BU-1611-M

PROFILING PHARMACY EXPENDITURES IN MANAGED HEALTH CARE: BAYESIAN INFERENCE FOR A TWO-PART HIERARCHICAL MODEL (DRAFT PAPER)

MIN ZHANG[1], ROBERT L. STRAWDERMAN[1], MARK COWEN[2], AND MARTIN T. WELLS[1]

ABSTRACT. Profiling is currently an important, and hotly debated, topic in health care and other industries that are looking for ways to control costs, increase profitability, and increase service quality. Profiling in health care began in earnest in 1987, when the Health Care Financing Administration attempted to quantify differences in mortality rates in hospitals and link these differences to hospital practice. Since then substantial effort and progress has been made in the areas of statistical modeling and the measurement and comparison of provider performance in more general settings. This paper focuses on the specific problem of developing statistical methods appropriate for profiling physician contributions to patient pharmacy expenditures incurred in a managed care setting. The model developed here accounts for both the skewed, zero-inflated nature of pharmacy expenditure data and the fact that patient pharmacy expenditures are correlated within physicians. For modeling expenditure data, we construct a two-part hierarchical model with a correlated random effect structure. Though our focus is on the problem described above, the model is completely general and can be employed in other settings. Inference is conducted in a Bayesian framework using Markov Chain Monte Carlo. Following Normand *et al.* (1997), provider-level performance indices appropriate for this model are constructed and their posterior distributions are obtained. Physicians are then ranked according to these performance measures, and the posterior distribution of these ranks subsequently serves as the basis for establishing a new financial incentive scheme for physicians working in a managed care setting. Finally, summary measures that attempt to capture the amount of variability in pharmacy expenditures attributable to physicians are also proposed and computed based on our model.

1. INTRODUCTION

In recent years, much effort has been devoted to measuring and comparing the performance of health care providers appropriately and efficiently. This practice is commonly referred to as profiling in the health care services literature, and represents the health care equivalent of producing "report cards" for providers of health care services. One prominent example of interest to the general public is the national rankings of hospitals published annually by US News and World Report. However, profiling in health care can be traced back at least as far as the major initiative launched by the Health Care Financing Administration (HCFA) in 1987 for evaluating hospital performance. HCFA derived estimates of hospital mortality rates based on a patient-level model of mortality, and compared the observed mortality rates to expected mortality rates (adjusted for patient mix and other factors). This was done in order to identify hospitals with excessive mortality rates, which were then subjected to further, more intense study for the purposes of identifying potential quality problems. More recent examples of government-led profiling initiatives include efforts by Pennsylvania and New York to assess the quality of care provided to its patients undergoing coronary artery bypass surgery (Pennsylvania Health Care Cost Containment Council, 1992, 1994a,b; NYS Dept of Health, 1993, 1995, 1996).

Key words and phrases. Profiling; Pharmacy expenditure; Managed health care; Hierarchical model; Intraclass correlation coefficient; Ranking.

Perhaps nowhere has there been more effort devoted to provider profiling than in managed care (Kerr *et al.*, 1995). Here, the primary focus is on measuring various aspects of physician, rather than institutional, performance and the ultimate association of these performance measures with the cost of providing care. The practice of profiling in managed care can have important financial implications for physicians because their compensation is often linked to the results of profiling processes (Cowen and Strawderman, 2002). For example, one area of managed care in which such arrangements are relatively common is pharmacy expenditures. In this article, we profile the "discretionary" aspect of physician prescribing. More precisely, we attempt to quantify the "physician contribution" to a patient's pharmacy expenditures. The results are used to identify physicians with "efficient" and "wasteful" prescribing practices. In a managed care context, these labels might subsequently be associated with financial rewards or penalties, providing some incentive for keeping escalating pharmacy expenditures under control.

The process of provider profiling, however implemented, involves the collection and statistical analysis of data. Objectives common to statistical profiling analysis include the following: (i) the identification of "high" and "low" performers according to some outcome measure(s) of interest; (ii) linking "process" (e.g., practice patterns) with "outcome" (e.g., mortality, cost); and, (iii) "explaining" the variability in outcomes across providers. Some representative examples of recent work either using, developing, or criticizing statistical methods for provider profiling include Chang and McCracken (1996), Cowen and Strawderman (2002), Fowles, Weiner, and Knutson (1994), Gatsonis *et al.* (1995), Goldstein and Spiegelhalter (1996), Hofer *et al.* (1999), Kassirer (1994), Marshall (1999), Miller, Hui *et al.* (1993), Normand *et al.* (1997), Palmer *et al.* (1996), Silber, Rosenbaum, and Ross (1995), and Tucker *et al.* (1996). The data involved in profiling analysis often has a natural hierarchical structure (e.g., patients served by a common physician, physicians within a single hospital, and so on). The use of a hierarchical model allows one to characterize the sources of variability in outcomes across providers: random error; systematic variability among providers caused by measurable factors not under provider control; and, variability among providers caused by measurable factors that are under provider control. An interpretable, quantitative summary of the last is often the main goal of the profiling process; however, the use of hierarchical models in health care profiling has only recently begun to gain favor (Burgess *et al.*, 2000; Cowen and Strawderman, 2002; Normand *et al.*, 1997; Silber, Rosenbaum, and Ross 1995). In the context of a mixed model, the desired component of variation is captured through provider-level random effects, and provider rankings can in whole or part be based on estimates of these effects; see, for example, Miller *et al.* (1993) or Cowen and Strawderman (2002). Statisticians have recently begun to provide convincing arguments for using a fully Bayesian hierarchical model for studying provider-level differences (Gatsonis *et al.*, 1995; Goldstein and Spiegelhalter, 1996; Marshall, 1999; Normand *et al.*, 1997). Fully Bayesian hierarchical models are useful in this setting because inference for random effects and interesting related quantities can be derived directly from the associated posterior distribution; see Normand *et al.* (1997) for further discussion.

To date, much of the statistical literature on profiling has involved comparatively simple hierarchical models. For example, Gatsonis *et al.* (1995) and Normand *et al.* (1997) consider fully Bayesian analysis

of hierarchical logistic regression models, and Goldstein and Spiegelhalter (1996) consider non-Bayesian and Bayesian analysis of simple linear mixed models. Cowen and Strawderman (2002) analyzed pharmacy expenditure data for patients in two managed care organizations (MCOs) using simple linear mixed models. A hierarchical model is clearly appropriate here since patients are nested within their primary care physicians. These authors go on to demonstrate (i) how the results obtained from a simple hierarchical model compare to the status quo in managed care (i.e., ordinary least squares) from the perspective of profiling pharmacy expenditures for physicians working in managed care; and, (ii) that the statistical method used could have a rather substantial impact on incentive schemes derived from the associated results. However, the basic mixed model on which these conclusions are based is technically inadequate for a number of reasons. First, the pharmacy expenditure data are severely skewed. Second, nearly 30% of the patients in the data set have zero pharmacy expenditures for the period under study, implying that the expenditure distribution has a significant point mass at \$0. These factors suggest that pharmacy expenditures, and possibly the underlying random effect distribution used to capture the physician effect, are not normally distributed. These observations are important because Cowen and Strawderman (2002) proposed to profile physicians using summaries derived directly from the estimated random effects, and such estimates can be adversely affected by lack of normality (Verbeke and Lesaffre, 1996). Such potential problems clearly exist more generally when clustering at the provider level is present and the response variable is "semicontinuous" (i.e., zero-inflated but otherwise continuous).

In this article, an appropriate fully Bayesian hierarchical "two-part" model is proposed for skewed semicontinuous response data like that encountered in Cowen and Strawderman (2002). In order to focus our discussion here and elsewhere, it is assumed that the response variable, say Z, represents pharmacy expenditure (hereafter: cost), that observations on Z come from patients nested within primary care physician, and that the goal of the analysis is to profile these physicians with respect to their contributions to patient pharmacy costs. In Section 2, a probability model for Z is devised by proposing separate models for the probability that $Z = 0$ (vs. $Z > 0$) and for the conditional density of $\log Z|Z > 0$. Correlated random intercepts are used in order to link these two models together; see Olsen and Schafer (2001) for a similar idea in the context of longitudinal data analysis. Each random intercept is assumed to consist of a "fixed" and "random" effect. The fixed intercept in the model for $Z = 0$ vs. $Z > 0$ describes the probability that the average physician will issue a prescription; the random effect may be interpreted as the additional propensity of a given physician to prescribe (i.e., compared to the average physician). Similarly, the fixed intercept in the model for $\log Z|Z > 0$ describes the contribution of the average physician to (log) costs among those having some pharmacy cost; the random effect quantifies the magnitude of a specific physician's contribution. The use of correlated random effects is natural here since a physician who is predisposed to prescribe is likely to have higher overall average costs than a physician who is less inclined to issue prescriptions.

Prior distributions appropriate for this two-part hierarchical model are discussed, as is the fitting of this model using Markov Chain Monte Carlo (MCMC) within the statistical package WinBUGS (Spiegelhalter, Thomas and Best, 2000). In Section 3, the focus turns to the specific problem of interest, which is profiling

physicians with respect to their contributions to pharmacy costs. In Section 3.1, a performance index analogous to that defined in Normand *et al.* (1997, §2.2) is proposed; Section 3.2 discusses how this is used to rank physicians. In Section 3.3, measures of the percentage of variability in costs that may be attributed to physicians are proposed based on our two-part model. As commented earlier, performance indicators in managed care may subsequently be used in devising physician incentive schemes. We propose a novel example of such an incentive scheme in Section 3.4 that accounts for the uncertainty in the ranking of a physician according to the chosen performance measure. We use the proposed methods to analyze the pharmacy cost dataset considered in Cowen and Strawderman (2002) in Section 4. Finally, we close the paper with some discussion in Section 5.

2. A HIERARCHICAL MODEL AND BAYESIAN INFERENCE

2.1. **A Two-Part Hierarchical Model.** In context of pharmacy cost data, assume that patient-level cost data are collected for M physicians. For $i = 1 \dots M$ and $j = 1 \dots n_i$, let Z_{ij} denote the observed costs for patient j clustered within physician i, X_{ij} denote a vector of patient specific characteristics, and W_i denote a vector of physician specific characteristics. It is assumed that pharmacy costs are *semicontinuous*; that is, observed costs are assumed to represent realizations of random variables whose probability distributions can be described by a finite mixture of a distribution degenerate at zero and a continuous distribution.

