.

Table 2. Bayes and Empirical Bayes Estimates of Log Slopes for Lung Cancer Risk in Humans: 5 × 9 Data Matrix

Environmental Emission	Posterior Mean	Posterior Standard Deviation	Lower Tail	Upper Tail
Roofing Tar				
Original Data	.50	1.41		
$\theta_{Bayes\ 1}$	.12	1.02	.010	.011
$\theta_{Bayes\ 2}$	.14	1.03	.010	.011
θ_{EB}	.12	1.01		
θ_{RSS}	.06	.95		
θ_{MLE}	−.01	.70		
Coke Oven				
Original Data	1.48	.34		
$\theta_{Bayes\ 1}$	1.38	.33	.010	.010
$\theta_{Bayes\ 2}$	1.38	.33	.010	.010
θ_{EB}	1.38	.33		
θ_{RSS}	1.35	.33		
θ_{MLE}	1.30	.31		
Diesel Engine A				
$\theta_{Bayes\ 1}$	−.46	1.45	.011	.012
$\theta_{Bayes\ 2}$	−.45	1.50	.011	.012
θ_{EB}	−.46	1.40		
θ_{RSS}	−.51	1.25		
θ_{MLE}	−.57	.80		

NOTE $\theta_{Bayes\ 1}$ assumes Prior 1 for $\pi(\sigma)$, and diffuse prior on β $\theta_{Bayes\ 2}$ assumes Prior 2 for $\pi(\sigma)$, and diffuse prior on β θ_{EB} assumes prior $\pi(\sigma)$ concentrated at $\hat{\sigma}_{EB} = 1.04$, and diffuse prior on β θ_{RSS} assumes prior $\pi(\sigma)$ concentrated at $\hat{\sigma}_{RSS} = 0\ 93$, and diffuse prior on β θ_{MLE} assumes prior $\pi(\sigma)$ concentrated at $\hat{\sigma}_{MLE} = 0\ 80$, and prior on β concentrated at $\hat{\beta}_{MLE}$. Lower Tail = $\Pr\{\theta \leq \theta^* - 2\ 326c^* \mid Y\}$, Upper Tail = $\Pr\{\theta \geq \theta^* + 2\ 326c^* \mid Y\}$

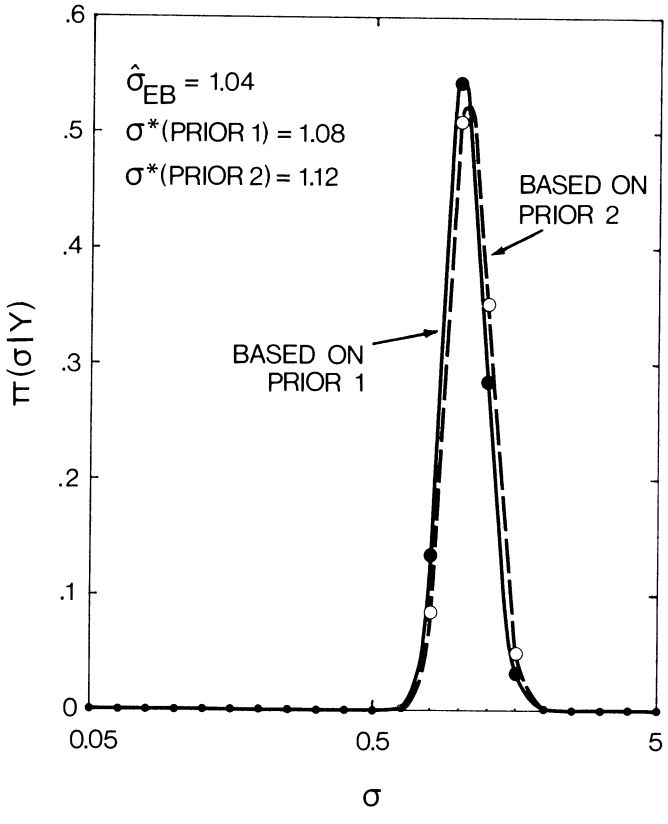


Figure 2. POSTERIOR DENSITIES OF log σ FOR THE DATA OF TABLE 1. The posterior densities are evaluated at points equally spaced (by .1 log 10) on the log scale.

Because the original coke oven data were relatively precise, the means and standard deviations of the coke oven log slope do not differ much from their original values. For roofing tar, however, there is a substantial improvement in precision. For the Bayes estimate based on Prior 1, the 95 percent credible interval for the carcinogenic potency of roofing tar (that is, e^{θ}) is narrowed from [.097, 27.8] to [.148, 8.66]. From the dosage measure given in Appendix A, the 95 percent credible interval for the relative risk of lung cancer from 20 years of occupational exposure to roofing tar is narrowed from [1.11, 31.6] with mode 2.80, to [1.16, 10.5] with mode 2.24.

The estimates for diesel engine A illustrate the dependence of the posterior precision of θ on our prior assumptions about σ and β. Thus, the Bayes estimate based on the conservative Prior 2 has a slightly larger posterior standard deviation than that based on Prior 1. Similarly, $\hat{\theta}_{EB}$ and $\hat{\theta}_{RSS}$, which assume fixed, known values of σ, have smaller posterior variances than the Bayes estimates. The estimate $\hat{\theta}_{MLE}$, based on the dubious assumption that both β and σ are fixed and known, yields on overly optimistic estimate of precision.

Despite the divergence of Priors 1 and 2, the posterior densities in Figure 2 are almost identical. This finding is to be contrasted with Figure 3, which displays the pos-

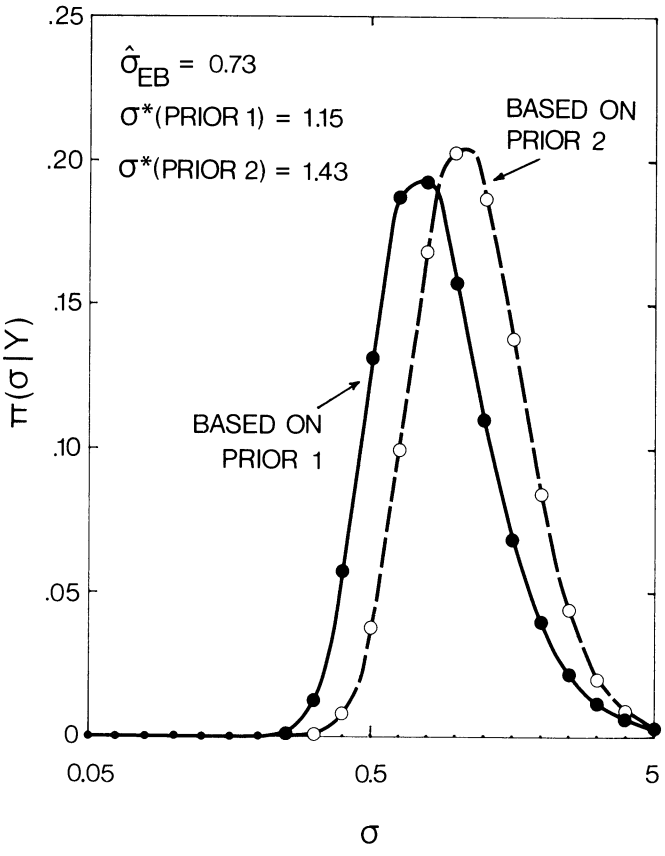


Figure 3. POSTERIOR DENSITIES OF log σ FOR THE 3 × 3 SUBMATRIX OF EXPERIMENTS IN FIGURE 1. The posterior densities are evaluated at points equally spaced (by .1 log 10) on the log scale.

terior densities of σ based solely on the upper left 3 × 3 submatrix of experiments, as displayed in Figure 1. With considerably fewer experiments, the posterior densities of σ now diverge substantially. The posterior standard deviations of the diesel A log potency are: 1.22 for the empirical Bayes estimate based on $\hat{\sigma}_{EB}$; 1.82 for the Bayes estimate based on Prior 1; and 2.24 for the Bayes estimate based on Prior 2. The corresponding 95 percent credible intervals for the diesel A dose-response slope are, respectively, [.139, 18.3], [.041, 59.3], and [.017, 132.].

