Empirical Bayes Ranking Methods
Author(s): Nan M. Laird and Thomas A. Louis
Source: *Journal of Educational Statistics*, Vol. 14, No. 1 (Spring, 1989), pp. 29-46
Published by: American Educational Research Association and American Statistical Association
Stable URL: http://www.jstor.org/stable/1164724
Accessed: 03/02/2010 13:30

Journal of Educational Statistics
Spring 1989, Vol. 14, No. 1, pp. 29–46

Empirical Bayes Ranking Methods

Nan M. Laird
Harvard School of Public Health
and
Thomas A. Louis
University of Minnesota School of Public Health

Key words: *ranking and selection, empirical Bayes*

Ranking problems arise in setting priorities for investigations, in providing a simple summary of performance, in comparing objects in a manner robust to measurement scale, and in a wide variety of other applications. Commonly, rankings are computed from measurements that depend on the true attribute. Using the Gaussian model, we propose and compare methods for using these measurements to estimate the ranks of the underlying attributes and show that those based on an empirical Bayes model produce estimates that differ from ranking observed data. These differences result both from the effect of shrinking posterior means towards a common value by an amount that depends on the precision of individual measurements and from the Bayes processing of the posterior distribution to produce estimates that account for the uncertainty in the distribution of the ranks. We illustrate different ranking methods using data on school achievement reported by Aitkin and Longford (1986). Mathematical and empirical results highlight the importance of using appropriate ranking methods and identify issues requiring further research.

Ranking and selection are related statistical problems with numerous applications. Ranking methods can provide key information in prioritizing chemicals for carcinogenicity testing, environmental monitoring of areas suspected to have elevated cancer rates (Lagakos, Wesson, & Zelen, 1986), and investigation of medical service regions suspected to have elevated surgical rates (McPherson, Wenneberg, Hovind, & Clifford, 1982). Selection rules play an important role in many settings, including breeding research programs (Gianola & Fernando, 1986) and clinical decision making.

Supported by grants DMS 8702402 and INT 8512148 from the National Science Foundation. This work was completed while the second author was in the Department of Biostatistics at the Harvard School of Public Health. We thank Nick Longford for background data, Jeff Longmate and David Wypij for their comments, Julia Bailey for the graphics, and Rosa Esparza for preparing the manuscript.

A recent paper by Aitkin and Longford (1986) on the evaluation of school effectiveness explored the use of various methods for ranking British elementary schools based on pupil performance data.

The books by Bechhofer, Kiefer, and Sobel (1968) and Gibbons (1971), a series of papers coauthored by Gupta (see Gupta & Hsiao, 1983), and the work of Pettitt (1982) provide general background in this area, but few publications discuss Bayes and empirical Bayes ranking methods. These methods estimate ranks for random effects in a variance components model. Portnoy (1982) discusses one aspect of the ranking problem in a random effects model; our approach can be applied more generally. The "histogram estimates" of Louis (1984) and Tukey (1974) are relevant, but do not explicitly estimate ranks. Recently, a Bayesian approach to selection of the maximum in ANOVA has been given by Berger and Deeley (1988). In this report, we discuss some conceptual and technical issues for this problem and use the Aitkin and Longford (1986) data as an example.

Although the thrust of this manuscript is methodological, it is important to note that educators disagree about the suitability of ranking schools, even adjusting for pupils' backgrounds. As Aitkin and Longford (1986) note, a potentially more informative way of evaluating school effectiveness is to use regression methods, such as those described in Laird and Ware (1982), to explain differences in school achievements as a function of additional pupil and school variables. If additional variables can be found that explain all the between-school variation, there is no basis for ranking. Our methodology is easily generalized to incorporate additional pupil and school variables and to produce ranks adjusted for these factors. If these factors do explain all the between-schools variation, the estimated ranks will be degenerate at the mean rank. One way of evaluating the relative importance of school and pupil variables would be to examine the impact on the ranks of excluding them from the model.

Finally, we note that ranking can play a valuable role in drawing attention to unusually good or poor performance, thus providing a mechanism for setting priorities for case studies or detailed investigation. For example, low-ranking and high-ranking schools can be compared in an attempt to discover programmatic, staff, or student determinants of these extreme ranks. In other contexts, ranks can be used, for example, to set priorities for investigation of small areas with apparently elevated cancer rates and studies of potential carcinogens, and to scrutinize hospitals with elevated death rates. Ranks provide a summary that avoids misrepresentation of the precision of estimation and the validity of measurement systems.

The generic ranking problem can be described as follows: K units are to be ranked on the basis of a set of unknown parameters, θ_k, $k = 1, \ldots, K$. The θ_ks may be performance potential, genetic merit, and so forth. Data are available for each unit (possibly multivariate observations that may be stochastically dependent from unit to unit). A statistical model is used to specify the distribution of the data, conditional on the θ_k and possibly other unknown parameters, say, ψ. Of course, if the θs are known, so are the ranks, but generally the "best" ranks do not correspond to the ranked estimates of the θs.

Aitkin and Longford (1986) use two general approaches to ranking: (a) Treat the θ_ks as fixed parameters, estimate the θ_ks (and ψ) in some appropriate way, and rank the estimated θ_ks, (b) Treat the θ_ks as random effects, estimate each θ_k by its conditional expectation given the data, and rank the conditional expectations. The second approach can be viewed as standard Bayes or empirical Bayes (EB), where we compute a posterior for each θ_k, conditional on data and prior parameters, then rank the θ_ks on the basis of the posterior mean. If the prior parameters are unknown and estimates based on the data are used in calculating the posterior distribution (as in Aitkin and Longford), the approach is empirical Bayes.