A two-stage structure is used to model the discontinuous nature of the distribution of Z_{ij}. Let $Z_{ij} = I(Y_{ij}^{(1)} > 0) \times \exp(Y_{ij}^{(2)})$, where $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ represent two correlated "latent" random variables. The first stage models patient-level costs, adjusting for patient mix via patient-level characteristics. Specifically:

Stage I. (Patient-Level/Within-Physician)

$$
\begin{aligned}
Z_{ij} &= I[Y_{ij}^{(1)} > 0] \times \exp(Y_{ij}^{(2)}), \\
Y_{ij}^{(1)} &= \alpha_0 + b_i^{(1)} + \alpha_{(1)}^T X_{1ij} + \epsilon_{1ij}, \\
Y_{ij}^{(2)} &= \beta_0 + b_i^{(2)} + \beta_{(1)}^T X_{2ij} + \epsilon_{2ij}, \\
\begin{pmatrix} \epsilon_{1ij} \\ \epsilon_{2ij} \end{pmatrix} &\overset{iid}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma_e^2 \end{pmatrix} \right).
\end{aligned} \tag{2.1}
$$

In the model above, X_{1ij} and X_{2ij} represent the observed characteristics of patient j within physician i. Note that X_{1ij} and X_{2ij} may comprise different subsets of X_{ij}; consequently their dimensions are allowed to be different and are assumed to be d_1 and d_2 respectively. Correspondingly, the dimension for $\alpha_{(1)}$ is d_1 and dimension for $\beta_{(1)}$ is d_2. In addition, the variance of ϵ_{1ij} is set equal to 1 for identifiability and it is assumed that ϵ_{1ij} and ϵ_{2ij} are independent. Finally, in this stage, $b_i^{(k)}$ $(k = 1, 2)$ represent the deviations of physician i from the "average" physician, as is respectively captured by α_0 and β_0. Our assumptions thus far imply that $Y_{ij}^{(1)} \perp Y_{ij}^{(2)}$ given $b_i^{(k)}$.

The second stage of the model links the latent variables $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ together by assuming that $b_i^{(1)}$ and $b_i^{(2)}$ are realizations of correlated random effects. This stage is also used to capture the "physician effect,"

which we decompose into systematic and random effects; see also Normand *et al.* (1997). Specifically:

Stage II. (Physician-Level/Between-Physician)

$$\begin{pmatrix} b_i^{(1)} \\ b_i^{(2)} \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha_{(2)}^T W_{1i} \\ \beta_{(2)}^T W_{2i} \end{pmatrix}, V \right).$$

Similarly to Stage I, the physician-level characteristics W_{1i} and W_{2i} may be different. Their dimensions, and those for $\alpha_{(2)}$ and $\beta_{(2)}$, are assumed to be l_1 and l_2 respectively. The 2×2 variance matrix V is assumed to be positive definite; v_{12} is allowed to be nonzero to account for the potential correlation between $b_i^{(1)}$ and $b_i^{(2)}$. The random effect $b_i^{(1)}$ models the tendency of physician i to prescribe medication for his/her patients; the random effect $b_i^{(2)}$ can be viewed as physician i's inclination to prescribe more expensive medication.

Throughout the remainder of this paper, let $\Phi(\cdot)$ and $\phi(\cdot)$ respectively denote the standard normal cdf and pdf. Let $Z_{ij} = (Z_{ij}^{(1)}, Z_{ij}^{(2)})^T$, where $Z_{ij}^{(1)} = I(Y_{ij}^{(1)} > 0)$ and $Z_{ij}^{(2)} = \exp(Y_{ij}^{(2)})$. Define $Z_i^{(k)} = (Z_{i1}^{(k)}, \ldots, Z_{i,n_i}^{(k)})^T$ for $k = 1, 2$ and $i = 1 \ldots M$. Define $\underline{\alpha} = (\alpha_0, \underline{\alpha}_{(1)}, \underline{\alpha}_{(2)})^T$; $\underline{\beta} = (\beta_0, \underline{\beta}_{(1)}, \underline{\beta}_{(2)})^T$; $\underline{b}_i = (b_i^{(1)}, b_i^{(2)})^T$; $V = \{(v_{rs}) : r = 1, 2.\ \ s = 1, 2\}$; and $\widetilde{\rho} = \frac{v_{12}}{\sqrt{v_{11} v_{22}}}$. Finally, define full data $= \{(Z_i^{(1)}, Z_i^{(2)}, \underline{b}_i),\ i = 1 \ldots M\}$. Then, if full data were completely observed, the associated likelihood for the model parameters $\underline{\alpha}$, $\underline{\beta}$, σ_e^2 and V (i.e., given all covariate information) would be

$$\begin{aligned}
\mathcal{L}(\underline{\alpha}, \underline{\beta}, \sigma_e^2, V | \text{full data}) \;\propto\; & \prod_{i=1}^{M} P(Z_i^{(1)}, Z_i^{(2)} | \underline{b}_i, \underline{\alpha}, \underline{\beta}, \sigma_e^2, V) \times P(\underline{b}_i | \underline{\alpha}, \underline{\beta}, \sigma_e^2, V) \\
= \; & \prod_{i=1}^{M} \prod_{j=1}^{n_i} P(Z_{ij}^{(1)} | \alpha_0, b_i^{(1)}, \underline{\alpha}_{(1)}) \times \prod_{i=1}^{M} \prod_{j=1}^{n_i} P(Z_{ij}^{(2)} | \beta_0, b_i^{(2)}, \underline{\beta}_{(1)}, \sigma_e^2) \\
& \times \prod_{i=1}^{M} P(\underline{b}_i | \underline{\alpha}_{(2)}, \underline{\beta}_{(2)}, V) \\
= \; & \prod_{i=1}^{M} \prod_{j=1}^{n_i} \left[\Phi(\alpha_0 + b_i^{(1)} + \alpha_{(1)}^T X_{1ij}) z_{ij}^{(1)} + \Phi(-\alpha_0 - b_i^{(1)} - \alpha_{(1)}^T X_{1ij})(1 - z_{ij}^{(1)}) \right] \\
& \times \prod_{i=1}^{M} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma_e^2}} \frac{1}{z_{ij}^{(2)}} \exp\left(- \frac{\left[\log z_{ij}^{(2)} - (\beta_0 + b_i^{(2)} + \beta_{(1)}^T X_{2ij}) \right]^2}{2\sigma_e^2} \right) \\
& \times \prod_{i=1}^{M} \frac{1}{2\pi\sqrt{v_{11} v_{22}(1 - \widetilde{\rho}^2)}} \exp\left(- \frac{1}{2(1 - \widetilde{\rho}^2)} \left[\frac{(b_i^{(1)} - \alpha_{(2)}^T W_{1i})^2}{v_{11}} \right.\right. \\
& \left.\left. - 2\widetilde{\rho} \frac{(b_i^{(1)} - \alpha_{(2)}^T W_{1i})(b_i^{(2)} - \beta_{(2)}^T W_{2i})}{\sqrt{v_{11} v_{22}}} + \frac{(b_i^{(2)} - \beta_{(2)}^T W_{2i})^2}{v_{22}} \right] \right).
\end{aligned}$$

This likelihood is similar to that derived in Olsen and Schafer (2001), who considered the case where longitudinal observations are available on the same subject. Olsen and Schafer (2001) also considered a logit, rather than probit, model for the binary portion of the model and based inference for the model parameters on a Laplace-approximated marginal likelihood (i.e., integrating out the random effects). A similar approach might be taken here; as described in Olsen and Schafer (2001, §2.6), inference for the random effects could then be conducted based on empirical Bayes estimates of the posterior means of $b_i^{(1)}$ and $b_i^{(2)}$ for $i = 1 \ldots M$. We instead propose to employ a Bayesian approach, which more closely aligns this

work with the profiling methodology developed in Normand *et al.* (1997). As will be discussed later, the use of the probit link for the binary component of the model has a significant advantage over the logit model in constructing the performance index ultimately used for ranking physicians.

2.2. **Bayesian Inference.**

2.2.1. *Prior choice: regression parameters.* As is clearly explained in Normand *et al.* (1997), a Bayesian approach to the problem of provider profiling has certain advantages not shared by non-Bayesian approaches. For example, in ranking physicians, the posterior distribution of the rank of a provider can be constructed based on some meaningful performance measure. The spread of this posterior distribution reflects the level of uncertainty in the physician's rank, and hence in the profiling process. Such uncertainty is, in our view, important to account for when profiling physicians according to pharmacy costs and in particular when assigning penalties or rewards based on such information.

Philosophical issues aside, the two greatest drawbacks involved in Bayesian inference are computation and prior specification. The posterior distribution of $\underline{\alpha}, \underline{\beta}, \sigma_e^2, V$ and $\underline{b}_i$, $i = 1 \ldots M$ given the observed data is very complicated and neither amenable to analytical calculation nor direct Monte Carlo sampling. Hence, Markov Chain Monte Carlo (MCMC) is used in order to approximate the desired posterior distributions. WinBUGS (Spiegelhalter, Thomas and Best, 2000) is used for this purpose. Normal prior distributions were chosen for the regression model parameters, and an Inverse Gamma prior was chosen for the variance parameter σ_e^2. More precisely, the priors used are as follows:

$$\begin{cases} \alpha_k \sim N(0, 10^6); k = 0, \ldots, d_1; \\ \beta_k \sim N(0, 10^6); k = 0, \ldots, d_2; \\ (\sigma_e^2)^{-1} \sim \Gamma(1, 10^{-3}); \end{cases}$$

where we assume that $\alpha_1, \ldots, \alpha_{d_1}; \beta_1, \ldots, \beta_{d_2}; \sigma_e^2$ are all mutually independent.