With even fewer experiments, the precision of the human risk estimates depends almost entirely on prior beliefs. Suppose that we predict the lung cancer risk of diesel A emissions solely from the human data on coke oven emissions in combination with data on their relative potencies in mice. The maximum likelihood estimate $\hat{\sigma}_{MLE}$ and the empirical Bayes estimate $\hat{\sigma}_{EB}$ would be identically zero. For the empirical Bayes estimate, the 95 percent credible interval for the diesel A dose-response slope (e^{θ}) would be [.549, 2.23]. For Prior 1, however, it would be [.0008, 1451] and, for Prior 2, it would be [3.3 × 10^{-6}, 3.7 × 10^{5}].

In our statistical model, we assumed that the standard errors c_i of the observed log slopes were known. To check the sensitivity of our results to this assumption, we repeated our analysis on the data of Table 1, arbitrarily doubling the c_i for all the nonhuman experiments. For Prior 1, we obtained σ* = .99; the posterior mean and standard deviation of the human carcinogenic potency of diesel A were −.61 and 1.35, respectively. The other estimated human carcinogenic potencies were likewise only slightly different from those in Table 2.

5. INCORPORATING PRIOR INFORMATION

Figure 4 diagrams a 5 × 10 array of experiments. In the figure, the columns of Table 1 have been rearranged so that the diesel experiments *A* through *D* are on the right. In addition, we have appended the results of an epidemiological study of men exposed to a fifth type of diesel emission (Diesel E). The estimation for this additional dose-response slope, taken from Harris's (1981,1983) analysis of lung cancer incidence among diesel bus workers, is summarized in Appendix A.

Figure 4 exemplifies the case where there are only nonhuman data on some agents and only epidemiological data on another related agent. Unless we have some prior information on the relation among the agents, the inclusion of the epidemiological data cannot improve our estimates of human cancer risk. Thus, in the present example, if we continued to assume a diffuse prior distribution on the parameters β, our inclusion of the epidemiological study on diesel E would not affect the results in Table 2. The extra column effect corresponding to diesel E would merely be set so that the observed data on diesel E in humans exactly fit the underlying constant relative potency model.

Accordingly, we now consider the prior assumption that the emissions from the various diesel engines have very similar biological effects. In the double-subscript notation of (3.1), this biological similarity corresponds to a positive correlation among the five diesel column effects $\{\gamma_6, \ldots, \gamma_{10}\}$. More specifically, we model the prior relation between the diesel column effects by

$$\gamma_l = \gamma_0 + \eta_l, \quad l = 6, \ldots, 10 \qquad (5.1)$$

where γ_0 is a component common to all diesel effects, and where $\{\eta_6, \ldots, \eta_{10}\}$ represent the deviations of each

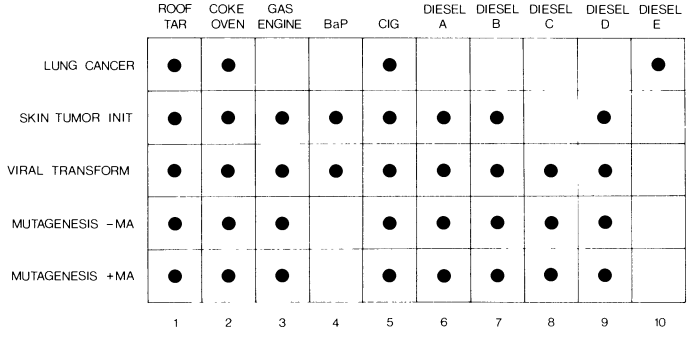
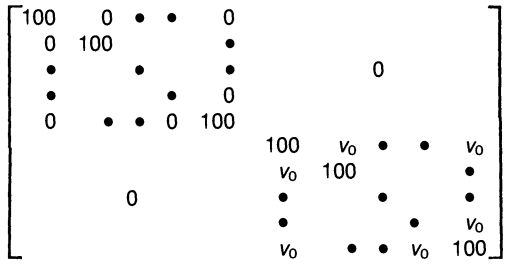


Figure 4. 5 × 10 DATA ARRAY. The columns of Table 1 have been permuted. An epidemiological study of diesel bus workers (Diesel E) has been appended. A filled circle means data available.

Table 3. Covariance Matrix V Under Informative Prior for Diesel Effects



diesel effect from the common mean. Each η_l, we assume, has prior mean zero, is independent of the other η_l and of γ_0, and has prior variance v_η. We denote the prior variance of γ_0 by v_0.

Since the covariance among the diesel column effects is our main interest, we specify an uninformative prior for the remaining parameters. Specifically, we set the prior covariances of all the parameters $\{\mu, \alpha_k, \gamma_l\}$ equal to 100. (Hence, $v_0 + v_\eta = 100$.) Moreover, except for the covariances v_0 among the diesel effects $\{\gamma_6, \ldots, \gamma_{10}\}$, we set the covariances among all the parameters $\{\mu, \alpha_k, \gamma_l\}$ equal to zero. The prior covariance matrix V thus takes the form given in Table 3, where the upper left 11 $\times$ 11 submatrix is the prior covariance of $\{\mu, \alpha_1, \ldots, \alpha_5, \gamma_1, \ldots, \gamma_5\}$ and the lower right 5 $\times$ 5 submatrix is the prior covariance of $\{\gamma_6, \ldots, \gamma_{10}\}$.

Thus, given σ, each θ_i has a prior variance $300 + \sigma^2$ or, equivalently, with probability 0.68, each dose-response slope could vary by a factor of $\exp\sqrt{300} \cong 3 \times 10^7$. (It is possible, as Smith 1973 shows in a simpler case, to let the variances $(v_0 + v_\eta)$ of the diesel effects go to infinity while v_η remains fixed.) Since the prior variances of the θ_i are so large, and since all the observed y_i are within ± 5, we can arbitrarily set $b = 0$ without affecting the conclusions below.

The prior relation between the diesel column effects depends critically upon v_η, for which we consider several possible values.

First, we consider $v_\eta = 100$, so that $v_0 = 0$ and $V = 100I$, where I is the 16 $\times$ 16 identity matrix. A priori, the diesel effects would be uncorrelated and diesel emissions would resemble each other no more than the other emissions.

Second, we consider $v_\eta = 0$, so that $v_0 = 100$ and V has rank 12. The diesel column effects would be a priori identical, and the number of distinct column effects would be reduced to 6. Any two diesel log potencies θ_i and $\theta_{i'}$ in the same species would have the same model predictions (that is, $x_i\beta = x_{i'}\beta$) and differ only in their model deviations (that is, δ_i and $\delta_{i'}$). This case resembles that analyzed by Yates and Cochran (1938) and Cochran (1980), in which experiments of very similar design differed primarily in their date and location. We might adopt this form of prior information when the various diesel emissions are thought to contain essentially the same biologically active components, but variations in engine type or operating conditions could produce idiosyncratic effects in some species.

