BIOMETRICS 56, 915–921
September 2000

When Should Epidemiologic Regressions Use Random Coefficients?

Sander Greenland

Department of Epidemiology, UCLA School of Public Health,
Los Angeles, California 90095-1772, U.S.A.

SUMMARY. Regression models with random coefficients arise naturally in both frequentist and Bayesian approaches to estimation problems. They are becoming widely available in standard computer packages under the headings of generalized linear mixed models, hierarchical models, and multilevel models. I here argue that such models offer a more scientifically defensible framework for epidemiologic analysis than the fixed-effects models now prevalent in epidemiology. The argument invokes an antiparsimony principle attributed to L. J. Savage, which is that models should be rich enough to reflect the complexity of the relations under study. It also invokes the countervailing principle that you cannot estimate anything if you try to estimate everything (often used to justify parsimony). Regression with random coefficients offers a rational compromise between these principles as well as an alternative to analyses based on standard variable-selection algorithms and their attendant distortion of uncertainty assessments. These points are illustrated with an analysis of data on diet, nutrition, and breast cancer.

KEY WORDS: Bayesian statistics; Causal inference; Empirical Bayes estimators; Epidemiologic methods; Hierarchical regression; Mixed models; Multilevel modeling; Random-coefficient regression; Relative risk; Risk assessment; Shrinkage; Variance components.

1. Introduction

When should epidemiologic regressions use random coefficients? I will argue that they are advisable whenever the analysis objective is estimation of multiple causal effects and some sort of dimensionality-reduction strategy is needed. My arguments are not of mathematical or simulation form because there are many technical studies that support my thesis (cf., the citations in Greenland (1998, p. 428–430)); I will instead focus on the scientific advantages of mixed modeling that those studies reflect. I have derived these arguments from writings of Box (1976), Leamer (1978), Good (1983), and other pragmatic Bayesians or Bayesians with reservations and compromises (e.g., Rubin, 1984; Draper, 1995), though any oversights are my own. What follows is an attempt to apply these ideas in epidemiology, an often controversial and idiosyncratic field whose importance is recognized but whose use of statistics remains largely primitive; implementation details can be found in textbooks under the topic of hierarchical modeling (e.g., Gelman et al., 1995, Section 13.4; Leonard and Hsu, 1999, Section 6.3) though not at a level accessible to most epidemiologists.

Causal effects are usually underidentified by epidemiologic data in that any realistic model for the effects cannot be fit without constraints. This underidentification is concealed by routine analysis strategies but can be addressed openly using models with random coefficients. The issue is important to society at large because of the seriousness with which the public and lay press often respond to epidemiologic studies (Taubes, 1995). For example, massive lawsuits often result from weak suggestions of hazards while dietary fads get launched by even weaker data. I attribute some of this problem to inappropriate modeling strategies that are common in epidemiology today. The example below is intended to show how these strategies lead to illusory significant results. I have encountered others in which this occurs, and I believe many reported findings (including several cited in Taubes (1995)) contain similar modeling artifacts.

I will not contrast fitting methods, which have been the focus of much research. That work, though important, has far outpaced work on connecting models to the scientific context (Hodges, 1996; Mallows, 1998). Nor will I address issues of model-form uncertainty or pure (noncausal) prediction modeling, as considered, e.g., in the literature on model averaging (e.g., Draper, 1995; Raftery, 1996; Buckland, Burnham, and Augustin, 1997), although mixed modeling can be viewed as a model-averaging method (Greenland, 1998, 1999).

2. Complete Confounding in a Study of Food Constituents and Breast Cancer

The example is from a case–control study of diet, food constituents, and breast cancer (Witte et al., 1994); controls are sisters of cases and so the data comprise matched sets with one to five sister controls. The variables include intakes of 35 food constituents (nutrients and suspected carcinogens) computed from 87 diet questionnaire items plus five potential confounders. This study is typical of many: The number of subjects (140 cases, 222 controls) is not much larger than the number of variables. (For further study details, see Ursin et al. (1992).) I will assume for now that only the food constituents are of interest. This still leaves a dimensionality problem, as one should expect with 35 primary plus 5 confounding covariates and only 140 cases (3.5 cases per covariate) available for analysis.

Standard analyses employ conditional logistic modeling with one of the following strategies:

(1) Use all 35 food constituents as candidate variables for some sort of data-based variable-selection procedure, such as stepwise regression, forcing in the five confounders (sometimes the confounders are also subject to selection based on significance testing, but this practice has been condemned for leaving important confounders uncontrolled (Greenland and Neutra, 1980)).