In this paper, we extend the second approach by treating the ranks of the θ_ks (denoted R_ks) as the parameters of interest, and developing ranking methods based on the conditional distribution of the ranks rather than the conditional distribution of the θ_ks. Our methods move the rank of a unit with a relatively high posterior variance toward the average rank. Thus, our extension of the Aitkin and Longford (1986) analysis underscores the importance of incorporating posterior variability into the ranking procedure. In addition, we show the effect of inflating the posterior variance to account for uncertainty in estimating the Bayes model, and produce easily applied confidence procedures for estimated ranks.

Our approach to ranking has the advantage of reporting the posterior means and variances of the ranks. Using these, rather than integer ranks, can give a much clearer picture of what differences there are (if any) among the schools. This clarity is illustrated in the Aitkin and Longford (1986) example.

Most applications require modeling to adjust for covariates and thus produce estimated posterior distributions. Our methods focus on using these estimated posteriors, and the importance of including estimation uncertainty in them. The exact nature of the covariance adjustment influences our analyses only through its effect on estimated posterior distributions. Aitkin and Longford (1986) discuss relevant issues in the school evaluation context, and we do not consider them further.

This report focuses on concepts and data analysis; the appendices contain most technical developments. The second section deals with ranking via posterior distributions. The third section describes empirical Bayes methods for producing posterior distributions, the fourth section contains an example using the Aitkin and Longford (1986) school data, and the fifth section summarizes results and conclusions.

Ranking via Posterior Distributions

General Issues

To focus on the basic issues in using a Bayes approach to rank components, we assume for simplicity that modeling and data adjustment result in independent Gaussian posterior distributions for the θ_k s:

$$\theta_k|\text{data} \sim \mathrm{N}(\mu_k, \tau_k^2), \qquad k = 1, \ldots, K \tag{1}$$

and our objective is to rank the θs. The independence assumption will often be reasonable for standard Bayesian analysis and a simplifying assumption for empirical Bayes models. We let S_k denote the ranks of the μ_k s. Let

$$R_k = \text{rank}(\theta_k) = \sum_{j=1}^{K} \mathrm{I}(\theta_k \leq \theta_j) = \sum_{j=1}^{K} \mathrm{I}_{jk}, \tag{2}$$

with $\mathrm{I}(\cdot)$ the indicator function and $\mathrm{I}_{jk} = \mathrm{I}(\theta_k \leq \theta_j)$. We assume that ties have probability 0, so the largest theta has rank 1 and so on.

The (empirical) Bayesian methodology can be used to generate the joint posterior distribution of the ranks or, more simply, relevant summaries. Evaluation of multidimensional Gaussian integrals is required to produce the posterior and would be necessary to summarize the posterior by finding the modal ranks. As Pettitt (1982) shows, finding the modal ranks ($\tilde{R}_k$) requires delicate numerical integrations and extensive computations. His approximations provide an attractive approach. To avoid these computations, we compute the posterior expected ranks ($\hat{R}_k$) and the associated covariance matrix. Unlike the modal ranks, the posterior expected ranks can be affected by nonlinear monotone transforms, and they will not necessarily be integers. To obtain integers, we can rank the $\hat{R}_k$, producing R_k^*.

In some situations these rankings methods agree. For example, if the τ_k s are all equal to a common τ, then $\tilde{R}_k = R_k^* = S_k$ (the modal and ranked expected ranks equal the ranks of μ_k s; see Appendix A). Portnoy (1982) generalizes this result for ranking random effects in the variance component model. His counterexamples highlight the conceptual and technical difficulties for arbitrary τ_k. Other loss functions can be considered. For example, we conjecture that the R_k^* are Bayes with respect to squared error loss subject to the estimates' being a permutation of the K integers 1 to K.

We use the $\hat{R}_k$ and R_k^*, because they are easily computed, and the $\hat{R}_k$ permit us to produce standard confidence statements. They also clearly exhibit the effect of unequal posterior variances in producing inferences different from those based on ranked means. The Expected Ranks subsection and Appendix B derive relatively straightforward properties of the $\hat{R}_k$ that display these differences. Gibbons (1971, section 7.5) and Pettitt and Siskind (1981) present similar results.

Expected Ranks

Let

$$\hat{R}_k = \mathrm{E}(R_k) = \sum_{j=1}^{K} P_{jk},$$

where $P_{kk} = 1$ and $P_{jk} = \mathrm{P}(\theta_k \le \theta_j)$ for all $j \ne k$. Assuming independence of the θ_ks,

$$P_{jk} = \mathrm{P}(\theta_k - \theta_j \le 0) = \Phi[(\mu_k - \mu_j)/\sqrt{\tau_k^2 + \tau_j^2}], \tag{3}$$

with Φ the standard normal distribution function. Note that $\Sigma \hat{R}_k = K(K + 1)/2$. Formula 3 shows that as $\tau_k \to 0$, $\hat{R}_k \to 1 + \sum_{j \ne k} \Phi(\Delta_{kj})$, where $\Delta_{kj} = (\mu_k - \mu_j)/\tau_j$, and as $\tau_k \to \infty$, $\hat{R}_k \to (K + 1)/2$. We denote the R_k^* the ranks of the $\hat{R}_k$. If some of the $\hat{R}_k$ are tied, we produce the relevant R_k^* in the usual manner by equating each to the average of the relevant ranks.

In the case where the θ_ks are independent, the discrepancy between, for example, R_k^* and S_k depends on the differences in the τ_ks. As shown in Appendix A, if the τ_ks are equal (producing stochastically ordered posteriors), the ranks agree. If the τ_ks differ, then the R_k^* ranks for relatively high τ_ks are pulled toward $(K + 1)/2$ compared to the S_k.