Finally, we consider $v_\eta = 1$ (so that $v_0 = 99$) and $v_\eta = 2$ (so that $v_0 = 98$). In these cases, the diesel effects would be a priori correlated. Although the prior correlations are .99 and .98, respectively, our setting $v_\eta = 1$ or 2 in fact permits a substantial range of relative potencies. To see this, note that for a given σ, the difference between any two diesel log slopes in the same species has variance $2(v_\eta + \sigma^2)$. Provided that the prior means of the diesel column effects are equal, the relative potency of any two diesel emissions is a priori $\exp[z(2(v_\eta + \sigma^2))^{1/2}]$, where z is distributed $N(0, 1)$. Since $\exp(z\sqrt{2}) \cong 10$ for $z = 1.6$, our setting $v_\eta = 1$ already allows for more than a 10 percent probability that one of the diesels is 10 times more potent than the other. For $v_\eta = 2$, this probability increases to more than 25 percent. We might adopt these less extreme forms of prior information when the diesel emissions are thought to share many biologically active components, but the overall composition of the mixture varies with engine type and operating conditions.

Table 4 shows the resulting estimates of σ and of the θ's for roofing tar, diesel A, diesel E, gasoline engine emissions, and benzo(a)pyrene. Since the assumption that $v_\eta = 100$ (diesel effects uncorrelated) approximates complete ignorance about β, the results in that column are quite close to those obtained in the diffuse prior case in Table 2. As expected, the diesel E epidemiological study contributes only minimally to the estimation of σ and to the remaining θ's.

The assumption that $v_\eta = 0$ (identical diesel column effects) results in a marked increase in σ^*. Any variation in the diesel θ's that is in fact due to the γ's is forced to be fit to the corresponding δ's. Hence, the posterior variance of δ increases. Although the posterior standard de-

Table 4. Bayes Estimates of Log Slopes for Human Lung Cancer Risk under Alternative Assumptions on the Relation Between Diesel Column Effects

v_η	100	0	1	2	Original Data
σ^*	1.08	1.39	1.10	1.09	
Roofing Tar					
θ^*	.13	.53	.38	.32	.50
c^*	1.02	1.11	1.02	1.01	1.41
Diesel Engine A					
θ^*	−.44	−.98	−.28	−.25	
c^*	1.45	1.68	1.43	1.42	
Diesel Engine E					
θ^*	1.87	.31	.35	.64	1.92
c^*	1.50	1.16	1.17	1.23	1.51
Gasoline Engine					
θ^*	−2.01	−1.29	−1.59	−1.68	
c^*	1.44	1.79	1.46	1.44	
Benzo(a)pyrene					
θ^*	5.85	6.30	6.01	5.92	
c^*	1.55	1.94	1.57	1.55	

viation for diesel E is reduced, the precisions of the other estimates deteriorate.

The assumptions that $v_\eta = 1$ and $v_\eta = 2$ (correlated diesel effects) do not have these limitations. The estimate σ^* is not substantially affected. The posterior standard deviations of the log potencies of roofing tar, diesel A emissions, gasoline engine emissions, and benzo(a)pyrene are not much different from those in the first column of the table. But the precision of the estimate for diesel E is improved, and the potencies of diesel A and diesel E are drawn toward each other. As expected, the effect of going from $v_\eta = 0$ to $v_\eta = 1$ is much more dramatic than the effect of going from $v_\eta = 1$ to $v_\eta = 2$. Comparing the original data for diesel E with the estimate under $v_\eta = 2$, and using the dosage measure given in Appendix A, we find that the 95 percent credible interval for the relative risk of lung cancer for a diesel bus garage worker with 20 years' occupational exposure is narrowed from [1.007, 3.95] to [1.003, 1.31].

These illustrative results by no means exhaust the potential uses of prior biological information in our statistical framework. The inclusion of several groups of chemically similar environmental agents, for example, entails a straightforward generalization of (5.1). Such an analysis would permit us to examine whether the observed strong correlations between short-term bioassays and animal carcinogenesis studies result merely from the comparison of many chemically similar compounds within only a few chemical classes (Rinkus and Legator 1979).

We could also incorporate prior information on the relation between species or bioassays. Suppose, for example, that the experimental data matrix included several biologically related mutagenesis tests (Skopek et al. 1978). Prior correlations among the row effects $\{\alpha_k\}$ might be appropriate when, say, some strains test frameshift mutations and others test base-substitution mutations.

Although the exchangeable prior on δ could be criticized as an overly naive expression of prior ignorance, we stress that this assumption is the central starting point in any approach to interspecies comparisons. Nevertheless, as we noted in Section 3, nearly all of our results can be generalized to the case where δ is distributed as $N(0, \Sigma)$. Thus, suppose that one species is considered to be more heterogeneous than the others. For example, the subjects observed in the epidemiological studies were surely more genetically diverse than Sencar mice, Syrian hamster cells, or mouse lymphoma cells. Then we might specify $\delta_i \sim N(0, \tau^2)$ for all the human studies and $\delta_{i'} \sim N(0, \sigma^2)$ for the remaining experiments, where the δ_i and $\delta_{i'}$ are independent and τ is a priori likely to exceed σ.

Moreover, a simple correlation of row or column effects (that is, a prior on β) will not capture instances where two related species are thought to be especially sensitive to one class of compounds and especially insensitive to another. In those cases, the off-diagonal elements of Σ will be nonzero. As we explained in Section 3, the assumption that β is a priori uncorrelated with σ can also be relaxed. Such a generalization might be appropriate when the overall accuracy of interspecies extrapolation is thought to be related to the relative potency of two agents.

6. SELECTING THE MOST RELEVANT EXPERIMENTS

So far, we have assumed that the data of Table 1 are given in advance. Although the experiments were selected according to reasonable scientific criteria, it is not obvious that these data are most relevant for predicting human lung cancer risks. Even if we regard Table 1 as an exhaustive listing of available data, it is unclear whether a subset of experiments might not perform better. The question is: which of all possible subsets should we choose and upon what basis?

One possible starting point is to examine the residuals from the full model. Accordingly, Figure 5 depicts the posterior mean values of the residuals $\delta = \theta - X\beta$ derived from the Bayes estimates in Table 2. (We return to the case of a diffuse prior on β in order to keep the interpretation simple.) As Figure 5 shows, the posterior mean residuals for cigarette smoke are in four of five cases relatively large in absolute value. Further, the range of the mean residuals is largest for the direct mutagenesis experiments in the absence of metabolic activator. By

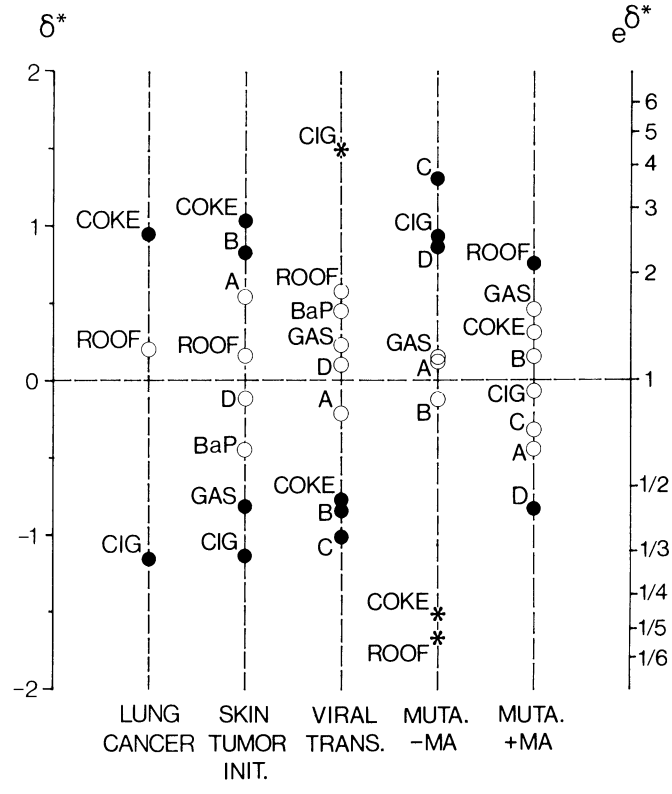


Figure 5. POSTERIOR MEAN RESIDUALS $\delta_i^* = \theta_i^* - x_i\beta^*$ DERIVED FROM BAYES ESTIMATES OF TABLE 1. Species-emission pairs with no original data, and therefore zero residuals, are omitted. An open circle means $|\delta_i^*| < \omega_i^*$; a closed circle means $\omega_i^* \le |\delta_i^*| < 2\omega_i^*$; and a star means $|\delta_i^*| \ge 2\omega_i^*$, where ω_i^* is the posterior standard deviation of δ_i.

contrast, the residuals for indirect mutagenesis with activator are more concentrated around the origin. A similar finding applies to those viral transformation residuals other than cigarette smoke.