(2) Force all 35 food constituents and the 5 potential confounders into a single model and (if it fits) base inference on this model.

Strategy 1 can be condemned on the grounds that (i) the food constituents are strongly correlated and hence estimates from reduced subsets may be confounded by excluded variables, even if the latter are nonsignificant, and (ii) data-based variable selection leads to nonnormal estimators and to severe downward bias in the P-values and standard errors that come from the final model (e.g., see the studies cited in Buckland et al. (1997) and Greenland (1998, p. 402)). Bootstrapping the selection procedure is occasionally used to address problem (ii), but this approach has its own problems (Freedman, Navidi, and Peters, 1988). Strategy 2 has also been promoted to avoid the shortcomings of strategy 1 but depends on asymptotics whose applicability is dubious given the case/covariate ratio (exact logistic programs exist, but such large problems remain beyond their reach). Here, however, I will focus on a major problem for causal inference that is overlooked by all these strategies, i.e., confounding by residual dietary effects.

To describe this problem, let X represent the 362 × 87 dietary data matrix, let W be the 362 × 5 confounder data matrix, and let $Z = \{z_{jk}\}$ be the 87 × 35 composition matrix for the diet items; element z_{jk} is the amount of constituent k found in one unit of diet item j. Thus, Z is the table of contents for the diet items and XZ is the 362 × 35 matrix giving the constituent intakes for the subjects. Letting Y be the vector of subject-specific disease indicators, the logistic model underlying the above strategies may be written

$$\ell \equiv \text{logit}\{\text{E}(Y \mid X, Z, W)\} = \alpha + XZ\pi + W\theta, \qquad (1)$$

where π is the target parameter vector of constituent coefficients and α is a vector of nuisance parameters that are constant within matched sets. Strategy 2 uses model (1) in its entirety, whereas strategy 1 uses the data to select columns of XZ for use in a reduced model. The models are fit by conditional maximum-likelihood to eliminate α, and effects are measured by the vector of odds ratios e^{π} (Breslow and Day, 1980).

The first column of Table 1 presents selected results from applying strategy 1 to the food constituents using backward deletion with α-to-remove = 0.10; 15 of the 35 constituents are retained, and 11 of these have $P < 0.05$. The second column presents conditional maximum-likelihood (CML) estimates of odds ratios from strategy 2 (fit the full model); only 2 of the 35 coefficients have $P < 0.05$. The first four food constituents are shown because they have received considerable publicity as potential factors in carcinogenesis (possibly protective for Ω3 fatty acids, β-carotene, and phytoestrogens and possibly causal for alcohol). The differences in the point estimates from strategies 1 and 2 are trivial relative to the confidence-interval widths, but the intervals from the full model are meaningfully wider for Ω3 fatty acids and for alcohol. The differences in widths are unsurprising given the downward bias in standard errors estimated from data-selected models. The latter consideration should be enough to make one prefer the full-model intervals over the backward-deletion intervals. I will argue, however, that even the full-model intervals are misleadingly narrow.

Use of model (1) implicitly assumes absence of any effects of the diet variables X beyond the logit-linear effects mediated through the constituents in Z. There is no scientific basis for this assumption, and there are good reasons to reject it.

Table 1

Estimates of odds ratios e^{π} *from conditional logistic regressions of breast cancer on food constituents (95% confidence limits in parentheses); five potential confounders forced into each model*

	Model			
			With random diet residuals	
	Backward deletion[a]	CML, all 35 constituents	$\tau^2 = 1/8$	$\tau^2 = 1/2$
Ω3 fatty acids	0.77	0.71	0.58	0.49
(g/day)	(0.65, 0.92)	(0.46, 1.1)	(0.17, 2.0)	(0.06, 4.3)
β-carotene	1.1	1.2	1.1	1.2
(mg/day)	(0.99, 1.2)	(1.01, 1.3)	(0.81, 1.6)	(0.64, 2.1)
Phytoestrogens	0.80	0.73	0.73	0.72
(mg/day)	(0.70, 0.92)	(0.58, 0.93)	(0.40, 1.3)	(0.26, 1.9)
Alcohol	0.94	0.89	0.93	0.91
(3 oz./day)	(0.88, 1.00)	(0.63, 1.3)	(0.37, 2.3)	(0.18, 4.6)
Carbohydrate	1	0.97	0.99	1.0
(100 g/day)	(deleted)	(0.79, 1.2)	(0.58, 1.7)	(0.39, 2.6)

[a] α-to-remove = 0.10; 15 food constituents retained.

Dietary factors that may influence health continue to be discovered, and their effects are not captured by π. While the individual effects of single omitted factors are likely to be small, so are the effects under study. Furthermore, the aggregate confounding due to the omitted effects may be important because of the high positive correlations among healthy dietary habits.