Now consider the case where the θ_ks are not necessarily independent. If all the θs have a common variance (τ^2) and a common correlation (ρ) that increases from 0 to 1, the $\hat{R}_k$ (and thus the R_k^*) converge to the S_k. This result along with the previous shows that the degree of shrinkage toward $(K + 1)/2$ depends on both the variance (relatively higher variance for some components producing more shrinkage) and the correlation. To understand the role of the correlation coefficient, consider the representation for compound symmetry with a single latent random variable. Then, differences in θs have mean $(\mu_k - \mu_j)$ and variance $2\tau^2(1 - \rho)$, so as $\rho \to 1$ the differences are degenerate at the difference in means. Thus the $\hat{R}_k$ converge to S_k as $\rho \to 1$.

Our method of computing the $\hat{R}_k$ ignores any correlation in the posterior distributions of the θ_ks that may be induced by using estimated prior distributions (Laird & Louis, 1987; Rubin, 1982) or that may be present in a more general variance components model. The generalization of (3) to the model where the θs have a joint multivariate normal distribution is straightforward.

Producing Posterior Distributions

In the Bayes model, the prior distribution for the θ_ks and any nuisance parameters will be specified fully. The joint posterior for the θ_ks, integrating out nuisance parameters, can be produced in a straightforward way. Some formulae in the preceding section will need modification if the joint posterior is not the product of independent normals, but the same conceptual approach applies.

Lindley and Smith (1972) give a Bayes approach for regression models. When the parameters of the prior are unspecified, a mixed model provides the appropriate structure. This model is also referred to as Model II in the analysis of variance, and empirical Bayes on random effects more generally. Laird and Ware (1982) give the structure and show how to estimate the prior distribution and adjust the procedure for having estimated the prior mean (equivalently the fixed effects). An additional adjustment is necessary for having estimated the prior covariance matrix, as we discuss later in this section.

The particular example we treat is bivariate; each school has an individual slope and intercept relating initial performance to achievement. Like Aitkin and Longford (1986), we will take θ_k to be expected achievement for a given initial performance. Our approach is to produce the bivariate posterior distribution of intercept and slope for each school. Then, assuming multivariate normality, the posterior for θ_k follows directly from the bivariate posterior of slope and intercept.

Assume for the kth school and the jth student that

$$y_{kj} = a_k + b_k X_{kj} + e_{kj} \qquad \begin{aligned} k &= 1, \ldots, K \\ j &= 1, \ldots, n_k, \end{aligned} \tag{4}$$

where the e_{kj}s are independent $N(0, \sigma^2)$ random variables and the random effects (a_k, b_k) have a known Gaussian distribution:

$$\begin{pmatrix} a_k \\ b_k \end{pmatrix} \sim \mathrm{N}_2\left[\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \Sigma\right] \tag{5}$$

with

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{\mathrm{I}}^2 & \sigma_{\mathrm{SI}} \\ \sigma_{\mathrm{SI}} & \sigma_{\mathrm{S}}^2 \end{bmatrix}.$$

In the school effectiveness example, the response (y_{kj}) is a measure of student achievement and X_{kj} is Verbal Reasoning Quotient (VRQ). The random intercept model of Aitkin and Longford (1986) results from setting $\sigma_{\mathrm{S}}^2 = \sigma_{\mathrm{SI}} = 0$. Schools are to be ranked on the basis of predicted outcome at a given input X: $\theta_k = a_k + b_k X$. We suppress dependence on X.

First, assume that all the fixed effects and variance components $(\alpha, \beta, \sigma^2, \Sigma)$ are known. Then the posterior distribution of (a_k, b_k) given the data and the fixed effects is:

$$\begin{pmatrix} a_k \\ b_k \end{pmatrix} \Bigg| \, \mathrm{data} \sim \mathrm{N}_2\,[\boldsymbol{m}_k, \boldsymbol{V}_k]$$

$$\boldsymbol{m}_k = \boldsymbol{V}_k\left[\boldsymbol{D}_k^{-1}\begin{pmatrix} \hat{a}_k \\ \hat{b}_k \end{pmatrix} + \boldsymbol{\Sigma}^{-1}\begin{pmatrix} \alpha \\ \beta \end{pmatrix}\right]$$

$$V_k = (D_k^{-1} + \Sigma^{-1})^{-1}, \tag{6}$$

where $(\hat{a}_k, \hat{b}_k)$ are the ordinary least squares (LS) estimates for the kth school, $D_k = \sigma^2(X_k^t X_k)^{-1}$, and X_k is the kth school's design matrix. Formula 6 shows that the posterior mean vector is closer to the prior mean vector than $(\hat{a}_k, \hat{b}_k)$ is, in that the posterior mean lies inside the circle of radius $[(\hat{a}_k - \alpha)^2 + (\hat{b}_k - \beta)^2]^{1/2}$ around (α, β). This mapping can produce nonintuitive results. For example, the posterior expected intercept need not be between $\hat{a}_k$ and α.

Because the y_{kj}s are independent given (a_k, b_k) and the (a_k, b_k)'s are independent, the θ_ks are a posteriori independent with:

$$\theta_k \,|\text{data} \sim \mathrm{N}(\mu_k, \tau_k^2),$$

where

$$\mu_k = m_{k1} + m_{k2}X$$
$$\tau_k^2 = (1 \;\; X)V_k\begin{pmatrix}1\\X\end{pmatrix}. \tag{7}$$