2.2.2. *Prior choice: covariance matrix.* Selecting a prior for the covariance matrix V turned out to be a more interesting and challenging problem. Initially, we fit models using an Inverse Wishart prior on V. We found that the posterior distributions of the variance components, which are of central importance in this investigation, are actually quite sensitive to the choice of hyperparameter. The reason for this turns out to be connected to the behavior of the prior itself. The Inverse Wishart prior depends on two parameters: the degrees of freedom and a scale matrix R. In our case, diagonal R matrices are used, and Tables 6 and 7 show that the latter plays a far more significant role than the former. The Inverse Wishart prior has a tendency to force the eigenvalues of the covariance matrix apart. This fact can be proved by demonstrating that the prior depends directly on a Vandermonde determinant that places greater probability density on covariance matrices with such behavior. Spreading out the eigenvalues rather than allowing shrinkage is contrary to typical prior beliefs. Nevertheless, no major difficulties are seen to arise in the case where R is a "small" diagonal matrix (i.e., with positive diagonal elements close to zero). Indeed, in direct analogy to the

Bayes version of estimating the covariance matrix for multivariate normal data, the posterior mean essentially approaches the MLEs of the variance components as R becomes small. However, as R increases, the variance components v_{11}, v_{22} and the correlation ρ are significantly affected. The form of the Inverse Wishart prior shows that the use of a "large" R will place significant probability mass only on "small" covariance matrices having a diffuse eigenstructure. As is discussed in Section 4.2.5, no difficulties were observed in the ability of WinBUGS to update the resulting sampling chains. Hence, we suspect that, despite an apparently ample sample size, the sharp information reflected in this prior was too much for the data to "swamp."

Implied in the above discussion is the fact that the proper Inverse Wishart priors commonly used in practice can have hidden undesirable features. Careful application of objective Bayesian tools can expose these features and allow for the development of better priors. We therefore experimented with various other priors, including the conditionally conjugate prior recently proposed in Daniels and Pourahmadi (2002) and also the reference prior of Yang and Berger (1994). The results of Yang and Berger (1994) suggest that the reference prior should outperform Jeffrey's prior due to the tendency of the latter to spread out the eigenvalues. Since the Inverse Wishart prior is approximately equal to the Jeffrey's prior for "small" R, we suspected that using the Yang and Berger prior may result in some improvement.

The reference prior of Yang and Berger (1994) avoids the problem of placing significant probability mass on covariance matrices with diffuse eigenstructures. However, a practical problem with this prior is its impropriety. Specifically, suppose that our 2×2 covariance matrix V is decomposed as $O'DO$, where O is a certain orthogonal matrix and D is a diagonal matrix with elements $d_2 \geq d_1 \geq 0$; see Yang and Berger (1994) for a more precise formulation. For a 2×2 matrix V, it is easy to show that

$$O = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

where $\theta \in (-\pi/2, \pi/2]$. With this choice of O, the reference prior for $V = O'DO$ (see Yang and Berger, 1994, eqn. 13) reduces to

$$\pi(D, O)(dD)(dO) \propto \frac{I\{d_1 \leq d_2\}}{d_1 d_2}(dD)(dO) \tag{2.2}$$

where $(dD) = (dd_1) \times (dd_2)$ and $(dO) = dO_{12}$. This prior evidently implies the independence of O and D, with the marginal distribution of the former being uniform and that of the latter being improper.

The marginal prior on O is being placed on "$-\sin\theta$" and the implied uniform distribution is therefore Uniform(-1,1) (i.e., a proper prior). As indicated above, the marginal prior on D is improper. However, because of the complicated form of the posterior and because MCMC is being used, it is desirable in practice to choose a proper prior. Hence we propose here a modified version of the reference prior of Yang and Berger (1994) for the 2×2 case whose qualitative behavior is similar to (2.2). Our proposal amounts to replacing the term $\frac{I\{d_1 \leq d_2\}}{d_1 d_2}$ in (2.2) by the joint density of $(\eta_{(1)}, \eta_{(2)})$, where $\eta_{(1)} = \min(\eta_1, \eta_2)$, $\eta_{(2)} = \max(\eta_1, \eta_2)$, and η_1 and η_2 are independent gamma random variables.

A more sampling-oriented version of the proposed prior is now described. With O and D defined as above, the decomposition $V = O'DO$ implies that we may rewrite

$$V = \begin{pmatrix} d_2 \cos^2\theta + d_1 \sin^2\theta & -(d_2 - d_1)\sin\theta\cos\theta \\ -(d_2 - d_1)\sin\theta\cos\theta & d_2 \sin^2\theta + d_1 \cos^2\theta. \end{pmatrix}$$

Let $\gamma = \sin\theta$ and $\delta = d_2 - d_1$; note that $\delta \geq 0$. Then, using the fact that $\sin^2\theta + \cos^2\theta = 1$ and $\cos\theta = \sqrt{1 - \sin^2\theta}$ for $\theta \in (-\pi/2, \pi/2]$, we obtain

$$V = \begin{pmatrix} d_2 - \delta\gamma^2 & -\delta\gamma\sqrt{1-\gamma^2} \\ -\delta\gamma\sqrt{1-\gamma^2} & d_1 + \delta\gamma^2. \end{pmatrix} \tag{2.3}$$

Based on (2.3), our proposed prior on V may then be summarized as the set of random matrices $\mathcal{V}$ such that

$$\mathcal{V} = \begin{cases} \begin{pmatrix} \eta_{(2)} - \Delta\Gamma^2 & -\Delta\Gamma\sqrt{1-\Gamma^2} \\ -\Delta\Gamma\sqrt{1-\Gamma^2} & \eta_{(1)} + \Delta\Gamma^2 \end{pmatrix} \\ \Gamma \sim \text{Uniform}(-1, 1); \\ \eta_1, \eta_2 \sim \text{Gamma}(0.5, 10^2); \\ \Delta = \eta_{(2)} - \eta_{(1)} \end{cases}$$

This prior is used in constructing Tables 1 and 5.

3. A PERFORMANCE INDEX AND RELATED MEASURES

Meaningful comparisons of physician performance require a representative, interpretable measure of performance. Comparison of these measures may then be used to rank physicians accordingly. An interpretable summary measure of the physician contribution to pharmacy costs is also helpful to a managed care organization in targeting directions for improvement. In this section, we propose and calculate such a performance index based on our two-part hierarchical model. We then discuss how to rank physicians in the context of an MCMC sampling scheme. In some managed care settings, including pharmacy costs, physicians may operate under incentive schemes that hold them responsible for some portion of costs deemed "excessive" ("deficit") or provide rewards for keeping costs down ("surplus"). We propose a novel method for calculating these surplus or deficits based on the posterior distribution of physician rankings. Summary measures intended to quantify the extent of variation explained by physician level factors are also defined based on the two-part hierarchical model.

3.1. Definition and Computation of Performance Index.

Following Normand *et al.* (1997), we define the expected outcome (i.e., cost) for the i^{th} physician after risk-adjustment (i.e., patient mix) as μ_i^A, and the standardized outcome (i.e., cost) of the i^{th} physician as μ_i^S. Use of the probit link for the binary stage of our model leads to simple formulas for μ_i^A and μ_i^S, greatly facilitating computation. In contrast, use of a logit link leads to formulas for μ_i^A and μ_i^S with no simple closed form. Our performance index is then defined as $\mu_i^A - \mu_i^S$, and measures the additional (expected) cost attributable to physician i beyond that expected for a

comparable "reference" physician; i.e., one with the same case mix load and physician characteristics. The direct use of $\mu_i^A - \mu_i^S$ as the performance index differs somewhat from Normand *et al.* (1997), who define performance indices derived from a comparison of $\mu_i^A - \mu_i^S$ to some internal or external benchmark.

3.1.1. *Expected Cost.* By definition, the expected cost is the conditional expectation of the physician-average cost, given the physician effect. Mathematically,

$$\mu_i^A = \frac{1}{n_i}\sum_{j=1}^{n_i} E[Z_{ij}|b_i^{(1)}, b_i^{(2)}, W_{1i}, W_{2i}, X_{1ij}, X_{2ij}]. \tag{3.1}$$

Note that μ_i^A also conditions on all model parameters ($\underline{\alpha}$, $\underline{\beta}$, V, σ_e^2); here and elsewhere in this section we do not explicitly indicate this in order to simplify notation. To further simplify notation, we also define

$$\mu_{ij}^A = E[Z_{ij}|b_i^{(1)}, b_i^{(2)}, W_{1i}, W_{2i}, X_{1ij}, X_{2ij}], \tag{3.2}$$

evidently, $\mu_i^A = \frac{1}{n_i}\sum_{j=1}^{n_i}\mu_{ij}^A$.

Using the Stage I model and noting that $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ are independent given $b_i^{(1)}$ and $b_i^{(2)}$, we have

$$\begin{aligned}\mu_{ij}^A &= E\big[\exp(Y_{ij}^{(2)}) \times I(Y_{ij}^{(1)} > 0)|b_i^{(1)}, b_i^{(2)}, W_{1i}, W_{2i}, X_{1ij}, X_{2ij}\big] \\ &= E[\exp(Y_{ij}^{(2)})|b_i^{(2)}, W_{2i}, X_{2ij}] \times E[I(Y_{ij}^{(1)} > 0)|b_i^{(1)}, W_{1i}, X_{1ij}].\end{aligned}$$

Evidently, $E[I(Y_{ij}^{(1)} > 0)|b_i^{(1)}, W_{1i}, X_{1ij}] = P\left(\epsilon_{1ij} > -\alpha_0 - b_i^{(1)} - \alpha_{(1)}^T X_{1ij}\middle|b_i^{(1)}, X_{1ij}\right)$. In addition, given $(b_i^{(2)}, W_{2i}, X_{2ij})$, $\exp(Y_{ij}^{(2)})$ has a log-normal distribution. Therefore,

$$\mu_i^A = \frac{1}{n_i}\sum_{j=1}^{n_i}\left[\exp\left\{\beta_0 + b_i^{(2)} + \beta_{(1)}^T X_{2ij} + \frac{\sigma_e^2}{2}\right\} \times \Phi\left(\alpha_0 + b_i^{(1)} + \alpha_{(1)}^T X_{1ij}\right)\right]. \tag{3.3}$$

The expectation μ_i^A is conditional on $b_i^{(1)}$ and $b_i^{(2)}$ and thus represents a physician-specific quantity.