Such an informal examination of residuals suggests that we should merely exclude those species-emission pairs with large values of $|\delta_i^*|$. But these individual interactions are what determine the relatedness of the various species and emissions. It would be more appropriate to assess whether a specific species or a specific agent is more or less relevant to the others.

Moreover, confining our attention solely to the model deviations δ ignores uncertainty in the model predictions $X\beta$. Consider a particular human log slope $\theta_i = x_i\beta + \delta_i$. The presence of less relevant experiments may accentuate the error δ_i, but it may also improve the posterior precision of $x_i\beta$. As we successively remove less relevant experiments, the variance of δ_i declines, but the variance of $x_i\beta$ increases and, at some point, the variance of θ_i also increases. In fact, when we have no initial data on θ_i, as in the case of Diesel A, the elimination of every conceivably irrelevant experiment leaves us with exactly what we had at the start—no information on θ_i at all.

We therefore consider the following procedure. Let Y_{k-} be the vector of log slopes formed by exclusion of all experiments involving species k, and define $\sigma^*_{k-} = E[\sigma^2 \mid Y_{k-}]^{1/2}$. Analogous definitions apply to Y_{-l} and σ^*_{-l} for each environmental agent l. Given an initial set of experiments Y, a prior density $\pi(\sigma)$, and a particular θ_i of interest, we first calculate σ^*_{k-} and σ^*_{-l} for each k and l, and determine the species k' or emission l' for which σ^*_{k-} or σ^*_{-l} is minimized. We then eliminate species k' or emission l' and iterate our procedure on the reduced set of experiments so long as (a) there exists a species k or emission l for which $\sigma^*_{k-} < \sigma^*$ or $\sigma^*_{-l} < \sigma^*$; (b) the species k' or emission l' does not correspond to θ_i; and (c) the elimination of species k' or emission l' reduces the posterior standard deviation of θ_i. If conditions (a), (b), and (c) are not satisfied, the procedure terminates.

We applied this procedure to the data of Table 1. We assumed Prior 1 for $\pi(\sigma)$ and focused on the human risks for roofing tar and diesel emission A. Table 5 shows the minimized values of σ^* and the corresponding values of θ_i^* and c_i^* for each iteration. Successive elimination of the row for mutagenesis without metabolic activator ($M-$), the column for cigarette smoke (CIG), and the row for skin tumor initiation in mice (SKIN) resulted in a reduction of σ^* from 1.08 to .48. Each of these steps also reduced c_i^* for both roofing tar and diesel A. In the final step, however, removal of the column for coke oven emissions (COKE) reduced σ^* to .395, but the posterior standard deviations of both roofing tar and diesel A increased. Hence, the procedure was terminated. As Table 5 indicates, the posterior mean values θ_i^* varied somewhat with successive iterations, but these variations were within $\pm c_i^*$.

Table 5. Results of Diagnostic Procedure[a]

	5 × 9	4 × 9	4 × 8	3 × 8	3 × 7	Original Data
Row/Column Eliminated		M−	CIG	SKIN	COKE	
σ^*	1.08	.93	.73	.48	.40	
χ^{*2}		13.5	11.5	15.8	2.5	
d.f.		7	3	6	2	
P		.06	.009	.01	.28	
P		.001	.003	.0004	.28	
Roofing Tar						
θ^*	.12	.31	.96	1.53	.50	.50
c^*	1.02	.96	.92	.74	1.41	1.41
Coke Oven						
θ^*	1.38	1.37	1.46	1.42		1.48
c^*	.33	.33	.33	.33		.34
Diesel Engine A						
θ^*	−.46	−.71	.21	.33	−.84	
c^*	1.45	1.30	1.16	.87	1.58	

[a] The columns labeled 5 × 9, 4 × 9, 4 × 8, 3 × 8, and 3 × 7 refer to successive iterations of the diagnostic procedure

NOTE M− denotes mutagenesis experiments in L5178Y mouse lymphoma cells without metabolic activator CIG denotes experiments on cigarette smoke SKIN denotes skin tumor initiation experiments in Sencar mice COKE denotes experiments on coke oven emissions See text for definitions of χ^{*2}, P, and p

The elimination of a particular subset of experiments is equivalent to estimating an augmented model of the form

$$Y = (X\ X_0)\begin{pmatrix}\beta\\ \beta_0\end{pmatrix} + \delta + \epsilon \qquad (6.1)$$

where the columns in the appended matrix X_0 are indicator vectors corresponding to all but one of the experiments in the subset to be eliminated, and where a diffuse prior is assumed on the corresponding parameters β_0. The hypothesis that a particular subset of experiments is less relevant to the others amounts to the assertion that β_0 is far from the origin. Let β_0^* and V_0^* denote, respectively, the posterior mean and covariance matrix of β_0, and let $\chi^{*2} = \beta_0^{*\prime} V_0^{*-1} \beta_0^*$. Since the posterior distribution of β_0 is approximately normal, the quantity $p = \exp -\frac{1}{2}\chi^{*2}$ is the approximate posterior density ratio comparing $\beta_0 = 0$ to $\beta_0 = \beta_0^*$. Moreover, $P = \text{Prob}\{\chi^2 \geq \chi^{*2}\}$ is the approximate posterior probability that β_0 is at least as far from β_0^* as β_0^* is from 0 (where distance is defined by the metric V_0^{*-1}, and where χ^2 is a chi-squared variable with degrees of freedom (df) equal to the number of omitted experiments minus 1). As a check on our results, we have therefore displayed these statistics for each iteration in Table 5. Note that the χ^2 tail probability P markedly increased when we eliminated the column for coke oven emissions.