When the fixed effects are unknown, Aitkin and Longford (1986) suggest using (7) for the posterior mean and variance of θ_k, replacing the unknown fixed effects and variance components by suitable (e.g., marginal maximum likelihood: MML) estimates. As pointed out by Laird and Ware (1982), Morris (1983), and Laird and Louis (1987), the resulting posterior variance will be too small, because it neglects any uncertainty about the estimated prior. To account for estimating the prior mean, Laird and Ware modify the posterior variance expression (6) by introducing a flat prior for (α, β) and producing the marginal posterior of the (a_k, b_k)'s conditional on the data, σ^2, and Σ. The resulting distribution for (a_k, b_k) is still bivariate normal with mean equal to $\hat{m}_k$ ($\hat{\alpha}$ and $\hat{\beta}$ being replaced by their generalized least squares estimates) and variance given by

$$V_k^* = \mathrm{var}\left(\begin{matrix}a_k\\b_k\end{matrix}\middle|\,\text{data}, \sigma^2, \Sigma\right) = V_k + V_k\Sigma^{-1}Q\Sigma^{-1}V_k \tag{8}$$

where

$$Q = \left[\Sigma^{-1}\sum_{k=1}^{K} V_k D_k^{-1}\right]^{-1}.$$

They replace σ^2, Σ with the restricted maximum likelihood estimates in computing (8). Harville (1976) and Tukey (1974) give sampling theory justifications for the modified variance formula.

The Laird and Ware approach needs to be extended to account for estimation of the variance components and to incorporate posterior correlation among schools. It is straightforward to incorporate the correlation resulting from estimating (α, β), but very complicated in general to adjust the posteriors to account for uncertainty about σ^2 and Σ. The use of a multivariate analogue of Morris (1983) or the bootstrap (Laird & Louis, 1987) would be required. These details are not central to our main point and our data analysis is primarily based on the "naive" approach. However, for completeness we include some details of this adjustment in Appendix C.

School Effectiveness: Results

In this section we use the school effectiveness data given in Aitkin and Longford (1986) to rank the 18 secondary schools on the basis of predicted performance at VRQ = 85 and 110. As noted in Aitkin and Longford, schools 17 and 18 are single-sex schools, and the rest are mixed-sex comprehensive schools. The single-sex schools have considerably higher intake scores. For purposes of illustration, we retain all 18 schools for ranking and

TABLE 1
Predicted response

	VRQ = 85				VRQ = 110			
	Least squares[a]		Posterior[b]		Least squares[a]		Posterior[c]	
School number	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	10.3	2.1	10.5	1.8	34.6	1.5	34.0	1.3
2	12.5	1.3	12.3	1.2	32.8	2.0	32.4	1.6
3	6.9	3.0	8.7	2.2	34.3	1.6	33.8	1.4
4	6.5	2.4	8.2	1.9	28.7	1.7	29.7	1.5
5	10.8	1.4	10.8	1.3	34.1	2.3	33.2	1.8
6	8.5	1.6	9.1	1.5	29.0	3.6	30.9	2.2
7	9.9	1.5	10.4	1.4	26.1	2.3	28.5	1.8
8	10.3	1.9	10.6	1.6	29.6	1.6	30.1	1.4
9	11.0	1.6	11.1	1.4	30.8	2.4	31.2	1.8
10	14.9	2.1	13.7	1.8	36.2	1.9	34.7	1.6
11	12.3	2.0	12.2	1.7	26.8	1.8	28.2	1.5
12	4.8	2.2	6.7	1.8	38.8	2.1	36.4	1.7
13	10.0	1.6	10.4	1.5	29.5	2.2	30.4	1.7
14	10.5	2.8	10.7	2.1	32.0	2.5	31.9	1.9
15	10.3	1.3	10.5	1.3	29.3	1.8	30.1	1.5
16	12.3	1.7	12.1	1.5	27.6	1.8	28.8	1.5
17	36.1	12.6	9.6	2.9	47.3	4.2	37.5	1.4
18	20.6	10.1	12.8	2.8	32.1	4.1	29.9	1.5

[a] Least squares depart from those in Table 6 of Aitkin and Longford (1986) because of rounding errors.
[b] Prior mean = 12.1; prior *SD* = 3.3.
[c] Prior mean = 31.8; prior *SD* = 2.8.

restrict attention to models that include both random slope and intercept, because this policy produces the greatest spread in the resulting posterior means and variances. If only the 16 comprehensives are ranked, most methods of ranking based on random effects models give nearly identical results, because the posterior variances are so similar.

Similarly, if we assume a single slope for all schools ($\sigma_S = \sigma_{IS} = 0$), there is less spread in posterior variances, making the rankings based on any empirical Bayes model very similar. For this reason, our rankings should not be compared with Aitkin and Longford (1986) ranks, because the only rankings they present for the variable-slope case are based on least squares (column 2 of our Tables 2 and 3). They argue instead for deleting school 12, whose slope is distinctly different from the others. Deleting it makes the resulting $\hat{\sigma}_S^2$ not significantly different from zero so that ranks can be based on a single-slope model. Our point here is not to draw different substantive conclusions, but rather to illustrate the performance of ranking methods in a more complex setting.

For the first analysis, we use the approach of Aitkin & Longford (1986) to produce the $\hat{\mu}_k$ and $\hat{\tau}_k$s (from formulae 6 and 7, replacing all fixed effects by their MML estimates provided by Aitkin and Longford). Table 1 gives the individual least squares estimates ($\hat{\theta}_k = \hat{a}_k + \hat{b}_k X$), using slopes from Aitkin and Longford's Table 5 and intercepts calculated from their Table 1. Discrepancies between our Table 1 and their Table 6 are due to rounding errors. The posterior means and standard deviations shown in our Table 1 are calculated using formula 6 and prior values supplied to us by Aitkin and Longford. As described in Appendix C we also calculated adjusted posterior variances that account for having estimated $(\alpha, \beta, \sigma^2, \Sigma)$ (not shown). The adjustment had very little effect on the posterior variances, except for schools 17 and 18, where increases in standard deviations ranged from 10 to 150% (the .8 for school 17 in Table 3 changes to 2.4). This illustrates that these increases can be very dramatic for points far from the center of the distribution.