To account for this confounding problem, consider the expanded model

$$\ell = \alpha + XZ\pi + X\delta + W\theta. \qquad (2)$$

The term $X\delta$ is intended to capture the residual diet-item effects. Because XZ is a linear function of X, however, the constituent and diet effects are completely confounded in that model (2) is not identified without side constraints. This nonidentification reflects the following fact: To control for other dietary effects using a fixed-effects-only model, one would have to measure the constituents responsible for those effects and add them to model (1); without such measurements, the effects of the measured constituents Z are not logically separable from other dietary effects because those constituents are measured only through diet variables in X. Standard analyses of nutrient effects dodge this logical problem by not looking beyond model (1). Of the two models, however, model (2) is the only scientifically reasonable one for effect estimation. Use of model (1) corresponds to imposing the implausible constraint $\delta = 0$ on model (2), which leads to understatement of uncertainty about e^{π}.

Underidentified structures like model (2) are common in epidemiology. Other examples include occupational studies in which X contains job histories and Z is a matrix of exposure levels within jobs, exercise studies in which X contains physical-activity histories and Z is a vector of metabolic expenditures of activities, and other studies in which X contains questionnaire items and Z is a matrix that transforms the items into quantities of focal interest. Most often, the potential effects of X items not captured by XZ are ignored; occasionally, items from X may be tested and added in a forward-selection strategy, although the number that can be added in this way is severely limited by the linear dependence of XZ on X.

3. A Mixed-Modeling Approach

3.1 *A Family of Estimators*

By treating δ as a vector of random coefficients, we can achieve identification using less restrictive and more plausible constraints than setting components of δ to 0. Perhaps the simplest way to do so is to treat model (2) as a mixed model by specifying $\delta \sim \text{MVN}(\mu, T)$, where μ and T are known or are simple functions of a few unknown parameters. I will here use $\mu = 0$, $T = \tau^2 I$; a more realistic prior would have the diagonal elements of T vary with diet item (indeed, Witte et al. (1994) constructed a more complex prior for δ based on extensive review of the background nutrition and epidemiology literature). The fact that the components of e^{δ} represent residual odds ratios after regressing out food-constituent effects makes the zero-correlation (diagonal T) assumption reasonable, because prior correlations among the diet-item effects are, for the most part, due to shared constituents. The normality of the prior is chiefly for computational ease and could be replaced by other assumptions if one had skill with software for Monte Carlo fitting. Assuming normality, however, leads to simple fitting methods such as restricted generalized least squares (Goldstein, 1995), restricted maximum likelihood (Wolfinger and O'Connell, 1993), penalized likelihood with a quadratic penalty for δ (Breslow and Clayton, 1993; Greenland, 1997), data augmentation (Bedrick, Christensen, and Johnson, 1996), and ridge regression with ridge parameters for δ proportional to $1/\tau^2$ (Titterington, 1985).

Discussions of penalized likelihood and ridge regression often treat $1/\tau^2$ as a tuning or smoothing parameter for solving an ill-conditioned regression problem rather than as an inverse variance component, and thus may appear to finesse the problem of specifying a coefficient distribution. Nonetheless, from a Bayesian perspective, such a distribution is implicit in these methods (Leamer, 1978) and the tuning parameter should reflect the precision of background information. I will thus use the prior information available in the example to assign plausible values to the prior variance of the residual effects in δ.

Let $\tilde{\pi}(\tau^2)$ denote the penalized conditional likelihood (PCL) estimator of π obtained from fitting model (2) with $\mu = 0$ and the prior variance fixed at τ^2. The third column of Table 1 gives results using $\tilde{\pi}(1/8)$, i.e., with $\tau^2 = \{\ln(2)/1.96\}^2 \doteq 1/8$. The latter number is derived from the context by noting that odds ratios below 1/2 or above 2 are extremely implausible because the components of e^{δ} are odds ratios for the residual effects for typical intakes of the dietary items in X after regressing out effects mediated by measured constituents. Taking $\tau^2 = 1/8$ corresponds to assigning 95% prior probability to the odds-ratio interval $\exp(0 \pm 1.96/8^{1/2}) = (1/2, 2)$ for each component of e^{δ}. The resulting point estimates differ little from those in the earlier columns, but the PCL intervals are considerably wider. Unlike the results from strategies 1 and 2, no mixed-model estimate has $P < 0.05$, and the precision of certain results in the first two columns apparently hinges on ignoring residual diet effects. Thus, mixed modeling indicates that there is little information in the data about effects of individual food constituents once we allow for the possibility of even small residual diet effects. As an added benefit, mixed modeling provides intervals for coefficients excluded by backward deletion.