The resulting tradeoff between predictive bias (σ^*) and predictive efficiency (c_i^*) is depicted graphically in Figure 6. Beyond the reduced 3 × 8 array, any further reduction in σ^* is at the cost of a marked loss of precision. To be sure, we cannot unequivocally conclude that the predictions resulting from the 3 × 8 matrix are most pre-

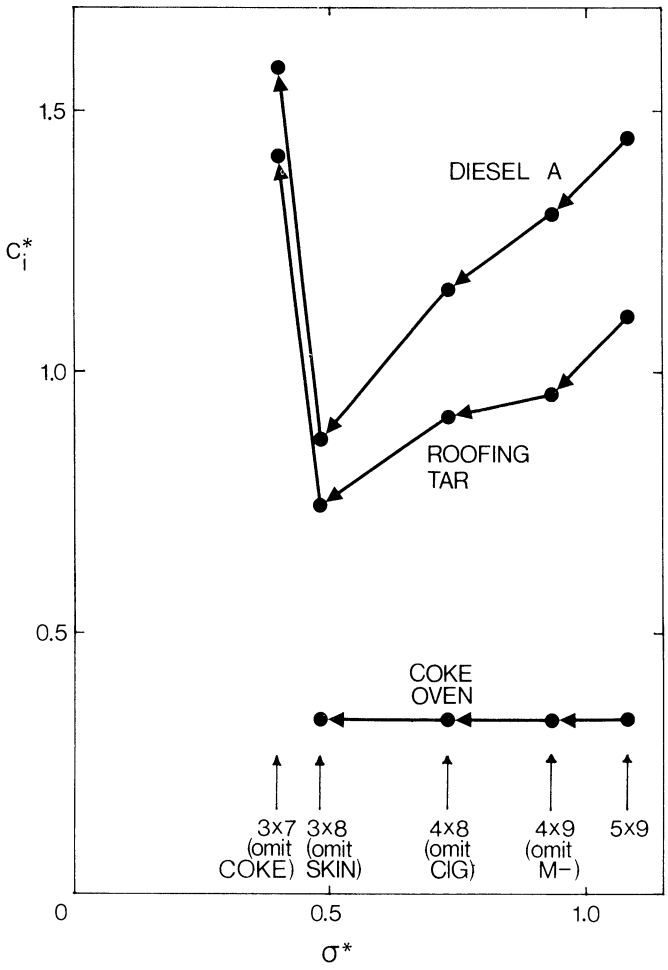


Figure 6. TRADEOFF BETWEEN σ AND c_i* FOR EACH ITERATION OF THE DIAGNOSTIC PROCEDURE. The excluded experiments and the resulting two-way tables at each iteration are given at the bottom of the figure.*

ferred. For many public health and environmental policy applications, however, a reduction in the extent of uncertainty about human risks is desired. To seek to eliminate less relevant experiments, so long as predictive efficiency is improved, appears to be appropriate for such applications.

Cigarette smoke, Figure 5 suggests, is a weaker human lung carcinogen, a weaker mouse skin tumor initiator, a more potent cell transforming agent, and a more potent direct mutagen than would be predicted from the underlying constant relative potency model. Peto (1977) has also noted that the mutagenic potency of cigarette smoke appears to be greater than would be expected from its systemic carcinogenicity. Such a finding may reflect chemical differences between cigarette smoke and the other combustion products in Table 1. (For example, cigarette smoke contains tobacco-specific nitrosamines and various heterocyclics.) Whatever the interpretation, our procedure leads us to eliminate a column containing the most precise human data. As a consequence, our estimate of σ is derived primarily from the nonhuman data. We see no completely satisfactory response to this limitation other than to suggest, where possible, the inclusion of other precise human data.

Our results necessarily depend upon the choice of prior density $\pi(\sigma)$. However, this dependence should be important only when there are few experiments. If we assume Prior 2 for σ, then the row for mutagenesis without activator, the column for cigarette smoke, and the row for skin tumor initiation are likewise successively eliminated, and σ* is reduced from 1.12 to .73. Thereafter, however, no further reduction in σ* is possible. Prior 2 is so conservative that any fewer experiments diminish our confidence in interspecies extrapolation.

Our procedure represents only one possible strategy of stepwise search through the subsets of a given data set. In the current example, there is in fact no row or column in the 3 × 8 array whose elimination reduces both σ* and the c_i*. But there could be a more efficient way of arriving at the frontier of the (σ*, c_i*) tradeoff. Adapting the methods in Efron and Morris (1973), we could replace our assumption that $V(\delta_i) = \sigma^2$ (for all i) with the assumption that $V(\delta_i) = \tau^2$ for some subset of experiments and then employ a joint prior distribution for (σ, τ) to derive the posterior distribution of θ. Unless the suspicious subset can be identified a priori, the present method appears to be much simpler in practice.

Despite such qualifications, we find the results of our diagnostic procedure intriguing. Assays for direct mutagenicity and tumor initiation have been excluded as less relevant. The remaining laboratory bioassays are designed to gauge an agent's interference with gene replication and cell differentiation. For the polyaromatic-containing emissions remaining in the 3 × 8 table, these biological processes could be the most relevant to human carcinogenesis.

7. CONCLUDING COMMENTS

The covariance components model of this article distinguishes between the measurement error of each experiment and an error of imperfect relevance among experiments. For the illustrative data set of Table 1, σ*, the standard deviation of the error of relevance, was about 1.1 on a log scale. That is, with 95 percent credibility, extrapolations from nonhuman data to human cancer risks can be accurate at best to a factor of $\exp(2\sigma^*) \cong 9$ in either direction. From this result, we computed credible intervals for the human carcinogenic potencies of various environmental agents for which there were little or no direct human observations. We next applied the Bayesian statistical framework to incorporate prior information on the structural similarity of various compounds. Finally, we explored the problem of deciding which experiments are most relevant to human cancer risks. Mammalian cell transformation studies and indirect mutagenesis studies appeared to be most relevant for the class of polyaromatic hydrocarbon-containing emissions considered here.

A Bayesian approach makes explicit the genuine un-

certainties in the problem at hand. When the scientist has a large body of data, as illustrated by Table 1, these uncertainties are likely to be reduced substantially. But when a compound's human cancer risk is predicted solely from experiments with that compound in a single nonhuman species such as rats (in effect, a 2×1 table with an empty cell), then the resulting human risk estimates will depend entirely on one's prior beliefs about the conversion factor between rats and humans or about the relevance of rat to man.

Moreover, performing many different nonhuman laboratory studies on only one environmental agent does not necessarily reduce the uncertainty about that agent's human health risks. Such an exercise corresponds to a $K \times 1$ case, where our statistical specification (3.2) reduces to $y_k = \mu + \alpha_k + \delta_k + \epsilon_k$. In this situation, profitably combining the experiments requires strong prior assumptions about the interspecies conversion factors $\{\alpha_k\}$. In the absence of such sharp prior information, one needs experiments in many species and many agents.

The results of combining experiments unavoidably depend upon the dose-response models used for each experiment. For example, we were constrained by the available data to apply a relative risk model to the human studies, where dosage was measured as cumulative lifetime exposure. We would have liked to estimate an additive risk model that distinguishes between the duration and intensity of exposure (Doll and Peto 1978). But our illustrative relative risk models did fit the original data reasonably (Harris 1981,1983).

Nor do we attach any special limitation to our reliance on the slope of a linear dose-response relationship. There are alternative methods of summarizing the results of an experiment, such as TD_{50}, the dose at which 50 percent of subjects develop tumor (Meselson and Russell 1977; Ames and Hooper 1978). A multivariate generalization could summarize each experiment by a vector of numbers. In this way, one could incorporate nonlinearities of dose-response or possible synergistic actions with other agents.

The validity of individual dose-response models is closely tied to the idea of exchangeability. The misspecification of the dose-response model for a particular experiment may affect the accuracy of the model relating all the experiments. Accordingly, exchangeability means that it is unknown a priori which dose-response model is more likely to be subject to misspecification. Perhaps misspecification is more probable in epidemiological studies based upon occupational exposures to relatively high doses. But empirical studies to support this presumption are not available.

Since there is uncertainty whether any single experiment on agent l in species k is representative of all such experiments on the same agent in the same species, it is arguable that the variances c_{kl}^2 are understated. The question is whether to allocate this source of uncertainty to the within-experiment errors ϵ_{kl} or the between-experiment errors δ_{kl}. We prefer the latter. When multiple experiments on a given agent in a given species have been performed in different laboratories, informative priors like those of Section 5 would apply. In any case, when we arbitrarily doubled the c_{kl} for all nonhuman experiments in Table 1, the resulting estimates of human carcinogenic potency were not substantially changed.