Tables 2 and 3 give ranks based on least squares estimates, posterior means, and posterior expected ranks. In evaluating formulas 3 and 4 we replace μ_k and τ_k with their estimates and use 9-point Gaussian quadrature.

Discussion

Apparent anomalies occur in Table 1, as can be seen by looking at schools 17 and 18. For example, for VRQ = 85 school 17's LS mean of 36.1 overshoots the prior mean of 12.1, when moving to the empirical Bayes posterior mean of 9.6. For VRQ = 110, school 18's LS mean (32.1) overshoots the prior mean (31.8) when moving to the empirical Bayes posterior mean of 29.7. These behaviors could not occur in univariate Bayes or empirical Bayes models. In multivariate models, however, though the pos-

TABLE 2
Ranking schools for VRQ = 85

School number	Ranks based on: LS mean response	Ranks based on: Posterior mean response (S_k)	Ranks based on: Posterior expected ranks (R_k^*)	Posterior ranks (R_k): Mean[a]	Posterior ranks (R_k): SD
17	1	14	14	11.3	5.4
18	2	2	3	5.4	4.9
10	3	1	1	3.2	2.8
2	4	3	2	5.1	2.8
16	5	5	4	5.8	3.5
11	6	4	5	5.9	3.8
9	7	6	6	8.3	3.7
5	8	7	7	9.0	3.6
14	9	8	8	9.2	4.8
8	10	9	9	9.7	4.1
15	11	10	10	9.9	3.7
1	12	11	11	9.9	4.4
13	13	13	13	10.2	3.9
7	14	12	12	10.2	3.7
6	15	15	15	13.2	3.3
3	16	16	16	13.4	4.2
4	17	17	17	14.5	3.4
12	18	18	18	16.7	2.0

[a] Ties in the posterior means are due to rounding. The mean rank is 9.5.

terior mean vector must be closer in Euclidean distance to the prior mean vector than is the observed mean vector, this relation need not hold for linear combinations.

Ranking posterior means produces a big change from ranks based on unadjusted least squares means (Tables 2 and 3). Note especially the change in rank for VRQ = 85 for school 17: Its least squares rank is 1, but its posterior mean rank is 11.3. The large standard deviation of its least squares estimate (attributable to extrapolation to a VRQ value outside the range of its pupil's values) produces the change. As with the usual EB adjustments, the posterior expected ranks ($\hat{R}_k$) are closer to $(K + 1)/2$ [9.5] than are the ranked posterior means (the S_k), with the discrepancy depending on the posterior variance. So, the ranked $\hat{R}_k$ (the R_k^*) can differ from the S_k. Note the differences for schools 2, 18, 11, and 16 in Table 2. The adjusted posterior variances are more heterogeneous than the unadjusted, producing a change in the rankings for schools 4, 12, 17, and 18

TABLE 3
Ranking schools for VRQ = 110

School number	Ranks based on: LS mean response	Ranks based on: Posterior mean response (S_k)	Ranks based on: Posterior expected ranks (R_k^*)	Posterior ranks (R_k): Mean[a]	Posterior ranks (R_k): SD
17	1	1	1	1.5	0.8
12	2	2	2	2.4	1.5
10	3	3	3	4.2	2.1
1	4	4	4	5.0	2.0
3	5	5	5	5.4	2.2
5	6	6	6	6.5	3.0
2	7	7	7	7.8	3.0
18	8	14	14	12.9	3.0
14	9	8	8	8.8	3.5
9	10	9	9	10.2	3.5
8	11	12	12	12.3	3.0
13	12	11	11	11.7	3.4
15	13	13	13	12.5	3.1
6	14	10	10	10.7	4.1
4	15	15	15	13.2	3.0
16	16	16	16	14.8	2.7
11	17	18	18	15.8	2.4
7	18	17	17	15.2	2.8

[a] Ties in the posterior means are due to rounding. The mean rank is 9.5.

for VRQ = 110 (not shown). In this data set, these changes in ranking are small compared to the big impact of using a Bayes versus a least squares approach. We conjecture that the same qualitative results will hold for posterior modal ranks.

Perhaps more important than the comparison of columns 3 and 4 (i.e., ranks of posterior means versus ranks of posterior mean ranks) is the message conveyed by examining the posterior mean and *SD*s of the ranks. For VRQ = 85 there are essentially three groups of indistinguishable schools: schools 18, 10, 2, 16, and 11 with mean rank between 3.2 and 5.9, schools 9, 5, 14, 8, 15, 1, 13, and 17 with mean rank close to $(K + 1)/2 = 9.5$, and schools 6, 3, 4, and 12 with mean rank between 11.3 and 16.7. For VRQ = 110, schools 17 and 12 stand out as being clearly superior.

To illustrate these relations, we plot the posterior expected rank (plus or minus two posterior standard deviations) versus the ranks of the posterior and least squares means for VRQ = 85 (Figures 1 and 2) and VRQ = 110

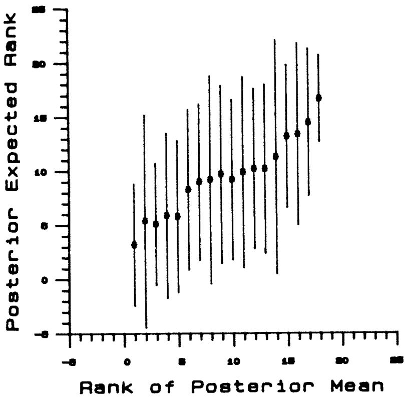

FIGURE 1. *Plot of the posterior mean rank plus or minus 2 standard deviations versus the rank based on the posterior mean predictor (VRQ = 85)*