The similarity of the point estimates in this example is not coincidental. A large change in point estimates upon variable deletion requires that the deleted variables have strong relations to both the outcome and the retained variables (cf., Breslow and Day, 1980, Chapter 2). Backward deletion with a high α-to-remove tends to delete only those variables with a weak relation to the outcome. Conversely, addition of random coefficients constrained by a small τ^2 tends to keep the added coefficients small. Hence, while large changes are possible, both the backward-deletion and the mixed-model point estimates tend to stay close to the full-model point estimates in this example. Nonetheless, the interval estimates differ profoundly, with the naive backward-deletion intervals shrinking as coefficients are removed and the mixed-model intervals growing as random coefficients are added, in accord with results on the impact of variable addition on logistic regression (Robinson and Jewell, 1991).

The mixed-model intervals are preferable for causal inference because model (2) better reflects current lack of knowledge about the diet residuals δ. The CML estimate $\hat{\pi}$ under strategy 2 equals $\tilde{\pi}(0)$, the mixed-model estimate obtained when δ is given a degenerate prior concentrated at zero. As uncertainty about the size of these residuals increases, so does uncertainty about π. This relation is illustrated by comparing the third column of Table 1 to the fourth column, which gives results using the contextually large value of $\tau^2 = \{\ln(4)/1.96\}^2 \doteq 1/2$; this τ^2 corresponds to assigning 95% prior probability to $\exp(0 \pm 1.96/2^{1/2}) = (1/4, 4)$ for each component of e^{δ}. The variances of the components of $\tilde{\pi}(\tau^2)$ increase without bound as $\tau^2 \to \infty$, reflecting the linear dependence of the constituents XZ on the diet items X.

As with δ, there is considerable prior information about π and θ in this example. Bayesian philosophy says one should use this information to add priors for π and θ to the analysis, while frequentist theory tells us that the resulting estimators may be superior to any above if that information is valid. Whether or not one finds these arguments compelling, they lack one crucial element in the argument for introducing the prior for δ: Some constraint on δ is needed to get a sensible estimate of π within model (2) whereas a prior for π or θ is not.

3.2 *Should the Prior Variance Be Estimated?*

What about uncertainty about τ^2 (or, more generally, T)? Because τ^2 is a parameter of the prior for δ, uncertainty about τ^2 is uncertainty about the uncertainty about δ, i.e., it is uncertainty about which prior we should use for δ. From a subjective Bayesian perspective, this hyperuncertainty concerns a parameter τ^2 that indexes different opinions about δ, and neither τ^2 nor a distribution for τ^2 have any objective meaning with respect to δ. In other words, uncertainty about τ^2 is nothing more than uncertainty about prior opinion. With this view, estimation of τ^2 is a pointless exercise; instead, uncertainty about τ^2 should be addressed by repeating the analysis using different values, as in the last two columns of Table 1. Those results suggest that, within the $\delta \sim \text{MVN}(0, \tau^2 I)$ prior specification, the main qualitative inference (no estimate appears incompatible with chance) should not vary among opinions with $\tau^2 > 1/8$.

Consider next a frequentist perspective in which one goal is to minimize expected loss in estimating π subject to the mixed-model specification. We don't know what value of τ^2 will minimize the expected loss of $\tilde{\pi}(\tau^2)$, so we might attempt to estimate it from the data. Because τ^2 controls the degree of shrinkage in $\tilde{\delta}(\tau^2)$, this approach accommodates intuitions that the data should have some say in how much to shrink δ. Unfortunately, common estimators for τ^2 can have very poor small-sample properties (Greenland, 1993, 1997); furthermore, the estimates they produce often equal no one's prior variance for δ, in which case the resulting odds-ratio estimates have no contextually relevant Bayesian interpretation. For this reason, if one feels compelled to estimate τ^2, I would recommend giving it a proper prior concentrated among contextually reasonable values.

3.3 *Mixed Coefficients*

So far, I have assumed that the analysis goal is to estimate effects of the composite covariates XZ, treating any residual effects of X as a source of bias. Suppose instead the goal is to estimate the effects of the basic covariates in X. A standard analysis would select columns of X for use in a logistic regression (strategy 1) or use all columns of X (strategy 2), as in the model

$$\ell = \alpha + X\beta + W\theta. \qquad (3)$$

In the example, this is a model for effects of the 87 diet items. Although β is identified without further specification, results from standard analyses are not credible: Upon fitting the full model, 29 components of the CML estimate $\hat{\beta}$ have $P < 0.05$ and many are absurdly inflated (Witte et al., 1994); after backward deletion with $\alpha = 0.10$, there is much less inflation, but 14 of the 20 retained components still have $P < 0.05$. For example, the full-model estimate of the odds ratio for eating two oranges per week is 3.1 (95% confidence limits: 1.2, 8.4); after backward deletion, the estimate becomes 1.6 (1.2, 2.2).