3.1.2. *Standardized Cost.* The standardized cost for the i^{th} physician is defined as the expected average cost assuming that the patients of physician i were treated by a comparable reference physician. In particular, the standardized cost, denoted μ_i^S, is

$$\mu_i^S = E\left[\mu_i^A\right] = \frac{1}{n_i}\sum_{j=1}^{n_i}\mu_{ij}^S \tag{3.4}$$

where

$$\begin{aligned}\mu_{ij}^S &= E\left[Z_{ij}|W_{1i}, W_{2i}, X_{1ij}, X_{2ij}\right] \\ &= E\left[E\left[Z_{ij}|b_i^{(1)}, b_i^{(2)}, W_{1i}, W_{2i}, X_{1ij}, X_{2ij}\right]\right] \\ &= E\left[\mu_{ij}^A|W_{1i}, W_{2i}, X_{1ij}, X_{2ij}\right]. \qquad (3.5)\end{aligned}$$

Equations (3.4) and (3.5) are obtained by integrating out the physician random effects. Similarly to (3.1), these expectations also condition on all model parameters ($\underline{\alpha}$, $\underline{\beta}$, V, σ_e^2), and this dependence is omitted throughout this section.

Observing that $\mu_{ij}^S = E[\mu_{ij}^A | W_{1i}, W_{2i}, X_{1ij}, X_{2ij}]$ is not particularly helpful in calculating μ_{ij}^S. Instead, we begin with the joint distribution of $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$. Assuming the model of Section 2,

$$\begin{pmatrix} Y_{ij}^{(1)} \\ Y_{ij}^{(2)} \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha_0 + \alpha_{(1)}^T X_{1ij} + \alpha_{(2)}^T W_{1i} \\ \beta_0 + \beta_{(1)}^T X_{2ij} + \beta_{(2)}^T W_{2i} \end{pmatrix}, \Lambda \right),$$

where

$$\Lambda = V + \Sigma = \begin{pmatrix} v_{11} + 1 & v_{12} \\ v_{21} & v_{22} + \sigma_e^2 \end{pmatrix}.$$

The definition of Σ is provided in (2.1). To simplify notation, let μ_1 and σ_1^2 denote the mean and variance for $Y_{ij}^{(1)}$, μ_2 and σ_2^2 denote the mean and variance for $Y_{ij}^{(2)}$, and ρ denote the correlation between $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$, i.e., assume

$$\begin{pmatrix} Y_{ij}^{(1)} \\ Y_{ij}^{(2)} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right).$$

Then,

$$Y_{ij}^{(2)} | Y_{ij}^{(1)} \sim N\left(\mu_2 + \frac{\rho\sigma_2(Y_{ij}^{(1)} - \mu_1)}{\sigma_1}, \sigma_2^2(1-\rho^2) \right).$$

Since

$$\mu_{ij}^S = E\left[E\left[\exp\left(Y_{ij}^{(2)}\right) \times I\left(Y_{ij}^{(1)} > 0\right) \middle| W_{1i}, W_{2i}, X_{1ij}, X_{2ij}, Y_{ij}^{(1)} \right] \right]$$

and $\exp(Y_{ij}^{(2)}) | Y_{ij}^{(1)}$ has a log-normal distribution,

$$\mu_{ij}^S = \int_0^\infty \exp\left(\mu_2 + \frac{\rho\sigma_2(y_{ij}^{(1)} - \mu_1)}{\sigma_1} + \frac{\sigma_2^2(1-\rho^2)}{2} \right) \times \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(- \frac{(y_{ij}^{(1)} - \mu_1)^2}{2\sigma_1^2} \right) dy_{ij}^{(1)}.$$

Let $z = \sigma_1^{-1}\left(y_{ij}^{(1)} - \mu_1\right) - \rho\sigma_2$; then, upon substitution and simplification, it can be shown that $\mu_{ij}^S = \exp(\mu_2 + \frac{\sigma_2^2}{2}) \times \Phi\left(\frac{\mu_1}{\sigma_1} + \rho\sigma_2\right)$. In our original notation,

$$\mu_{ij}^S = \exp\left(\beta_0 + \beta_{(1)}^T X_{2ij} + \beta_{(2)}^T W_{2i} + \frac{v_{22} + \sigma_e^2}{2} \right) \times \Phi\left(\frac{\alpha_0 + \alpha_{(1)}^T X_{1ij} + \alpha_{(2)}^T W_{1i} + v_{12}}{\sqrt{v_{11} + 1}} \right). \tag{3.6}$$

and we set $\mu_i^S = n_i^{-1} \sum_{j=1}^{n_i} \mu_{ij}^S$.

3.2. **Ranking Physicians via $\mu_i^A - \mu_i^S$.** The performance measure by which physicians will be ranked is $\mu_i^A - \mu_i^S$, and measures the incremental per-patient contribution of physician i to pharmacy costs. We interpret a negative difference as saying that a physician prescribes less than expected (i.e., costs the organization less money than expected); this is called a surplus. A positive difference (or deficit) means the opposite. Sorting the differences $\mu_i^A - \mu_i^S$ from lowest (negative) to highest (positive) thus ranks physicians by their relative contributions to per-patient pharmacy costs.

In fitting the model by MCMC, each physician is assigned a rank at each iteration. With M physicians, there are M rank chains, each of which can subsequently be used to approximate the marginal posterior distribution of the rank of each physician. This is one of the advantages of using a Bayesian approach; rather than obtaining a single point estimate of a physician's rank, we obtain the posterior probability associated with each possible rank. This allows us to derive any number of interesting statistics. For example, if we

are interested in the probability that a physician belongs to a certain quantile, we can simply approximate the probability that the i^{th} physician is in the corresponding rank interval R_k as follows:

$$\frac{1}{N}\sum_{t=1}^{N} I(rank[i]^{(t)} \in R_k) \longrightarrow P(rank[i] \in R_k) \quad as \quad N \to \infty, \tag{3.7}$$

where $rank[i]^{(t)}$ is the rank of i^{th} physician at the t^{th} draw of the associated sampling chain, and N is the total number of iterations. Increasing the number of iterations N leads to a more precise sampling-based estimate of this probability.

3.3. **Summarizing the Physician Effect.** One of the goals in Cowen and Strawderman (2002) was to evaluate the extent to which physicians contribute to the overall variation in pharmacy costs. In the context of a mixed effects model, Cowen and Strawderman (2002) proposed using the intraclass correlation coefficient (ICC) to estimate the proportion of variability in a single observation that is attributable to a physician. We now attempt to derive a sensible analog to ICC, referred to as physician explained variation, for the hierarchical two-part model.

3.3.1. *Defining "Explained Variation".* Cowen and Strawderman (2002) considered a linear mixed model having a random intercept term, and assumed that the random intercept fully captured the physician effect. The ICC based on such a model is given by

$$ICC = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}, \tag{3.8}$$

where σ_b^2 denotes the variance of the random intercept term and σ_e^2 denotes the variance of the model error term; see, for example, Crowder and Hand (1990) or Hofer *et al.* (1999). As defined here, ICC captures the physician contribution to the variance in cost for a randomly selected patient. If, for example, $Z_{ij} = \gamma_0 + b_i + \gamma_{(1)}^T X_{2ij} + \gamma_{(2)}^T W_{2i} + \epsilon_{ij}$, (3.8) may also be written

$$\frac{Var(E[Z_{ij}|b_i])}{Var(Z_{ij})}.$$

Importantly, (3.8) is independent of both i and j in the case of the random intercept model. However, if Z_{ij} is more complicated, (3.8) is in general no longer independent of i and j. For example, if Z_{ij} follows the two-part hierarchical model in Section 2, then

$$ICC_{ij} = \frac{Var(E[Z_{ij}|\theta_i])}{Var(Z_{ij})} \tag{3.9}$$

where $\theta_i = (b_i^{(1)}, b_i^{(2)})^T$. The measure (3.9) is a highly nonlinear function of X_{1ij}, X_{2ij}, W_{1i} and W_{2i} and thus depends on both i and j. Consequently, there is no longer an obvious, single summary measure of a physician's contribution to variability. One possibility here is to instead compute ICC_{ij} at one or more "reference" patients and physicians (i.e., fix X_{1ij}, X_{2ij}, W_{1i} and W_{2i}). However, this approach seems questionable unless there is a particular justification for considering particular patient/physician types.

Perhaps the most obvious way to proceed is to average (3.9) over all patients and physicians. For example, consider the class of weighted measures

$$APEV = \frac{1}{M}\sum_{i=1}^{M} w_i \sum_{j=1}^{n_i} ICC_{ij},$$

where $w_1 \dots w_M$ are positive weights and ICC_{ij} is defined in (3.9). $APEV$ can be considered as a measure of the average physician effect on patient-level costs. Because ICC_{ij} is clearly restricted to lie in $[0,1]$, the same bounds apply to $APEV$ provided only that $\sum_{i=1}^{M} w_i n_i = M$. In particular, $APEV$ can equal zero (one) if and only if ICC_{ij} equals zero (one) for all i and j. In addition, in the case where $Z_{ij} = \gamma_0 + b_i + \gamma_{(1)}^T X_{2ij} + \gamma_{(2)}^T W_{2i} + \epsilon_{ij}$ (i.e., a linear mixed model), $APEV$ reduces to ICC as defined in (3.8) upon taking $w_i = n_i^{-1}$. Setting $w_i = n_i^{-1}$ is therefore an intuitively appealing choice, weighting the average contribution of each physician equally (i.e., each physician-averaged ICC receives the same weight). However, there are clearly other sensible choices, including e.g. $w_i = M n_i (\sum_i n_i^2)^{-1}$. This choice would assign greater weight to physicians with larger patient panels, and does not reduce to (3.8) in the case where ICC_{ij} equals (3.8) for all i and j. However, it would continue to satisfy the natural constraint of being equal to zero (one) if and only if ICC_{ij} equals zero (one) for all i and j.