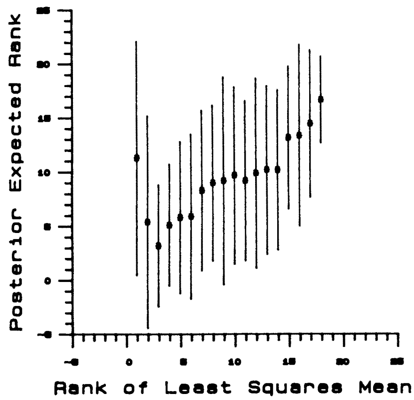

FIGURE 2. *Plot of the posterior mean rank plus or minus 2 standard deviations versus the rank based on the least squares predictor (VRQ = 85)*

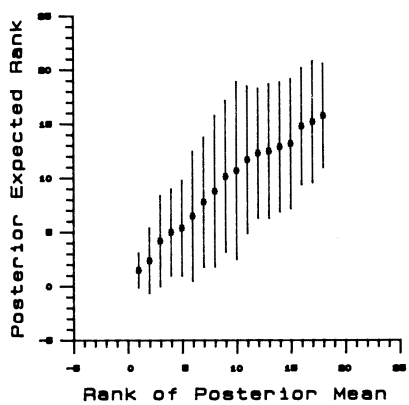

FIGURE 3. *Plot of the posterior mean rank plus or minus 2 standard deviations versus the rank based on the posterior mean predictor (VRQ = 110)*

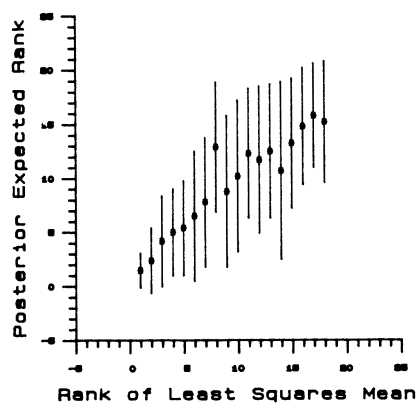

FIGURE 4. *Plot of the posterior mean rank plus or minus 2 standard deviations versus the rank based on the least squares predictor (VRQ = 110)*

(Figures 3 and 4). Figures 1 and 3 illustrate three features of using posterior expected rank instead of the rank of the posterior mean. The shrinkage of the posterior expected ranks to 9.5 is seen through a slope less than one in the trend line through the plotted points. Nonmonotonicity in the plotted points shows that the ranked posterior expected ranks (R_k^*) will not coincide with the ranked posterior means (S_k). The standard errors of the R_k

depend on the P_{jk}s (formula 3) and will generally be smallest for high- or low-ranking schools (as with a binomial-type variance). Figures 3 and 4 show the same general features exaggerated by the differences between least squares and posterior means.

Because the posterior variances and covariances influence the ranks, it can be important to inflate them to account for uncertainty in the marginal maximum likelihood estimate of the fixed effects and variance components. The variance adjustments do influence the ranks somewhat in the school example and can have more dramatic effects in data sets with an estimated prior variance that is small relative to the sampling variance. The adjustment has its most dramatic effect on the estimated posterior *SD* of the rank. For example, the *SD* of .8 for school 17 in Table 3 jumps to 3.0, and the posterior mean goes from 1.5 to 2.7.

Even with the adjustments for using estimates of the prior, the posterior distributions used in the school effectiveness example fail to incorporate correlation among the θ_k and departures from the Gaussian distribution (e.g., skewness) induced by estimation. The correlation can be handled with little additional computation, and the bootstrap (see Laird & Louis, 1987) provides one approach for inducing departures from the Gaussian posterior. The theoretical and empirical effect of these modifications remains to be determined.

Our results promote the concept that ranking procedures should be based on the posterior distribution of the ranks, which conveys more information than integer ranks of parameters. Further research into methods for producing valid confidence intervals for ranks in the empirical Bayes setting is desirable.

APPENDIX A
Equality of the least squares, modal, and expected ranks

Theorem A. If $\theta_1, \ldots, \theta_k$ are independent and stochastically ordered, then

$$\tilde{R}_k = R_k^* = S_k$$

Proof: Let $\theta_k \sim G_k$ and $\bar{G}_k = 1 - G_k$. Without loss of generality take $\bar{G}_k > \bar{G}_{k+1}$. Then, it is easy to show that for $k < m$:

$$P_{mk} < \frac{1}{2}$$

$$P_{jk} < P_{jm} \qquad \text{for all } j \neq k \text{ or } m.$$

Therefore,

$$\begin{aligned}\hat{R}_m = \sum_j P_{jm} &= P_{mm} + P_{km} + \sum_{j \neq k \text{ or } m} P_{jm} \\ &> P_{mm} + P_{km} + \sum_{j \neq k \text{ or } m} P_{jk} \\ &= \hat{R}_k + (P_{km} - P_{mk}) > \hat{R}_k.\end{aligned}$$

Because this is true for any (m, k) pair, it follows that $R_k^* = S_k$. To show $\tilde{R}_k = S_k$, consider the case where $K = 3$ and compare the ranks $R_k \equiv k$ to $R_1 = 1$, $R_2 = 3$, $R_3 = 2$. We have:

$$\begin{aligned}\mathrm{pr}(R_k \equiv k) &+ \mathrm{pr}(R_1 = 1,\ R_2 = 3,\ R_3 = 2)\\ &= \int_{-\infty}^{\infty} g_1(s) \int_{-\infty}^{s} \{\bar{G}_2(t)g_3(t) + \bar{G}_3(t)g_2(t)\}dt\, ds\\ &= \int_{-\infty}^{\infty} g_1(s)[1 - \bar{G}_2(s)\bar{G}_3(s)]ds\end{aligned}$$

and

$$\begin{aligned}\mathrm{pr}(R_k \equiv k) &= \int_{-\infty}^{\infty} g_1(s) \int_{-\infty}^{s} \bar{G}_2(t)g_3(t)dt\, ds\\ &> \int_{-\infty}^{\infty} g_1(s) \int_{-\infty}^{s} \bar{G}_3(t)g_3(t)dt\, ds\\ &= \frac{1}{2}\int_{-\infty}^{\infty} g_1(s)[1 - \bar{G}_3^2(s)]ds\\ &> \frac{1}{2}\int_{-\infty}^{\infty} g_1(s)[1 - \bar{G}_2(t)\bar{G}_3(t)]dt\end{aligned}$$