Much more plausible results can be obtained by exploiting the information in Z about food composition to shrink the CML estimate of β toward the value expected under model (1), in which foods have no effect beyond that conferred by their measured constituents. Model 2 with $\delta \sim \text{MVN}(0, T)$ is equivalent to a two-stage hierarchical (multilevel) model in which the first stage is model (3) and the second stage is

$$\beta = Z\pi + \delta. \qquad (4)$$

β is now a combination of fixed and random coefficients; an independence structure for the random part, δ, implies that any prior correlations among the diet effects in β are entirely explained by known differences in constituents of the diet items. This implication is a scientific proposition that was evaluated against background literature (Witte et al., 1994).

The mixed coefficient β can be estimated by plugging the mixed-model (model (2)) estimates $\tilde{\pi}(\tau^2)$ and $\tilde{\delta}(\tau^2)$ into equation (4). Since the estimated random vector $\tilde{\delta}(\tau^2)$ is shrunk toward the zero vector, $\tilde{\beta}(\tau^2) = Z\tilde{\pi}(\tau^2) + \tilde{\delta}(\tau^2)$ is an estimate of β that is shrunk toward $Z\tilde{\pi}(\tau^2)$, that portion of the estimated dietary effects due to the constituents Z. With $\tau^2 = 1/8$, the overall results appear much more ambiguous than those from CML or backward deletion; e.g., only 4 of the 87 components of $\tilde{\beta}(1/8)$ have $P < 0.05$, and the estimate of the odds ratio for eating two oranges per week is reduced to 1.4 (0.93, 2.0). The degree of shrinkage is controlled by τ^2: $\tilde{\beta}(0) = Z\hat{\pi}$, where $\hat{\pi}$ is the CML estimate of π under model (1), whereas $\tilde{\beta}(\tau^2)$ approaches $\hat{\beta}$ as τ^2 increases. (For further illustration of these points in the example, see Witte et al. (1994).)

Use of model (4) does not require prior information as detailed as a diet-nutrient matrix. If that matrix had been unavailable for our analyses, we would have used other, more crude information to construct a second stage (prior) design matrix Z. For example, we could group the coefficients by food type (vegetables, fruits, white meats, red meats, etc.); Z would then be the matrix of group indicators. As before, the objectives of the prior grouping would be to produce uncorrelated or exchangeable priors for residual effects not captured by the grouping and to minimize bias in any one coefficient as a result of shrinkage toward an inappropriate mean (Greenland, 1992). Because we would expect greater heterogeneity of effects within food-based than within constituent-based groups, however, we would have used a larger value of τ^2 (or a prior for τ^2 with a larger mean) with a food-based grouping.

4. Discussion

4.1 *Mixed Modeling as an Extension of Established Methods*

Epidemiologic regressions occasionally include random effects that are coefficients for a set of group or cluster membership indicators, where the groups are families, geographic areas, or sets of repeated observations on single individuals. The group-indicator coefficients are treated as i.i.d. or as having a specified correlation structure (e.g., exchangeable) because the groups are too numerous and small to allow stable estimation of their coefficients without constraints. In other words, the indicator coefficients are assumed random for the same reason as δ was assumed random in the above example. This assumption is easily accepted in group-indicator cases because (i) the group coefficients are usually regarded as nuisance parameters, which makes the assumption seem to be of only indirect importance, and (ii) the groups constitute a single natural partition of the observations, which makes the prior correlational assumptions seem natural in that the latter reflect symmetries in prior information about the group effects.

Mixed models extend standard random-effects models to include prior information about causal and measurement processes in a model for the source of effect correlations (Greenland, 1992; Searle, Casella, and McCulloch, 1992, p. 330). Such modeling can provide shrinkage estimators superior to the original ridge and James–Stein estimators, which only shrink toward the origin. Consider estimation of dietary effects (β in model (3)). Shrinking β toward the origin is equivalent to using model (4) with $\pi = 0$, an incorrect restriction. Mixed modeling allows shrinkage of β toward a manifold $Z\pi$ that is contextually determined, which increases coherence of the analysis with prior information. By dropping the incorrect restriction, we should also expect less bias (at a cost of greater variance) in mixed modeling than in classical shrinkage while retaining lower mean-squared error than unconstrained ML estimation.