Is the normal prior structure of our model particularly objectionable? Deviations from the underlying constant relative potency model may certainly arise from biological processes that are non-Gaussian. But since the normal distribution has smaller tails than other likely candidates, any outliers from the underlying normal model will have a stronger contribution to the overall estimate of σ. In this respect, the use of the normal model is more conservative. On the other hand, the model may be misleading about the posterior probabilities of extreme values of the log carcinogenic potencies θ, such as $\Pr\{\theta > X\beta + 3\sigma\}$.

The statistic σ^{*2} corresponds to the mean squared error of prediction (on the log scale) for an experiment yet to be performed. It could thus be interpreted as the Bayesian risk of interspecies extrapolation under the loss function $(\theta_i - x_i\beta^*)^2$, where θ_i is estimated by $x_i\beta^*$. Although this loss function is symmetric with respect to proportional errors of prediction, it attaches greater weight to absolute underprediction of cancer risks. We regard this characteristic of σ^* as appropriately conservative.

The present statistical model resolves the difficulty, encountered by Crouch and Wilson (1979,1981), of having to perform separate comparisons of carcinogenic potency in different pairs of species. It also satisfies these authors' desire for a systematic method of identifying potential exceptions to the underlying extrapolative model. Since the use of informative priors permits us to include multiple experiments on the same agent in the same species, we avoid the problem, encountered by these authors, of deciding which of several experiments to incorporate in the analysis. The use of informative priors could also incorporate information about the faulty design or execution of an experiment. Furthermore, a multivariate generalization of our model could incorporate the incidences of tumors of different sites. This would avoid the additional difficulty, encountered by Crouch and Wilson, of deciding which of several endpoints to choose.

The Bayesian paradigm provides a unified statistical theory for previous studies of interspecies extrapolation, in particular, comparisons of carcinogenic potency in mammals with genotoxic potency in short-term tests (Ames and Hooper 1978; Clive 1977; McCann et al. 1975,1981; Meselson and Russell 1977; Poirier and de Serres 1979; Purchase et al., 1978; Rinkus and Legator 1979). Instead of individual two-way comparisons of, say, a particular tumorigenesis model with a particular strain of Salmonella, it is possible to analyze multiple tests of carcinogenicity and genotoxicity simultaneously. The use of informative priors can take account of known differences in route of exposure or organ sensitivity in carcinogenesis studies and source of metabolic activator or tar-

get genetic locus in mutagenesis studies. A variant of the diagnostic procedure in Section 6 might be valuable in determining which genotoxic test is most relevant to animal carcinogenicity.

Finally, it is very often infeasible to conduct a precise epidemiological study of the human cancer risks of a particular agent. In such cases, the Bayesian model can help to decide which of many other unperformed experiments might be most informative about the agent's carcinogenicity.

APPENDIX A: DOSE-RESPONSE MODELS

The slopes in Table 1 were estimated from linear dose-response models. In the nonhuman experiments, we observed n_j, the number of positive responses (skin papillomas, transformed cells, mutant colonies), and N_j, the number of surviving experimental sites (surviving mice, surviving SHE cells in culture, surviving L5178Y cells in culture), at each of several doses d_j. For each experiment, we assumed that the n_j were independently Poisson distributed with mean values $N_j(\zeta + \xi d_j)$, where ζ and ξ are unknown parameters.

In the epidemiological studies, we assumed that the relative risk of lung cancer was a linear function of cumulative lifetime exposure. For both the roofers study and the diesel bus workers study introduced in Section 5, we observed n_{ij}, the number of lung cancer cases, E_{ij}, the number of expected cases based on direct age standardization, and d_{ij}, the corresponding cumulative lifetime dose for subjects observed during calendar period i in exposure group j. (The exposure groups reflected varying durations of union membership in the roofers study and diverse job categories in the bus workers study.) For these two studies, we assumed that, conditional upon $n_{i\cdot} = \sum_j n_{ij}$, the n_{ij} were multinomially distributed with mean values $\zeta_i E_{ij}(1 + \xi d_{ij})$, where the ζ_i are constants of proportionality and ξ is an unknown parameter. For both the coke oven worker and cigarette smoking studies, we observed n_{ij}, the number of lung cancer cases, N_{ij}, the number of person-years at risk, and d_{ij}, the corresponding cumulative lifetime dose for subjects in age group i and exposure category j. (The exposure categories reflected various job groups in the coke oven worker study and differences in smoking habits in the cigarette smoking study.) For these two studies, we assumed that, conditional upon $n_{i\cdot} = \sum_j n_{ij}$, the n_{ij} were multinomially distributed with mean values $\zeta_i N_{ij}(1 + \xi d_{ij})$ where, again, the ζ_i are constants of proportionality and ξ is an unknown parameter.

The data for our analysis are based on the maximum likelihood estimates and standard errors of the dose-response slopes ξ. The estimation algorithms are given in DuMouchel (1981). Except for the column corresponding to cigarette smoke, the results are reported in Harris (1981,1983).

In our analysis of these experiments, the information about the log slope $i(\theta) = \partial^2 L(\theta)/\partial\theta^2$ (where $L(\theta)$ is the log likelihood for the experiment) always varied less, as a function of θ, than did the information about the slope, namely $i(\theta)e^{-2\theta}$. We therefore found it more plausible that the estimates of θ, rather than of e^{θ}, have known variances. The parameterization in terms of log slopes excludes negative carcinogenic potencies. Unless an agent is suspected to be anti-carcinogenic, we do not regard this restriction as especially serious.

If θ denotes the log carcinogenic potency of a particular emission in humans, and if d is the cumulative lifetime exposure to extractable organics (or whole condensate in the case of cigarettes), then the relative risk of lung cancer is $1 + de^{\theta}$. In units of $10^4\ \mu g/m^3 \times$ years, typical values of the cumulative dose d are: .30 for a coke oven worker aged 45–54 years; 1.1 for a roofer with 20 years' occupational exposure; .021 for a diesel bus garage worker with 20 years' exposure (diesel E introduced in Section 6); and about 210 for a smoker of one pack per day for 30 years. For an increase in the market share of light-duty diesel vehicles (such as diesel A) to 25 percent over a 20-year period, an urban resident would typically receive an ambient cumulative dose d equal to about 5×10^{-4}. These dosage values are included solely to facilitate interpretation of our results. They are not intended as precise estimates of exposure to be used in computing excess cancer rates in selected populations.

APPENDIX B: PROOF OF PROPOSITION 2

Lemma. Let U be an $n \times m$ matrix of rank $m < n$, I be the $n \times n$ identity martix, and t be a scalar. Then as $t \to \infty$,

$$(I + tUU')^{-1} = I - U(U'U)^{-1}U' + t^{-1}(UU')^{+} + O(t^{-2}), \tag{B.1}$$

$$|I + tUU'| = t^m |U'U| [1 + t^{-1}\mathrm{tr}(UU')^{+} + O(t^{-2})], \tag{B.2}$$

where A^{+} is the Moore-Penrose pseudoinverse of A.