So $\mathrm{pr}(R_k \equiv k) > \mathrm{pr}(R_1 = 1, R_2 = 3, R_3 = 2)$. In this manner we can prove for a general K that $\tilde{R}_k \equiv k$ is the mode. All we need show is that if $r_k > r_{k+1}$, then

$$\begin{aligned}&\mathrm{pr}(R_1 = r_1, \ldots, R_k = r_k, R_{k+1} = r_{k+1}, \ldots, R_K = r_K)\\ &\qquad < \mathrm{pr}(R_1 = r_1, \ldots, R_k = r_{k+1}, R_{k+1} = r_k, \ldots, R_K = r_K)\end{aligned}$$

A sequence of single exchanges shows that $\tilde{R}_k \equiv k$ is the mode.

Corr.: If $\tau_1 = \tau_2 = \cdots = \tau_k$, then $\tilde{R}_k = R_k^* = S_k$

Proof: If the τs are equal, the Gaussian distributions are stochastically ordered.

APPENDIX B
Variance and covariance of the $\mathrm{R_k}$

Again using the assumption of independent θ_ks, we produce the posterior variance of R_k and the covariance of R_j and R_k.

We have

Theorem:

$$V(R_k) = \sum_{j=1}^{K} P_{jk}(1 - P_{jk}) + 2\sum_{j<1} (C_{jlk} - P_{jk}P_{lk})$$

$$\mathrm{cov}(R_k, R_m) = \sum_{j=1}^{K} (C_{kmj}^* - P_{jk}P_{jm})$$

where

$$C_{jlk} = \mathrm{P}(\theta_k \leq \min(\theta_j, \theta_1))$$

and

$$\begin{aligned} C^*_{kmj} &= \mathrm{P}(\theta_j \geq \max(\theta_k, \theta_m)) \\ &= C_{kmj} \text{ computed under } (-\mu_k, -\mu_m, -\mu_j). \end{aligned}$$

Note that $C_{jjk} = P_{jk}$, $C_{lkk} = P_{lk}$, and $C_{kjk} = P_{jk}$.

Proof:

$$\begin{aligned} \mathrm{V}(R_k) &= \sum_{j=1}^{K} \mathrm{V}(I_{jk}) + 2 \sum_{j<1} \mathrm{cov}(I_{jk}, I_{lk}) \\ &= \sum_{j=1}^{K} P_{jk}(1 - P_{jk}) + 2 \sum_{j<1} [\mathrm{pr}(\theta_k \leq \min(\theta_j, \theta_1)) - P_{jk} P_{jl}], \end{aligned}$$

and

$$\begin{aligned} R_k R_m &= \left(\sum_j I_{jk}\right)\left(\sum_l I_{lm}\right) \\ &= \sum_{j=1}^{K} I_{jk} I_{jm} + \sum_{j \neq 1} I_{jk} I_{lm}. \end{aligned}$$

So,

$$\begin{aligned} \mathrm{E}(R_k R_m) &= \sum_{j=1}^{K} \mathrm{P}(\theta_j \geq \max(\theta_k, \theta_m)) + \sum_{j \neq 1} P_{jk} P_{lm} \\ &= \sum_{j=1}^{K} C^*_{jlk} + \sum_{j \neq 1} P_{jk} P_{lm}. \end{aligned}$$

And

$$\begin{aligned} \mathrm{cov}(R_k, R_m) &= \mathrm{E}(R_k R_m) - \mathrm{E}(R_k)\mathrm{E}(R_m) \\ &= \sum_{j=1}^{K} C^*_{jlk} + \sum_{j \neq 1} P_{jk} P_{lm} - \left(\sum_j P_{jk}\right)\left(\sum_l P_{lm}\right). \end{aligned}$$

Substitutes of the relations among C, C^*, and the Ps produces the covariance. When the θs are correlated, similar, but more complicated, expressions result.

APPENDIX C
Extending the Laird–Ware approach

The approach in Laird and Ware (1982) can be extended in a straightforward way to account for posterior correlation between schools. Doing so, we find that

$$\begin{pmatrix} a_k \\ b_k \end{pmatrix} \text{ and } \begin{pmatrix} a_j \\ b_j \end{pmatrix} \quad k \neq j$$

are jointly multivariate normal, with covariance matrix

$$\begin{bmatrix} V^*_k & C^*_{kj} \\ C^{*t}_{kj} & V^*_j \end{bmatrix}$$

Where

$$C^*_{kj} = V_k \Sigma^{-1} Q \Sigma^{-1} V_j.$$

This implies multivariate normality for the θ_ks, with the same means μ_k, but

$$\tau_k^{*2} = (1X) V_k^* \begin{pmatrix} 1 \\ X \end{pmatrix}$$

and

$$\tau^*_{kj} = \text{cov}(\theta_k, \theta_j | \text{data}) = (1X) C^*_{kj} \begin{pmatrix} 1 \\ X \end{pmatrix}.$$

This approach accommodates uncertainty about (α, β), but not σ^2 and $\boldsymbol{\Sigma}$. Accounting for this latter uncertainty will be important for small to modest K. Whereas (8) dampens the differences in posterior variance and covariance, this adjustment would increase the posterior variance and covariance as a function of the distance between the individual-school estimated parameters and $(\hat{\alpha}, \hat{\beta})$. Because using the Laird–Ware method to account for estimating σ^2 and $\boldsymbol{\Sigma}$ is very difficult, we suggest instead an approximate multivariate δ-method.