4.2 *The Constraints of Unconstrained ML*

Standard epidemiologic analyses often begin and usually end with fixed-effects logistic regression fit by unconstrained maximum likelihood (ML). Unconstrained ML is often defended against shrinkage and Bayesian estimation with claims that it is unbiased and free from dependence on prior information. These claims are misleading because they are based on the assumption that the correct model is known and is the only model used in the analysis. In epidemiology, this assumption is always highly unrealistic, as in the example of estimating the constituent effects in π. Unconstrained ML forces use of an inadequately small fixed-effects model (such as model (1) or a backward-deletion model), whereas shrinkage allows use of a much richer mixed model (such as model (2) with random δ). In practice, then, ML tends to suffer from more bias due to model restrictions.

There is a sense in which this bias reflects an enhanced dependence of unconstrained ML on prior information. Every nonexperimental inference is a function of prior information and data. Although unconstrained ML uses no explicit prior, it does use a prior in the form of restrictions on the class of models available for the analysis (Leamer, 1978; Robins and Greenland, 1986). Mixed modeling expands that class and thus can reduce bias from incorrect model restrictions while facilitating use of plausible restrictions, as in the above example. Classical solutions to such problems involve sharp constraints, such as setting coefficients to zero or imposing absolute bounds, which do not reflect the vagueness of true prior information and which make valid uncertainty evaluation difficult. A smooth prior can be viewed as a probabilistic constraint requiring no sharp bounds. Mixed modeling represents a convenient means of imposing such fuzzy constraints.

4.3 *The Parsimony Problem and Model Selection in Causal Inference*

The above arguments for model expansion using random coefficients oppose the usual parsimony principle, which says to seek the simplest model for the job. When one attempts a causal analysis of complex and poorly understood relations from observations made without the benefit of randomization (as in most of epidemiology), models need to be complex to capture uncertainty about the relations. In other words, an honest uncertainty assessment requires parameters for all effects that we know may be present. This advice is implicit in an antiparsimony principle often attributed to L. J. Savage, "All models should be as big as an elephant" (see Draper, 1995). When we attempt to operationalize this advice with conventional regression tools, however, we run into another problem—you can't estimate anything well if you try to estimate everything simultaneously without constraints (illustrated by the fact that π in model (2) is not even identified without constraints). This problem drives analysts to search for simple models even if they do not explicitly adopt parsimony as a principle. Mixed models offer an alternative to purely data-driven model simplification and consequent uncertainty understatement.

Results will be sensitive to reasonable model choices whenever one can envision more important parameters than can be identified from the data. This problem is often handled with mechanical selection algorithms that ignore all context and produce models that exclude important parameters. The true sensitivity of causal inferences to model choice is concealed because these algorithms avoid the territory of underidentified models. To address this problem, some authors add a single unidentified parameter for unmeasured effects to a simple model and examine sensitivity of results to variations in this parameter (Rosenbaum, 1995; Copas and Li, 1997; Robins, Rotnitzsky, and Scharfstein, 1999), analogous to the use of τ^2 above. These methods are a welcome advance beyond the usual approach, but as implemented to date, they do not incorporate prior information as rich as that in the above example.

Many analysts recognize that no causal inference is possible from nonexperimental data without external identifying constraints. Placing distributions on coefficients provides more flexible and hence less unrealistic constraints than excluding them entirely. This flexibility can also be advantageous in pure prediction problems, for it allows one to move beyond the all-or-none approach of variable selection. Under a mean-zero, variance τ^2 specification for a coefficient,

$\tau^2 = \infty$ corresponds to its inclusion as a fixed effect and $\tau^2 = 0$ corresponds to its exclusion. By allowing τ^2 to be between these extremes, mixed models allow a smooth tradeoff between the excessive variability of estimators from complex models and the bias of estimators from oversimplified models.

Of course, there is always a limited number of parameters that can be estimated meaningfully in an analysis model; mixed modeling only raises this limit, and initial screening of candidate effects will still be needed in many problems. Here again, however, no effect should be excluded if such exclusion is contextually implausible, and Bayesian screening can outperform conventional stepwise procedures (Faraggi and Simon, 1997).