Proof. The $n \times n$ matrix UU', which has rank m, can be represented as

$$UU' = \sum_{j=1}^{m} \lambda_j u_j u_j',$$

where $\{\lambda_j\}$ are the strictly positive characteristic roots of UU' and $\{u_j\}$ are the corresponding characteristic vectors. The $n \times n$ identity matrix can be represented as

$$I = \sum_{j=1}^{m} u_j u_j' + \sum_{j=m+1}^{n} v_j v_j',$$

where the unit vectors $\{v_j\}$ are all orthogonal to the characteristic vectors $\{u_j\}$. Combining these two expressions, we have

$$I + tUU' = \sum_{j=1}^{m} (1 + t\lambda_j) u_j u_j' + \sum_{j=m+1}^{n} v_j v_j'. \tag{B.3}$$

Hence

$$(I + tUU')^{-1} = \sum_{j=1}^{m} (1 + t\lambda_j)^{-1} u_j u_j' + \sum_{j=m+1}^{n} v_j v_j'.$$

As $t \to \infty$,

$$\sum_{j=1}^{m} (1 + t\lambda_j)^{-1} u_j u_j' = t^{-1} \sum_{j=1}^{m} \lambda_j^{-1} u_j u_j' + O(t^{-2}).$$

Equation (B.1) now follows from our recognition that $\sum_{j=m+1}^{n} v_j v_j'$ is the orthogonal projection operator which maps R^n onto the subspace of R^n orthogonal to the columns of U (namely $I - U(U'U)^{-1}U'$), while $\sum_{j=1}^{m} \lambda_j^{-1} u_j u_j'$ defines the pseudoinverse. Similarly, from (B.3),

$$| I + tUU' | = \prod_{j=1}^{m} (1 + t\lambda_j)$$

$$= t^m \prod_{j=1}^{m} \lambda_j \left\{ 1 + t^{-1} \sum_{j=1}^{m} \lambda_j^{-1} + O(t^{-2}) \right\}.$$

Equation (B.2) follows from our recognition that $| U'U | = \prod_{j=1}^{m} \lambda_j$ and $\mathrm{tr}(UU')^{+} = \sum_{j=1}^{m} \lambda_j^{-1}$.

The proof of Proposition 2 requires successive applications of this lemma. To show that $(Y - Xb)'[C + \sigma^2 I + XVX']^{-1}(Y - Xb)$ in (3.4) reduces to $Y'SY$ in (3.7), we use expansion formula (B.1), setting $U = W^{1/2}XV^{1/2}$ and $S = W^{1/2}[I - U(U'U)^{-1}U']W^{1/2}$. (The quadratic form $Y'SY$ in (3.7) is the sum of squared residuals of the weighted least squares regression of Y on the columns of X, where the weights are the diagonal elements of W.) To show that the determinant $| C + \sigma^2 I + XVX' |^{-1/2}$ in (3.4) becomes proportional to $| W |^{1/2} | X'WX |^{-1/2}$ in (3.7), we use expansion formula (B.2) under the same definition of U. Expression (3.8) follows from our setting $V^{-1} = 0$ in (3.5). That expression (3.6) reduces to (3.9) is a result of formula (B.1), where we set $U = XV^{1/2}$ in order to evaluate the terms $(XVX' + \sigma^2 I)^{-1}$ in (3.6).

[*Received February 1981. Revised June 1982.*]

REFERENCES

AMES, B.N., and HOOPER, K. (1978), "Does Carcinogenic Potency Correlate with Mutagenic Potency in the Ames Assay?" *Nature*, 274, 19–20.

CASTO, B.C., HATCH, G.C., HSUNG, S.L., HUISINGH, J.L., NESNOW, S., and WATERS, M.S. (1980), "Mutagenic and Carcinogenic Potency of Extracts of Diesel and Related Environmental Emissions: In Vitro Mutagenesis and Oncogenic Transformation," *Health Effects of Diesel Engine Emissions: Proceedings of an International Symposium* (2 vols.) eds. W.E. Pepelco, R.M. Danner, and N.A. Clarke, Cincinnati: U.S. Environmental Protection Agency.

CLIVE, D. (1977), "A Linear Relationship Between Tumorigenic Potency *In Vivo* and Mutagenic Potency at the Heterozygous Thymidine Kinase ($TK^{+/-}$) Locus of L5178Y Mouse Lymphoma Cells Coupled with Mammalian Metabolism," *Progress in Genetic Toxicology*, eds. D. Scott, B.A. Bridges, and F.H. Sobels, Amsterdam: Elsevier/North-Holland.

COCHRAN, W.G. (1980), "Summarizing the Results of a Series of Experiments," *Proceedings of the 25th Conference on Design of Experiments in Army Research, Development and Testing*, ARD Report 80-2, Washington, D.C.: U.S. Department of Defense, pp. 21–34.

CROUCH, E., and WILSON, R. (1979), "Interspecies Comparison of Carcinogenic Potency," *Journal of Toxicology and Environmental Health*, 5, 1095–1118.

——— (1981), "The Regulation of Carcinogens," *Risk Analysis*, 1: 47–57.

DE FINETTI, B. (1937), "La Prévision: Ses Lois Logiques, Ses Sources Subjectives," *Annales de l'Institut Henri Poincaré*, 7, 1–68. Reprinted in English in *Studies in Subjective Probability* (1964), eds. H.E. Kyburg, Jr. and H.E. Smokler, New York: John Wiley.

DEMPSTER, A.P., RUBIN, D.B., and TSUTAKAWA, R.K. (1981), "Estimation in Covariance Components Models," *Journal of the American Statistical Association*, 76: 341–353.

DOLL, R., and PETO, R. (1978), "Cigarette Smoking and Bronchial Carcinoma: Dose and Time Relationships Among Regular Smokers and Lifelong Nonsmokers," *Journal of Epidemiology and Community Health*, 32, 302–313.

DUMOUCHEL, W.H. (1981), "Documentation for CATDATA," Cambridge, Massachusetts: Massachusetts Institute of Technology, Statistics Center Technical Report.

EFRON, B. and MORRIS, C. (1973), "Combining Possibly Related Estimation Problems" (with Discussion), *Journal of the Royal Statistical Society*, Ser. B, 35, 379–421.

FREIREICH, E.J., GEHAN, E.A., RALL, D.P., SCHMIDT, L.H., and SKIPPER, H.E. (1966), "Quantitative Comparison of Toxicity of Anti-Cancer Agents in Mouse, Rat, Hamster, Dog, Monkey, and Man," *Cancer Chemotherapy Reports*, 50, 219–245.

HAMMOND, E.C., SELIKOFF, I.J., LAWTHER, P.L., and SEIDMAN, H. (1976), "Inhalation of Benzpyrene and Cancer in Man," *Annals of the New York Academy of Sciences*, 271, 116–124.

HARRIS, J.E. (1981), *Potential Risk of Lung Cancer from Diesel Engine Emissions*, Washington, D.C.: National Academy Press.

——— (1983), "Diesel Emissions and Lung Cancer" (with Discussion)," *Risk Analysis*, 3, 1, in press.

HARVILLE, D.A. (1974), "Bayesian Inference for Variance Components Using Only Error Contrasts," *Biometrika*, 61, 383–385.

——— (1977), "Maximum Likelihood Approaches to Variance Components Estimation and to Related Problems" (with Discussion), *Journal of the American Statistical Association*, 72, 320–340.

HUISINGH, J.L., BRADOW, R.L., JUNGERS, R.H, HARRIS, B.D., ZWEIDINGER, R.B., BUSHING, K.M., GILL, B.E., and ALBERT, R.E. (1980), "Mutagenic and Carcinogenic Potency of Extracts of Diesel and Related Environmental Emissions: Study Design, Sample Generation, Collection, and Preparation," *Health Effects of Diesel Engine Emissions: Proceedings of an International Symposium* (2 vols.), eds. W.E. Pepelco, R.M. Danner, and N.A. Clarke, Cincinnati: U.S. Environmental Protection Agency.

KAHN, H.A. (1966), "The Dorn Study of Smoking and Mortality Among U.S. Veterans: Report on Eight and One-Half Years of Observations," *Epidemiological Approaches to the Study of Cancer and other Chronic Diseases*, ed. W. Haenszel, National Cancer Institute Monograph No. 19, pp. 1–125.