As a second possible measure, we consider constructing the ratio

$$PEV = \frac{\sum_{i=1}^{M} \frac{1}{n_i} \sum_{j=1}^{n_i} Var(E[Z_{ij}|\theta_i])}{\sum_{i=1}^{M} \frac{1}{n_i} \sum_{j=1}^{n_i} Var(Z_{ij})} \tag{3.10}$$

Since $Var(Z_{ij}) = Var(E[Z_{ij}|\theta_i]) + E[Var(Z_{ij}|\theta_i)]$, we may view PEV as being equal to

$$\frac{\sigma^2_{phys}}{\sigma^2_{phys} + \sigma^2_{pat}}$$

where

$$\sigma^2_{phys} = \frac{1}{M}\sum_{i=1}^{M} \frac{1}{n_i} \sum_{j=1}^{n_i} Var(E[Z_{ij}|\theta_i])$$

and

$$\sigma^2_{pat} = \frac{1}{M}\sum_{i=1}^{M} \frac{1}{n_i} \sum_{j=1}^{n_i} E[Var(Z_{ij}|\theta_i)].$$

Here, σ^2_{phys} represents the "average variability in cost attributable to physicians" and $\sigma^2_{phys} + \sigma^2_{pat}$ is the variation in total cost, averaged over all patients. PEV reduces to (3.9) in the case of the random intercept model and obeys the natural restrictions specified above. Weighted versions of this measure could be considered along the lines of $APEV$.

To this point, our discussion has focused on obtaining a point estimate. However, since the expectations and variances involved in the above definitions are in fact conditional on model parameters, there is no single point estimate of PEV and $APEV$. More precisely, these quantities also have posterior distributions. Consequently, relevant posterior summaries (e.g., mean, median, standard deviation) must be used. Notably, posterior summaries of other interesting measures, such as $\text{ICC}_i = n_i^{-1} \sum_{j=1}^{n_i} ICC_{ij}$ (i.e., the contribution of physician i to the variation in cost of his/her own patients), might also be computed, potentially adding an additional dimension to the overall profiling process.

3.3.2. *Computation of APEV and PEV.* Using the definition of μ_{ij}^A in (3.2) and the fact that $\mu_{ij}^S = E[\mu_{ij}^A]$, we may write

$$Var\left(E\left[Z_{ij}|\theta_i\right]\right) \;=\; Var\left(\mu_{ij}^A\right) \;=\; E\left[\left(\mu_{ij}^A\right)^2\right] - \left(\mu_{ij}^S\right)^2.$$

Similarly,

$$Var\left(Z_{ij}\right) \;=\; E\left[\left(\mu_{ij}^A\right)^2\right] - \left(\mu_{ij}^S\right)^2 + \left(E\left[Z_{ij}^2\right] - E\left[(\mu_{ij}^A)^2\right]\right),$$

and consequently

$$APEV \;=\; \frac{1}{M}\sum_{i=1}^{M} w_i \sum_{j=1}^{n_i} \frac{E[(\mu_{ij}^A)^2] - (\mu_{ij}^S)^2}{E[Z_{ij}^2] - (\mu_{ij}^S)^2}$$

and

$$PEV = \frac{\sum_{i=1}^{M} \frac{1}{n_i} \sum_{j=1}^{n_i} \left\{E\left[\left(\mu_{ij}^A\right)^2\right] - \left(\mu_{ij}^S\right)^2\right\}}{\sum_{i=1}^{M} \frac{1}{n_i} \sum_{j=1}^{n_i} \left\{E\left[\left(\mu_{ij}^A\right)^2\right] - \left(\mu_{ij}^S\right)^2 + \left(E\left[Z_{ij}^2\right] - E\left[(\mu_{ij}^A)^2\right]\right)\right\}}$$

Earlier, we showed how to compute μ_{ij}^S; see (3.6). The following result provides the required formulas for the remaining unknown quantities, $E[\left(\mu_{ij}^A\right)^2]$ and $E[Z_{ij}^2]$; this result is proved in Appendix A.

Theorem 3.1. *Define*

$$\mu_{ij}^A = E[Z_{ij}|\underline{\alpha}, \underline{\beta}, \underline{b}_i, W_{1i}, W_{2i}, X_{1ij}, X_{2ij}],\ j = 1 \ldots n_i.$$

Then,

$$E\left[\left(\mu_{ij}^A\right)^2\right] \;=\; \exp\left(2\beta_0 + 2\beta_{(1)}^T X_{2ij} + 2\beta_{(2)}^T W_{2i} + \sigma_e^2 + 2v_{22}\right) \times \int_{-\infty}^{+\infty} \Psi_{ij}(z)\ \phi(z)\ dz \tag{3.11}$$

for

$$\Psi_{ij}(z) = \left[\Phi\left(z\sqrt{v_{11}} + 2v_{12} + \alpha_{(2)}^T W_{1i} + \alpha_0 + \alpha_{(1)}^T X_{1ij}\right)\right]^2$$

and,

$$E\left[Z_{ij}^2\right] \;=\; \exp\left\{2(\beta_0 + \beta_{(1)}^T X_{2ij} + \beta_{(2)}^T W_{2i} + v_{22} + \sigma_e^2)\right\} \times \Phi\left(\frac{\alpha_0 + \alpha_{(1)}^T X_{1ij} + \alpha_{(2)}^T W_{1i} + 2v_{12}}{\sqrt{v_{11}+1}}\right). \tag{3.12}$$

Computational formulas for $APEV$ and PEV can now be obtained using (3.6), (3.11), and (3.12). These formulas involve computing $\int_{-\infty}^{+\infty} \Psi_{ij}(z)\phi(z)dz$ for $i = 1 \ldots M$ and $j = 1 \ldots n_i$, integrals that do not exist in closed form. A well-known method for integrating functions of the indicated form is Gauss-Hermite quadrature (e.g., Davis and Rabinowitz, 1984). However, Gauss-Hermite integration also requires abscissa and weight computations that can be easily avoided in this case. Specifically, since

$$\int_{-\infty}^{+\infty} \Psi_{ij}(z)\phi(z)dz = \int_0^1 \Psi_{ij}(\Phi^{-1}(s))ds,$$

a simpler and more robust quadrature rule, such as an extended trapezoidal or Simpson rule, can be used to compute the desired integrals. Both methods are used in the data analysis.

3.4. **Proposed Incentive Scheme.** Physicians who participate in risk sharing arrangements with a managed care organization may, as appropriate, either be expected to pay some or all of their deficit or receive as a reward some or all of their surplus. Cowen and Strawderman (2002) discuss several possibilities for computing a surplus or deficit (SOD). If a standard performance criteria is available, the SOD can be obtained by taking the difference between the actual cost and the standard cost. Such a standard can be defined under different modeling frameworks (Cowen and Strawderman, 2002). For example, in ordinary least squares regression (Cowen and Strawderman, 2002), the SOD for a given physician is usually calculated as the sum of the actual costs of the patients minus the sum of their predicted costs. Alternatively, the SOD may be calculated by taking the estimated per-patient contribution to costs and multiplying it by physician panel size. Cowen and Strawderman (2002) propose estimating the per-patient contribution via the random intercept term in a linear mixed model. However, per-patient contributions to cost can clearly be defined and estimated under any model framework.

In non-Bayesian approaches to this problem, the surplus or deficit (SOD) is generally computed based on a single point estimate. Typically, the uncertainty associated with these estimates is ignored. Such a system can work well for physicians having large patient panels, and hence rather precise estimates of SOD; however, physicians with smaller patient panels may be unjustly penalized (or rewarded). The Bayesian approach allows one to better incorporate uncertainty in physician rankings into an incentive scheme. Basing the SOD calculation on the posterior distribution of the performance index is one possibility. For example, one could employ the posterior mean of the performance index; however, this does not acknowledge the spread of the posterior distribution. We instead propose to rank physicians according to the performance index, and then use the posterior distribution of the rank to define a per-patient contribution that better takes uncertainty into account (albeit indirectly). We propose two related approaches below.

Suppose that there are M physicians. Partition the set of possible ranks, or $\{1, \ldots, M\}$, into K intervals $R_1, \ldots, R_K$ such that $R_i \cap R_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^{K} R_k = \{1, \ldots, M\}$. It is further assumed that the ranks in R_i are smaller than the ranks in R_j if $i < j$. Suppose that the managed care organization pre-defines a per-patient amount D_k that represents the financial penalty/reward assigned to a physician whose rank falls in the interval R_k. The values of D_k are assumed to be strictly increasing as k runs from 1 to K. For example, if K is odd, the middle interval $R_{\frac{K+1}{2}}$ might receive a weight of $D_{\frac{K+1}{2}} = 0$, representing no reward or penalty for such physicians; in contrast, intervals R_k for k less (greater) than $(K+1)/2$ might receive negative (positive) contributions D_k. Then, one could compute

$$SOD_i = \sum_{k=1}^{K} D_k \times P(rank[i] \in R_k), \tag{3.13}$$

as the "expected" per-patient SOD for physician i. This computation accounts for the uncertainty in physician rank by assigning different penalty/reward amounts physician-specific weights that are derived directly from the posterior probability that a physician's rank respectively falls in $R_1 \ldots R_K$. As described, a negative value of SOD_i corresponds to a surplus, or reward; a positive value represents a deficit, or penalty.

In reality, a managed care organization may budget a fixed amount for pharmacy costs, possibly based on amounts spent in previous years. The above incentive scheme could in principle be modified to adapt the penalty/reward scheme to account for this. For example, suppose that K is odd and $D_{\frac{K+1}{2}} = 0$. Suppose further that $D_k = D \times w_k$, where $w_k = k - \frac{K+1}{2}$ and $D > 0$ is a constant to be determined. Here, the tacit assumption being made is that the increments in D_k move in lockstep away from 0 as the rank moves away from the median. If we assume that physicians are only held responsible for the difference between what was budgeted (T) and the amount actually spent (S), then one may compute D by solving

$$S - T = \sum_{i=1}^{M} n_i \times SOD_i = \sum_{i=1}^{M} n_i \sum_{k=1}^{K} D_k \times P(rank[i] \in R_k) = D \sum_{i=1}^{M} n_i \sum_{k=1}^{K} w_k \times P(rank[i] \in R_k)$$

for D; i.e., by setting

$$D = \frac{S - T}{\sum_{i=1}^{M} n_i \sum_{k=1}^{K} w_k \times P(rank[i] \in R_k)}. \tag{3.14}$$

With D computed, SOD_i can be obtained using $D_k = D \times w_k$ as demonstrated earlier.