4.4 *The Multiple-Comparisons Issue*

The example might be viewed as a multiple-comparisons problem and tempt one to use classical adjustments (such as α-level reduction) to screen the results from a standard analysis. Prominent epidemiologists have condemned such adjustments as unscientific and have even denied there are multiple-comparisons problems (Rothman, 1990; Cole, 1993; Savitz and Olshan, 1995). Others recognize the problems but join in criticizing classical solutions (Greenland and Robins, 1991; Greenland, 1993; Thompson, 1998). Giving the target parameters random components (as in model (4)) treats the problem with a global loss function quite different from that in classical adjustment: Mixed modeling of the sort described here attempts to minimize estimation error by using additional background information, while classical methods only attempt to preserve global α-levels through purely arithmetic adjustments. It should be no surprise, then, that critics of the latter find mixed modeling more acceptable (Poole, 1991; Savitz and Olshan, 1995).

5. Concluding Remarks

The view of variable selection as an inadmissible form of shrinkage has been expounded for decades (e.g., Leamer 1978, Sections 5.2 and 5.3), yet naive modeling strategies based on pretest variable selection or other data-driven predictive approaches continue to dominate teaching, software, and practice. Meanwhile, strategies better suited for causal analysis (such as mixed modeling) go unmentioned in basic regression texts or are subsumed under specialty topics (like variance components) that focus on estimating τ^2, not β (although the empirical-Bayes and best-linear-unbiased-prediction (BLUP) literatures are noteworthy exceptions). As a consequence, mixed models are rarely used in fields like epidemiology that need them, even though there are many packages for generalized linear mixed modeling (Zhou, Perkins, and Hui, 1999).

I believe statisticians have a professional responsibility to distinguish causal from purely predictive modeling and to integrate methods suitable for causal modeling of non-experimental data into the basic teaching and practice of epidemiologic analysis. This integration will require that teachers and practitioners learn how to begin modeling with an underidentified model (a scientifically rich model that one would use if given enough information, such as models (2) or (4)) and then develop identifying constraints that are plausible. This strategy will require more attention to analysis context than is the current norm in statistics texts, but this requirement should be viewed as a benefit, not a burden.

Although mixed modeling is not the only scientifically sound approach to identification problems, it does have the advantage of using extensions of standard models in which the coefficients retain their familiar log-relative-risk interpretation. Mixed modeling can also encompass certain other flexible approaches, such as regression smoothing (through use of saturated splines with random coefficients). Nonetheless, like all methods, mixed modeling has limitations. One is its greater complexity and hence greater opportunity for misunderstanding relative to standard methods; others will no doubt become apparent with wider use. Such problems may, however, lead to improvements in the method and motivate implementation of other scientifically defensible approaches to underidentification.

Acknowledgements

This work originated as a presentation invited by the ASA CDC Chapter, Atlanta. The author is grateful for valuable criticisms received from Keith O'Rourke, the associate editor, and the referees.

Résumé

Les modèles des régressions à coefficients aléatoires apparaissent de façon naturelle à la fois dans les problèmes d'estimations fréquentistes et Bayesiens. Ils sont en voie de devenir disponibles dans les logiciels standards sous la dénomination de modèles linéaires généralisés mixtes, de modèles hiérarchiques, et de modèles multi-états. Je soutiens ici que de tels modèles offrent un cadre scientifiquement plus défendable pour l'analyse en épidémiologie, que les modèles à effets fixes, actuellement les plus utilisés. L'argument fait appel au principe d'anti-parcimonie, attribué à L. J. Savage: les modèles devraient être suffisamment riches pour refléter la complexité des relations à l'étude. Il fait aussi référence au principe de compensation selon lequel on ne peut rien estimer lorsqu'on cherche à tout estimer (souvent utilisé pour justifier la parcimonie). La régression à coefficients aléatoires offre un compromis rationnel entre ces principes, de même qu'une alternative aux analyses basées sur des algorithmes standards de sélection de variables, avec les distorsions des évaluations d'incertitude qui les accompagnent. Ces points sont illustrés par une analyse de données sur le régime, la nutrition et le cancer du sein.

References

Bedrick, E. J., Christensen, R., and Johnson, W. (1996). A new perspective on generalized linear models. *Journal of the American Statistical Association* **91,** 1450–1460.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association* **71,** 791–799.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88,** 9–25.

Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research,* Volume I, *The Analysis of Case-Control Studies.* Lyon: IARC.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics* **53,** 603–618.

Cole, P. (1993). The hypothesis generating machine. *Epidemiology* **4**, 271–273.

Copas, J. B. and Li, H. G. (1997). Inference for non-random samples (with discussion). *Journal of the Royal Statistical Society, Series B* **59**, 55–95.

Draper, D. (1995). Assessment and propagation of uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B* **57**, 45–97.

Faraggi, D., and Simon, R. (1997). Large sample Bayesian inference on the parameters of the proportional hazard model. *Statistics in Medicine* **16**, 2573–2585.