LINDLEY, D.W., and SMITH, A.F.M. (1972), "Bayes Estimates for the Linear Model" (with Discussion), *Journal of the Royal Statistical Society*, Ser. B, 34, 1–41.

LLOYD, W.J. (1971), "Long-Term Mortality Study of Steelworkers v. Respiratory Cancer in Coke Plant Workers," *Journal of Occupational Medicine*, 13, 53–68.

MAZUMDAR, S., REDMOND, C., SOLLECITO, W., and SUSSMAN, N. (1975), "An Epidemiological Study of Exposure to Coal Tar Pitch Volatiles Among Coke Oven Workers," *Journal of the Air Pollution Control Association*, 25, 382–389.

MCCANN, J., CHOI, E., YAMASAKI, E., and AMES, B.N. (1975), "Detection of Carcinogens as Mutagens in the Salmonella/Microsome Test: Assay of 300 Chemicals," *Proceedings of the National Academy of Sciences, U.S.A.*, 72, 5135–5139.

MCCANN, J., HORN, L., LITTON, G., KALDOR, J., MAGAW, R., BERNSTEIN, L., and PIKE, M. (1981), "Short-Term Tests for Carcinogens and Mutagens: A Data-Base Designed for Comparative, Quantitative Analysis," *Structure Activity Correlation as a Predictive Tool in Toxicology: Fundamentals, Methods, and Applications*, ed. L. Goldberg, Washington, D.C.: Hemisphere Press.

MESELSON, M., and RUSSELL, K. (1977), "Comparisons of Carcinogenic and Mutagenic Potency," *Origins of Human Cancer*, eds. H.H. Hiatt, J.D. Watson, and J.A. Winsten, Cold Spring Harbor: Cold Spring Harbor Laboratory.

MITCHELL, A.D., EVANS, E.L., JOTZ, M.M., RICCIO, E.S., MORTELMANS, K.E., and SIMMON, V.F. (1980), "Mutagenic and Carcinogenic Potency of Extracts of Diesel and Related Environ-

mental Emissions: In Vitro Mutagenesis and DNA Damage," *Health Effects of Diesel Engine Emissions: Proceedings of an International Symposium* (2 vols.) eds. W.E. Pepelco, R.M. Danner, and N.A. Clarke, Cincinnati: U.S. Environmental Protection Agency.
NESNOW, S., TRIPLETT, L.L., and SLAGA, T.J. (1981), "Tumorigenesis of Diesel Exhaust, Gasoline Exhaust, and Related Emission Extracts on Sencar Mouse Skin," *Short-Term Bioassays in the Analysis of Complex Environmental Mixtures II*, eds. M.D. Waters, S.S. Sandhu, J.L. Huisingh, L. Claxton, and S. Nesnow, New York: Plenum Press.
PATTERSON, H.D., and THOMPSON, R. (1971), "Recovery of Interblock Information When Block Sizes Are Unequal," *Biometrika*, 58, 545–554.
PETO, R. (1977), "Epidemiology, Multistage Models, and Short-Term Mutagenicity Tests," *Origins of Human Cancer*, eds. H.H. Hiatt, J.D. Watson, and J.A. Winsten, Cold Spring Harbor: Cold Spring Harbor Laboratory.
POIRIER, L.A., and DE SERRES, F.J. (1979), "Initial National Cancer Institute Studies on Mutagenesis as a Prescreen for Chemical Carcinogenesis: An Appraisal," *Journal of the National Cancer Institute*, 62, 919–926.
PURCHASE, I.F.H. (1980), "Inter-Species Comparisons of Carcinogenicity," *British Journal of Cancer*, 41, 454–468.
——— (1982), "An Appraisal of Predictive Tests for Mutagenicity," *Mutation Research*, in press.
PURCHASE, I.F.H., LONGSTAFF, E., ASHBY, J., STYLES, J.A., ANDERSON, D., LEFEVRE, P.A., and WESTWOOD, F.R. (1978), "An Evaluation of Six Short-Term Tests for Detecting Organic Chemical Carcinogens," *British Journal of Cancer*, 37, 873–903.
RINKUS, S.J., and LEGATOR, M.S. (1979), "Chemical Characterization of 465 Known or Suspected Carcinogens and Their Correlation with Mutagenic Activity in the Salmonella Typhimurium System," *Cancer Research*, 39, 3289–3318.
SKOPEK, T.R., LIBER, H.L., KADEN, D.A., and THILLY, W.G. (1978), "Relative Sensitivities of Forward and Reverse Mutation Assays in *Salmonella typhimurium*," *Proceedings of the National Academy of Sciences, U.S.A.*, 75, 4465–4469.
SMITH, A.F.M. (1973), "Bayes Estimates in One-Way and Two-Way Models," *Biometrika*, 60, 319–329.
U.S. ENVIRONMENTAL PROTECTION AGENCY (1979), "The Carcinogen Assessment Group's Final Risk Assessment on Coke Oven Exposure," Technical Report, Washington, D.C.
YATES, F., and COCHRAN, W.G. (1938), "The Analysis of Groups of Experiments," *Journal of Agricultural Science*, 28, 556–580, reprinted in F. Yates (1970), *Experimental Design: Selected Papers of Frank Yates*, London: Griffin.

Comment

Carcinogenic Risk Assessment

DANIEL KREWSKI*

Increasing public concern over the presence of potentially carcinogenic agents present in the environment has led to a tremendous increase in research on carcinogenesis and in efforts towards the development of new strategies for the regulation of carcinogens (Krewski and Brown 1981). In Canada, cancer currently accounts for about 19 and 22 percent of all deaths among males and females, respectively (HWC 1979). Working with United States statistics, Doll and Peto (1981) recently speculated that environmental contamination may account for only about two percent of the total cancer burden, while occupational exposure was thought to be responsible for a further two to eight percent. Other opinions, however, have been somewhat higher (Epstein 1979).

In pragmatic terms, the process of carcinogenic risk assessment may be described in terms of four distinct phases (Whyte and Burton 1980). Initially, *risk identification* involves demonstrating that a well-defined hazard exists. Once a carcinogenic agent has been clearly identified, *risk estimation* techniques may in theory be used to determine the actual magnitude of the risk posed to man. Next, *risk evaluation* procedures may be employed to assess the societal significance of this risk in light of any perceived benefits associated with the agent in question. Finally, suitable strategies for *risk management* may be invoked in order to control exposure to environmental carcinogens to the extent possible.

The four major methods for identifying chemical carcinogens are (a) examination of the physical and chemical properties of the test compound, (b) short-term in vitro and in vivo tests designed to detect certain critical events of the carcinogenic process, (c) long-term bioassay experiments using laboratory animals, and (d) epidemiological studies of human populations (OTA 1981). Risk identification based solely on an analysis of chemical structure and activity is unfortunately of only limited value since slight alterations in molecular configuration can have marked effects on the carcinogenic potential of a compound (Lawley 1976). Although over 100 rapid and relatively inexpensive short-term tests for specific mutagenic effects have been described in the literature (Hollstein and McCann 1979), the predictive value of such tests with respect to carcinogenicity in mammalian species remains to be more clearly established (Krewski et al. 1982). Currently, the most widely employed method of identifying chemical carcinogens is the long-term carcinogen bioassay in small rodents (Gart et al. 1979; IARC 1980a). Although epidemiological investigations can provide the most direct evidence of carcinogenic effects in man, studies of human populations are often of limited sensitivity due to low levels of human exposure and may be subject to unknown biases due to the presence of con-

* Daniel Krewski is Chief, Biostatistics and Computer Applications Division, Environmental Health Directorate, Health and Welfare Canada, Ottawa, Ontario, Canada K1A OL2.