One advantage of the proposed incentive scheme is that the actual magnitude of a physician's surplus or deficit is only used to determine a physician's rank; the associated penalty or reward is then determined based solely on the rank. This is in stark contrast to incentive schemes that are currently being used in the industry, some of which are based directly on the OLS residual in a regression analysis employing minimal case mix adjustment. Because the afore-described incentive scheme distributes rewards and penalties symmetrically about the median rank, physicians with small patient panels and hence highly uncertain ranks are unlikely to be penalized or rewarded significantly; on the other hand, physicians with larger patient panels and hence more precise rankings are more likely to be identified with a surplus or deficit. However, the possibility of extreme rewards or penalties is greatly reduced because this incentive scheme limits the maximum per-patient reward (penalty) to be D_1 (D_K).

4. ANALYSIS OF PHARMACY EXPENDITURE DATA

Cowen and Strawderman (2002) analyzed pharmacy cost data taken from two geographically distinct (i.e., western and mid-western) US MCOs at risk for medical and pharmacy expenditures. These authors considered several linear statistical models, and evaluated their respective impacts on cost predictions relevant to the physician profiling process. They also attempted to quantify the level of variability in costs that can be attributed to physicians. As discussed earlier, the models used in that paper are not really appropriate for the zero-inflated, skewed nature of pharmacy cost data. In this section, we apply the model framework of Section 2 to a version of the mid-western MCO data considered in Cowen and Strawderman (2002). We rank all the physicians and compute measures of the "physician-explained variation" (i.e., $APEV$ and PEV) as described in Section 3.3. We also provide some example calculations for the incentive scheme discussed in Section 3.4. The utility of these results relies in large part on both the success of the MCMC sampling scheme and model validity; we also explore these issues to some extent.

4.1. **Data Description.** The original study cohort included 98,382 patients clustered within 252 physicians. The physicians belong to a mid-western US MCO at risk for medical and pharmacy costs (Chernew *et al.* 2000). Each physician serves a different number of patients, ranging from 1 to 1501, with most physicians having $200 - 500$ patients. Information about physician gender, age, and time after graduation from medical school was obtained from the MCO.

The patient-level outcome in this study is total pharmacy cost, defined as the "ingredient cost," or contracted average wholesale price minus a network discount, incurred in 1999. At the time a patient is enrolled into the MCO, he/she is designated a primary care physician to serve as the main coordinator of care (primary care physicians were defined as pediatricians, family physicians and internists). The data was collected by following the convention of assigning pharmacy costs to the primary care physician even if the prescriptions were written by a different provider (Chang *et al.*, 1996; Tucker *et al.*, 1996), as might occur due to a referral.

Cowen and Strawderman (2002) analyzed data on patients aged 18-64. Some exploratory data analysis using both principal components analysis (e.g., see Dunteman, 1989) and SiZer (Chaudhuri and Marron, 1999) confirmed that the cost distributions for pediatricians and physicians treating adult patients were quite distinct. The results reported below are based on physicians who mostly treat adult patients. There are a total of 180 such physicians having 56,522 patients. The patients range in age from 18 to 65 years old, and are included if enrolled in this mid-western MCO for at least 7 months of the year; patients with less than 12 months enrollment had their costs annualized.

As described in Cowen and Strawderman (2002), the available independent patient-level variables include patient age, patient gender and 37 "dummy variables" defining diagnosis groups associated with prescriptions provided to these patients over the course of the study period. Examples of such categories include infection, HIV, liver and pancreas, cancer, coronary artery disease, cystic fibrosis, affective disorders, and so on. The diagnosis categories are constructed using data collected on patient diagnoses from several MCOs during an earlier study period (Cowen, Dusseau, Toth, Guisinger, Zodet 1998). In that study, patient diagnoses were initially identified according to ICD9 codes. Dummy variables representing the major diagnostic categories were then created. The first step was to map the thousands of possible ICD9 diagnoses into one of 260 mutually-exclusive categories of the Clinical Classification Software (CCS or CCHPR) (Elixhauser, Andrews, Fox 1993; Elixhauser 1996). The CCS categories were subsequently collapsed into 37 diagnostic groups using both clinical and statistical criteria. For the current data, patient diagnoses are identified from their ICD9 codes; values for each of the 37 binary covariates (i.e., medical diagnosis groups) are then determined according to whether a patient had at least one medical claim billed with the associated diagnosis during the study period. Further details on the actual process leading up to the identification of these variables can be found in Cowen *et al.*(1998) and Cowen and Strawderman (2002).

Because of limitations with WinBUGS, these 37 diagnostic covariates had to be pared down further in order to fit the model of Section 2. Using an approach similar to that described above, the 37 diagnostic

groups were further refined into 10 diagnostic groups represented by 9 binary covariates. These refined groupings were created to represent those diseases and conditions expected to delineate the major cost categories. We also attempted to maintain some level of clinical similarity within the grouped diagnoses. The ten diagnosis categories ultimately used for adult patients are: infection, diabetes, affective disorders, neuropsychological disorders and HIV (N_PSY_HIV), anxiety, hypertension/hyperlipidemia (HTN_LIPD), heart, asthma and chronic obstructive pulmonary disease (Asthcopd), collagen, and the implied baseline category "other/none." It is noted here that it is possible for patients to have no assigned diagnosis but some positive pharmacy costs; this might happen, for example, if a physician phoned in a prescription for a patient. It is also possible for patients to have some diagnosis but no pharmacy costs; this might happen because a physician recommended an over-the-counter medication or none at all. Also, because of the more limited case-mix adjustment, the total pharmacy cost for each patient is therefore capped at $2,500 in order to prevent unduly penalizing physicians that treat high cost patients. This cap affected 3.8% of the 56,522 patients ultimately used in the analysis.

Further exploratory data analysis using linear and generalized linear mixed models was done to investigate the need for additionally including physician-level characteristics and also interaction terms between patient level covariates. It was determined that interaction terms were not required, and that among physician-level characteristics, only physician gender had a significant impact on the results. Interestingly, patient gender carried little influence. The final set of covariates thus included physician gender, the 9 diagnostic covariates, and patient age. The lone continuous covariate, patient age, was centered at its mean in order to reduce autocorrelation and improve the overall performance of the MCMC sampling scheme (Gelfand, Sahu, Carlin 1995). In fact, the Markov chains for several parameters, including the intercept terms in both parts of the model and the coefficient for age, had convergence problems when using the uncentered age variable.

4.2. **Bayesian Inference on Physician Performance.** WinBUGS version 1.3 (Spiegelhalter, Thomas and Best 2000) was used to fit the mode Bayesian hierarchical model of Section 2. The results reported below use the prior selections discussed in Section 2.2; later, we discuss the impact of these selections on the results.

4.2.1. *Model Estimation.* Using information obtained from the MCMC convergence diagnostics provided by WinBUGS, we determined that it was sufficient to burn-in the first $10,000$ samples and then take the following $10,000$ samples for inference. The posterior means of all parameters, as well as several other posterior summaries (including a 95% credible interval [CI]) , are provided in Table 1.

The regression results reflect the rather unsurprising fact that the probability of having pharmacy costs, as well as total costs incurred, tend to increase with age. The results for the different diagnostic groups, which were selected to reflect our expectations of the major categories of disease/conditions that contribute to pharmacy costs, are somewhat more interesting. For example, it is noted that the posterior mean of the regression coefficient for each of the nine diagnostic covariates is positive. Given the stated expectations,

TABLE 1. Two-Part Hierarchical Model Parameter Estimation

Variable	Parameter	Mean	SD	Mean/SD	Median	95% CI
Intercept	α_0	0.2696	0.0242	11.1636	0.2698	(0.2215,0.3169)
Infection	$\alpha_{(1)1}$	0.7832	0.0131	59.9694	0.7832	(0.7574,0.8095)
Diabetes	$\alpha_{(1)2}$	0.4916	0.0341	14.3995	0.4927	(0.4258,0.5583)
Affective Disorder	$\alpha_{(1)3}$	0.5804	0.0448	12.9438	0.5807	(0.4941,0.6691)
N_PSY_HIV	$\alpha_{(1)4}$	0.4719	0.0735	6.4187	0.4738	(0.3319,0.6143)
Anxiety	$\alpha_{(1)5}$	0.4969	0.0324	15.3364	0.4969	(0.4344,0.5611)
HTN_LIPD	$\alpha_{(1)6}$	0.4791	0.0187	25.6340	0.4787	(0.4432,0.5163)
Heart	$\alpha_{(1)7}$	0.2190	0.0455	4.8164	0.2190	(0.1300,0.3078)
Asthcopd	$\alpha_{(1)8}$	0.5945	0.0283	21.0219	0.5949	(0.5383,0.6499)
Collagen	$\alpha_{(1)9}$	0.5477	0.0906	6.0426	0.5494	(0.37614,0.7300)
Age(patient)	$\alpha_{(1)10}$	0.0106	0.0005	19.9102	0.0106	(0.0096,0.0117)
Gender(physician)	$\alpha_{(2)}$	-0.1527	0.0290	-5.2710	-0.1533	(-0.2087,-0.0947)
Intercept	β_0	4.5950	0.0310	148.1302	4.5950	(4.5320,4.6540)
Infection	$\beta_{(1)1}$	0.3679	0.0156	23.5682	0.3675	(0.3387,0.3997)
Diabetes	$\beta_{(1)2}$	0.9768	0.0315	31.0490	0.9755	(0.9164,1.0420)
Affective Disorder	$\beta_{(1)3}$	1.1950	0.0383	31.1929	1.1940	(1.1180,1.2670)
N_PSY_HIV	$\beta_{(1)4}$	1.2520	0.0646	19.3838	1.2540	(1.1210,1.3790)
Anxiety	$\beta_{(1)5}$	0.7587	0.0300	25.2563	0.7584	(0.7001,0.8167)
HTN_LIPD	$\beta_{(1)6}$	0.6664	0.0199	33.5549	0.6664	(0.6273,0.7051)
Heart	$\beta_{(1)7}$	0.6173	0.0403	15.3367	0.6163	(0.5390,0.6943)
Asthcopd	$\beta_{(1)8}$	0.7873	0.0250	32.4543	0.7876	(0.7366,0.8376)
Collagen	$\beta_{(1)9}$	1.3290	0.0721	18.4379	1.3290	(1.1850,1.4710)
Age(patient)	$\beta_{(1)10}$	0.0382	0.0007	53.3231	0.0382	(0.0368,0.0396)
Gender(physician)	$\beta_{(2)}$	-0.1817	0.0372	-4.8844	-0.1819	(-0.2552,-0.1104)
$\text{Var}(b_i^{(1)})$	v_{11}	0.0175	0.0029	6.1247	0.0173	(0.0126,0.0237)
$\text{Var}(b_i^{(2)})$	v_{22}	0.0335	0.0052	6.4506	0.0332	(0.0243,0.0445)
$\text{Corr}(b_i^{(1)}, b_i^{(2)})$	$\tilde{\rho}$	0.7562	0.0698	10.8369	0.7624	(0.5998,0.8722)
Random Error	σ_e^2	2.3410	0.0166	141.4502	2.3420	(2.3090,2.3740)