Freedman, D. A., Navidi, W., and Peters, S. C. (1988). On the impact of variable selection in fitting regression equations. In *On Model Uncertainty and Its Statistical Implications,* T. K. Dijlestra (ed), 1–16. Berlin: Springer-Verlag.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis.* New York: Chapman and Hall.

Goldstein, H. (1995). *Multilevel Statistical Models,* 2nd edition. London: Edward Arnold.

Good, I. J. (1983). *Good Thinking.* Minneapolis: University of Minnesota Press.

Greenland, S. (1992). A semi-Bayes approach to the analysis of correlated multiple associations, with an application to an occupational cancer mortality study. *Statistics in Medicine* **11**, 219–230.

Greenland, S. (1993). Methods for epidemiologic analyses of multiple exposures: A review and comparative study of maximum likelihood, preliminary-testing, and empirical-Bayes regression. *Statistics in Medicine* **12**, 717–736.

Greenland, S. (1997). Second-stage least squares versus penalized quasi-likelihood for fitting hierarchical models in epidemiologic analyses. *Statistics in Medicine* **16**, 515–526.

Greenland, S. (1998). Introduction to regression modeling. In *Modern Epidemiology,* K. J. Rothman and S. Greenland (eds), 401–432. Philadelphia: Lippincott-Raven.

Greenland, S. (1999). Multilevel modeling and model averaging. *Scandinavian Journal of Work Environment and Health* **25**(Suppl. 4), 43–48.

Greenland, S. and Neutra, R. R. (1980). The control of confounding in the assessment of medical technology. *International Journal of Epidemiology* **9**, 361–367.

Greenland, S. and Robins, J. M. (1991). Empirical-Bayes adjustments for multiple comparisons are sometimes useful. *Epidemiology* **1**, 43–46.

Hodges, J. S. (1996). Statistical practice as argumentation: A sketch of a theory of applied statistics. In *Modeling and Prediction: Honoring Seymour Geisser,* J. C. Lee, W. D. Johnson, and A. Zellner (eds), 19–45. New York: Springer.

Leamer, E. (1978). *Specification Searches*. New York: Wiley.

Leonard, T. and Hsu, J. S. J. (1999). *Bayesian Methods.* Cambridge: Cambridge University Press.

Mallows, C. (1998). The zeroth problem. *The American Statistician* **52**, 1–9.

Poole, C. (1991). Multiple comparisons? No problem! *Epidemiology* **2**, 241–243.

Raftery, A. E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalized linear models. *Biometrika* **83**, 251–266.

Robins, J. M. and Greenland, S. (1986). The role of model selection in causal inference from nonexperimental data. *American Journal of Epidemiology* **123**, 392–402.

Robins, J. M., Rotnitzsky, A., and Scharfstein, D. O. (1999). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Methods in Epidemiology,* M. E. Halloran and D. A. Barry (eds), 1–92. New York: Springer-Verlag.

Robinson, L. D. and Jewell, N. P. (1991). Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review* **58**, 227–240.

Rosenbaum, P. R. (1995). *Observational Studies.* New York: Springer-Verlag.

Rothman, K. J. (1990). No adjustments for multiple comparisons are needed. *Epidemiology* **1**, 43–46.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations. *Annals of Statistics* **12**, 1151–1172.

Savitz, D. A. and Olshan, A. F. (1995). Multiple comparisons and related issues in the interpretation of epidemiologic data. *American Journal of Epidemiology* **142**, 904–908.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components.* New York: Wiley.

Taubes, G. (1995). Epidemiology faces its limits. *Science* **269**, 164–169.

Thompson, J. R. (1998). Invited commentary. Re: Multiple comparisons and related issues in the interpretation of epidemiologic data (with discussion). *American Journal of Epidemiology* **147**, 801–815.

Titterington, D. M. (1985). Common structure of smoothing techniques in statistics. *International Statistical Review* **53**, 141–170.

Ursin, G., Aragaki, C. C., Paganini-Hill, A., Siemiatycki, J., Thompson, W. D., and Haile, R. W. (1992). Oral contraceptives and premenopausal bilateral breast cancer: A case–control study. *Epidemiology* **3**, 414–419.

Witte, J. S., Greenland, S., Hale, R. W., and Bird, C. L. (1994). Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. *Epidemiology* **5**, 612–621.

Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *Journal of Statistical Computing and Simulation* **48**, 223–243.

Zhou, X.-H., Perkins, A. J., and Hui, S. L. (1999). Comparisons of software packages for generalized linear multilevel models. *The American Statistician* **53**, 282–290.

Received June 1999. *Revised February* 2000.
Accepted February 2000.