Approximate Multiviarate δ-method

We wish to approximate the posterior distribution of a linear functional of a multivariate parameter. Our approximation uses naive empirical Bayes (no account of having estimated the fixed effects and variance components) to produce the multivariate posterior distribution. Then, this distribution produces the naive univariate distribution of the functional, and the method in Morris (1983) provides the adjustment. There are two approximations in this approach: Adjustments are made after applying the functional, and the Morris method produces a Gaussian distribution, even though the true distribution is not Gaussian (see Laird & Louis, 1987).

Specifically, we start with (7) using MML estimates of fixed effects and variance components as the naive estimates. Then, let

$$\hat{\mu} = (1X) \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix},$$

$$b_k = 1 - \frac{\tau_k^2}{(1X) \boldsymbol{\Sigma} \begin{pmatrix} 1 \\ X \end{pmatrix}},$$

and

$$\sigma_k^2 = (1X) D_k \begin{pmatrix} 1 \\ X \end{pmatrix}.$$

We now apply the Morris procedure to the univariate problem, with prior mean $\hat{\mu}$, shrinkage factor b_k, and sampling variance σ_k^2. Thus, the adjusted posterior variance is:

$$\sigma_k^2(1 - b_k) + \frac{\sigma_k^2 b_k}{K} + \frac{2\Delta_k^2 b_k^2}{K - 3},$$

where

$$\Delta_k^2 = (1X) B_k \begin{bmatrix} \hat{a}_k - \hat{\alpha} & \hat{a}_k - \hat{\alpha} \\ \hat{b}_k - \hat{\beta} & \hat{b}_k - \hat{\beta} \end{bmatrix}^t B_k^t \begin{pmatrix} 1 \\ X \end{pmatrix},$$

with, from (6): $B_k = V_k \Sigma^{-1}$.

Other approximations could be used. One defines b_k to be compatible with the relation between the prior and posterior means, and then proceeds as above. Another employs the Laird and Ware (1982) method for accounting for estimating the fixed effects and continues as before, dropping the middle term in the adjusted posterior variance. Further investigation is required to compare these methods to each other and to either a quadrature or Monte-Carlo computation of the moments.

References

Aitkin, M., & Longford, N. (1986). Statistical modeling issues in school effectiveness studies (with discussion). *Journal of the Royal Statistical Society, (A) 149,* 1–43.

Bechhofer, R. E., Kiefer, J., & Sobel, M. (1968). *Sequential identification and ranking procedures.* Chicago: University of Chicago Press.

Berger, J. O., & Deeley, J. (1988). A Bayesian approach to ranking selection of related means with alternatives to analysis-of-variance methodology. *Journal of the American Statistical Association, 83,* 364–373.

Gianola, D., & Fernando, R. L. (1986). Bayesian methods in animal breeding theory. *Journal of Animal Science, 63,* 217–244.

Gibbons, J. D. (1971). *Nonparametric statistical inference.* New York: McGraw-Hill.

Gupta, S. S., & Hsiao, P. (1983). Empirical Bayes rules for selecting good populations. *Journal of Statistical Planning and Inference, 8,* 87–101.

Harville, D. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *Annals of Statistics, 4,* 384–395.

Lagakos, S. W., Wesson, B. J., & Zelen, M. (1986). An analysis of contaminated well water and health effects in Woburn, Massachusetts. *Journal of the American Statistical Association, 81,* 583–614.

Laird, N. M., & Louis, T. A. (1987). Bootstrapping empirical Bayes estimates to account for sampling variation. *Journal of the American Statistical Association, 82,* 739–757.

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics, 38,* 963–974.

Lindley, D. V., & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, B 34*(1), 1–18.

Louis, T. A. (1984). Bayes and empirical Bayes estimates of a population of parameter values. *Journal of the American Statistical Association, 79,* 393–398.

McPherson, K., Wenneberg, J. E., Hovind, O. B., & Clifford, P. (1982). Small area variations in the use of common surgical procedures: An international comparison of New England, England and Norway. *New England Journal of Medicine, 307,* 1310–1314.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association, 78,* 47–59.

Pettitt, A. N. (1982). Parametric tests for agreement amongst groups of judges. *Biometrika, 69,* 365–375.

Pettitt, A. N., & Siskind, V. (1981). Effect of within-sample dependence on the Mann-Whitney-Wilcoxon statistic. *Biometrika, 68,* 437–441.

Portnoy, S. (1982). Maximizing the probability of correctly ordering random variables using linear predictors. *Journal of Multivariate Analysis, 12,* 256–269.

Rubin, D. (1982). Estimation in parallel randomized experiments. *Journal of Educational Statistics, 6,* 377–401.

Tukey, J. W. (1974). Named and faceless values: An initial exploration in memory of Prasant C. Mahalanobis. *Sankhya, 36,* 125–176.

Authors

NAN M. LAIRD, Professor of Biostatistics, Harvard School of Public Health, 677 Huntington Ave., Boston, MA 02115. *Specializations:* statistics and biostatistics.

THOMAS A. LOUIS, Professor and Head, Division of Biometry, School of Public Health, A460 Mayo Memorial Bldg., Box 197, 420 Delaware St. SE, Minneapolis, MN 55455. *Specializations:* statistics, public health, environmental health.