these results are not unexpected. However, the relative magnitude of the regression coefficients across the model stages also make a great deal of sense. Consider, for example, the regression coefficients for Infection and Affective Disorders (e.g., depression). Since infections generally require treatment with antibiotics, the probability of receiving a prescription with an infection should be high; however, because infections tend to occur sporadically and typically have both a relatively short duration and cost, the pharmacy cost associated with treating them should be comparatively small. In contrast, affective disorders (e.g., depression) are less prevalent and in many cases do not require a prescription; however, in cases where prescriptions are warranted, treatment and hence medication tends to be both chronic and rather expensive. These trends are reflected in the comparison of regression coefficients, where it is seen that $\alpha_{(1)1} > \alpha_{(1)3}$ but $\beta_{(1)1} < \beta_{(1)3}$.

One interesting and rather surprising result is that patients of female physicians have a higher probability of receiving a prescriptions as well as a higher overall cost. It might be conjectured that female physicians tend to have female patients; since female patients tend to see their doctors more often than male patients, it could be further conjectured that such patients are more likely to have higher pharmacy costs. However, the patient gender by physician gender interaction was deemed insignificant in the exploratory phase of the analysis. Alternatively, it could be conjectured that male physicians are less nurturing than female physicians, and are thus less likely to issue prescriptions. Several other scenarios, none of which can be investigated in useful detail with the available data, are also possible. Whatever the reason, further investigation into this phenomenon seems warranted.

As discussed in Section 2, if $b_i^{(1)}$ is large and positive, this implies that the i^{th} physician tends to issue prescriptions more often than average (all else being the same), resulting in some pharmacy cost; a large

value of $b_i^{(2)}$ implies that his/her patients tend to receive more expensive medication, leading to higher overall pharmacy costs. For the current dataset, this interpretation is not quite so clear cut because the response variable is an aggregated pharmacy cost measure. For example, at this time, it cannot be determined whether the incurred cost for a given patient is due to receiving a few costly medications or several moderately priced prescriptions (e.g., as might be the case for a patient with chronic disease). Nevertheless, it is reasonable to observe the significant positive correlation ($\tilde{\rho}$) between these two random effects, which can be interpreted as saying that physicians who are prone to prescribing more than the average physician are also prone to having higher costs than the average physician. However, it can also be inferred from this table that the effect of physicians on costs pales in comparison to the effect of individuals; we investigate this a bit further in Section 4.2.3 and comment on its implications.

4.2.2. *Evaluation of Physician Performance.* As described in Section 3.1, the performance index for each physician can be calculated for each sampled parameter vector. Hence, physicians may be ranked by the corresponding performance indices at each step. The physician-performance-index chains and rank chains can then be used to study questions related to the physician performance. For example, the rank chain can be used to obtain the estimated posterior distribution for the rank of each physician; the spread of this physician-specific posterior distribution reflects the level of uncertainty in the rank of that physician according to the chosen performance measure.

We computed the performance index $\mu_i^A - \mu_i^S$ for each physician. In Figure 1, we plot the posterior mean $\pm$ posterior standard deviation. It is noted that the y-axis indicates the physician, sorted by associated panel size (in parenthesis). The y-axis sorts physicians by their panel size; while all physicians are represented on the plot, the labels appear only for every sixth physician. This performance plot shows that the adjusted per-patient cost for most physicians are within \$500 of that expected for the "average physician." The extremes are represented by physicians 34 and 59, who respectively spent an average of \$956 less and \$325 more per patient (i.e., as measured by the posterior mean). It is observed that the length of the interval generally decreases as the physician panel size increases, reflecting the fact that the fewer patients a physician has, the greater the spread of the resulting posterior. The plot also shows that one should be particularly careful in drawing conclusions about prescribing behavior of physicians with small patient panels because the posterior spread of the performance measure still varies significantly. However, the intervals for the physicians with the largest panel sizes are still quite wide, especially when viewed in terms of dollars per patient.

Interestingly, the estimated performance of physicians with 500 or more patients is mostly negative (i.e., below average). It is not clear why this is the case. One plausible explanation may be that the average cost for physicians with large patient panels may be diluted by the presence of greater proportions of patients with zero costs. However, the plot of the proportion of zero cost patients versus physician panel size does not reflect any such relationship. A more subtle and interesting explanation, not testable with currently available data, is that panel size effectively serves as a barrier to entry. For example, it is possible that patients served by a physician with a large patient panel have difficulties getting appointments, and thus

have less opportunity to be issued a prescription. Alternatively, it is possible that patients who are more acutely or chronically ill (hence, more likely to require prescriptions) may seek out a physician that serves a smaller number of patients in order to improve their access to care.

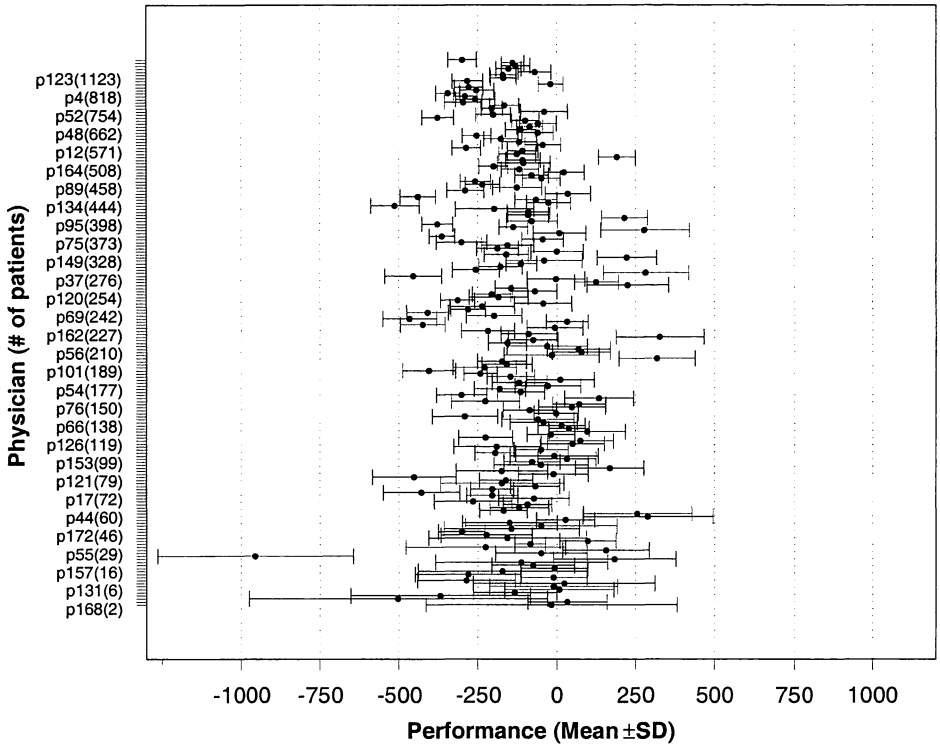

FIGURE 1. Performance plot for all physicians

We also ranked all physicians using the performance index $\mu_i^A - \mu_i^S$. Figure 2 shows the posterior mean and standard deviation for the physicians. The x-axis represents physician rank, and the y-axis again indicates the physician, sorted by associated panel size. In this plot, a smaller rank means "better" performance, in the sense that lower ranks correspond to smaller values of $\mu_i^A - \mu_i^S$. The three dotted vertical lines divide the 180 possible ranks into quartiles. Most of the intervals are wide, extending over two or three quartiles. Out of 180 physicians, only 14 physicians fall completely within the upper quarter and 17 within the lower quarter.

It is observed that the magnitude of a physician's rank is not dependent on his/her panel size (i.e., as measured through its posterior mean); that is, there is no indication that a larger (smaller) panel size leads to higher (lower) rank. However, as was observed in Figure 1, the posterior spread tends to decrease as panel size increases. Furthermore, the posterior distribution for physicians with large numbers of patients are appear close to Gaussian, while others are more skewed (data not shown). Correspondingly, the posterior

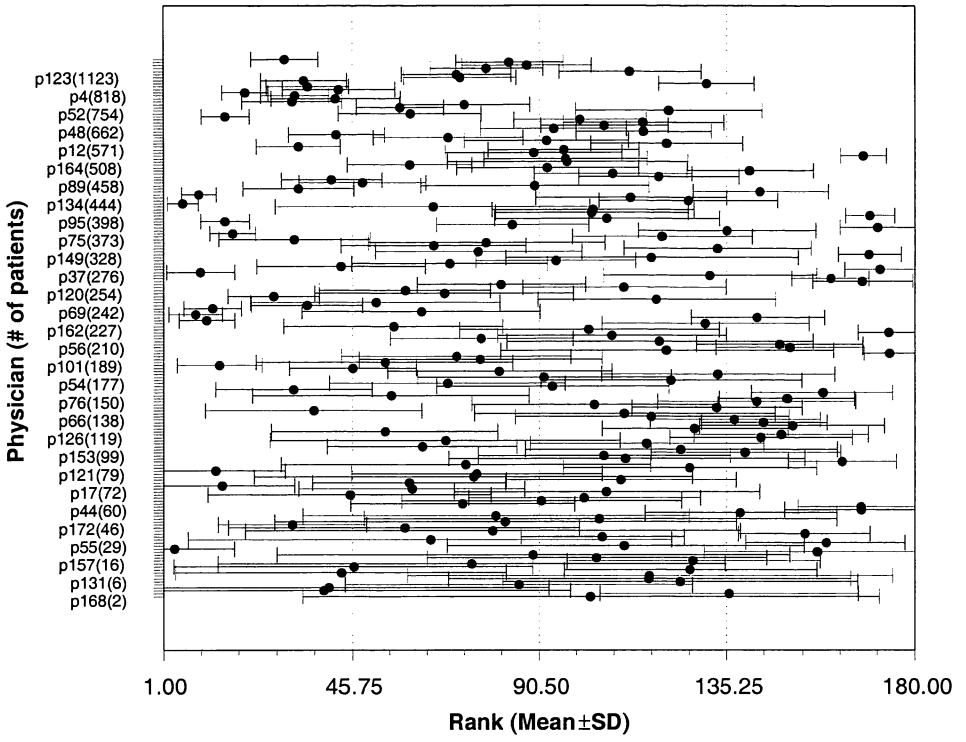

FIGURE 2. Rank plot for all physicians

median rank for physicians with more patients tends to be close to the posterior mean and more discrepant among physicians having smaller patient panels.

The posterior spread of physician rank for physicians with small panel sizes tends to be very broad. For example, the distributions for P168 and P152 span much of the possible rankings (1 to 180). Consequently inference based on a single point estimate of the rank (as might be done in a non-Bayesian setting) could be very misleading, particularly if a physician participates in a risk sharing arrangement with an incentive scheme that depends upon such a point estimate and ignores its uncertainty.

4.2.3. *Quantifying the Physician Effect.* Understanding and quantifying sources of variation is one of the major roles of statistics and also an important goal of this study. In the present context the results may assist a MCO in evaluating whether the modification of the prescribing behavior of physicians is likely to have a significant impact on costs. Two summary measures for the model of Section 2 were defined in Section 3.3; Table 2 shows the results for $APEV$ and PEV.

As discussed in Section 3.3, $APEV$ and PEV have reasonably straightforward interpretations, characterizing the percentage contribution of physicians to the variability in the prescription cost associated with

TABLE 2. Posterior Summaries for PEV and APEV

Parameter	Integration Method†	Mean	SD	Median
APEV ($w_i = n_i^{-1}$)	Simpson	0.0038	0.0006	0.0038
APEV ($w_i = n_i^{-1}$)	Gauss-Hermite	0.0038	0.0006	0.0038
PEV	Simpson	0.0036	0.0005	0.0035
PEV	Gauss-Hermite	0.0036	0.0005	0.0035

† 20-point Gauss-Hermite rule; 19-point extended Simpson rule.

treating a given patient. Consistent with Cowen and Strawderman (2002), the extent of the physician effect is very small by either measure, suggesting that attempts to modify the prescribing behavior of physicians is likely to have little impact on the pharmacy-related operating expenses of the MCO.

4.2.4. *Example Incentive Scheme Computations.* As discussed in Section 3.4, surplus or deficit (SOD) calculations can be used in constructing an incentive scheme that rewards or penalizes physicians for "good" or "bad" performance. The importance of the performance index, as well as its uncertainty, in such a scheme should be obvious. We now illustrate the procedure proposed in Section 3.4 for the 8 physicians considered in Table 3. The following should only be regarded as an example of how one might implement a rank-based incentive scheme in practice.

Suppose we group the set of all possible ranks for any physician into quintiles. With a total of 180 physicians, the corresponding rank intervals are: (0,36], (36,72], (72,108], (108,144], (144,180]. In the notation of Section 3.4, $K = 5$ and these 5 intervals would respectively be denoted $R_1 \dots R_5$. Now, let D_i denote the surplus or deficit assigned to $R_1 \dots R_5$. In this example, we take $D_1 = -\$50$, $D_2 = -\$25$, $D_3 = \$0$, $D_4 = \$25$, and $D_5 = \$50$. Consistent with the notion of surplus or deficit defined in Section 3.4, these amounts represent a per-patient reward (−) or penalty (+), and limits the liability (or reward) to be at most \$50 per patient. The SOD column in Table 3 is now computed as in (3.13). The total SOD, denoted as TSOD in Table 3, represents the total amount a physician would receive (negative) or owe (positive) under this incentive scheme, and is obtained as follows:

$$TSOD_i = n_i \times SOD_i = n_i \times \left(\sum_{i=1}^{5} p_i D_i \right)$$

where n_i denotes the number of patients served by physician i.

TABLE 3. Posterior Rank Estimation and SOD Computation for Selected Physicians

		Rank Distribution						
Physician	Panel Size	Quartile	Mean	SD	Median	SOD	TSOD	$\mu_i^A - \mu_i^S$
P10	438	1st	5.60	3.67	5	-\$49.99	-\$21,899	-\$512.50
P180	2	1st	39.05	58.97	4	-\$30.00	-\$60	-\$502.60
P83	760	2nd	57.16	10.47	57	-\$23.54	-\$17,894	-\$205.20
P13	20	2nd	88.97	61.11	84	-\$1.15	-\$23	-\$112.60
P115	987	3rd	130.50	11.34	131	\$26.60	\$26,264	-\$19.35
P168	2	3rd	102.80	68.76	121	\$7.50	\$15	-\$18.13
P114	535	4th	167.70	5.531	168	\$49.96	\$26,730	\$190.20
P173	2	4th	135.80	30.75	144	\$31.00	\$62	\$32.31

As an example, let us consider physician P168. To approximate the probability that the rank of P168 lies in each of these respective intervals, we used (3.7) based on $N = 10,000$ iterations. For P168, the resulting

probabilities, say p_i, of the rank falling in interval i are: $p_1 = 0.285$, $p_2 = 0.101$, $p_3 = 0.080$, $p_4 = 0.107$ and $p_5 = 0.427$. The SOD and TSOD for this physician are then computed as just described. For example,

$$TSOD_{168} = n_{168} \times SOD_{168} = 2 \times (-\$7.50) = -\$15.00.$$

Based on this calculation, P168 has a deficit of \$15, and would owe this amount to the MCO.

For comparison, the posterior mean of the performance index $\mu_i^A - \mu_i^S$ is provided in the last column; this number should be compared with the column labeled SOD. Multiplying this amount by the panel size leads to a different example of an incentive scheme. Notice, for example, that the associated SOD amounts would be vastly different. For example, P10 would receive \$224,475 (a ten-fold increase) and P115 would receive \$19,098 (instead of owing \$26,264). In the latter case, the posterior standard deviation of $\mu_i^A - \mu_i^S$ is 39.41, indicating that there is also significant probability of having a positive value of $\mu_i^A - \mu_i^S$. In contrast, the posterior standard deviation of this physician's rank, whose mean and median are both well above the median rank of 90.5, is considerably smaller. It is intuitively sensible to expect that the magnitude of $\mu_i^A - \mu_i^S$ may be more adversely affected than the rank by factors like model misspecification, etc ...; hence the proposed incentive scheme might also help insulate physicians from the effects of model-induced errors.

Finally, we provide an example calculation for an incentive scheme involving a budgeted amount. For the dataset being analyzed, the aggregate amount spent by the organization for all patients/physicians is \$22,141,336. Suppose the organization had budgeted \$17,750,000. Then, the value of D in (3.14) is approximately \$25 and the TSOD column of Table 3 reflects what these 8 physicians would owe or receive under such an arrangement.

4.2.5. *Model Evaluation.* We evaluated the model of Section 2 in several ways. First, empirical methods and diagnostic tests for evaluating the convergence of the sampled Markov chains showed no evidence of a lack of stationarity. As a secondary check we also separately fit a generalized linear mixed model (GLMM) to the binary portion of the two-stage model and a linear mixed model (LMM) to the continuous portion. As expected, the posterior means of the regression coefficients, σ_e^2, v_{11} and v_{22} in Table 1 are very similar to their counterparts derived obtained under these models (results not shown). Second, we also employed the logit rather than probit link for binary data, and reran the model. With few exceptions, the estimated probabilities of having positive costs under these two models were found to be quite similar. Finally, we specified different prior distributions for the variance and variance-covariance parameters to evaluate the effects of prior choice on parameters of direct interest. Since variances are non-negative, we chose gamma and inverse gamma distributions with different location and scale parameters to evaluate the effect on the posterior distribution. Table 4 shows the results for different distributions for σ_e^2, and demonstrates a distinct lack of effect on the results. More precisely, we can say that the posterior is insensitive to both conjugate and non-conjugate priors on σ_e^2.

As discussed in Section 2.2.2, we also investigated the sensitivity of the posterior for the variance-covariance matrix V (and other results reported in Table 1) using different prior distributions. The story here becomes

TABLE 4. Posterior Estimation for Different Priors on σ_e^2

Parameter	Γ(1,10)	Γ(10,1)	Γ(10,10)	$I\Gamma$(.01,.01)	$I\Gamma$(.1,.1)
α_0	0.2696(0.0229)	0.2681(0.0234)	0.2699(0.0242)	0.2699(0.0230)	0.2707(0.0223)
$\alpha_{(1)1}$	0.7823(0.0130)	0.7827(0.0131)	0.7824(0.0133)	0.7825(0.0130)	0.7828(0.0125)
$\alpha_{(1)2}$	0.4904(0.0356)	0.4926(0.0354)	0.4927(0.0355)	0.4912(0.0360)	0.4918(0.0363)
$\alpha_{(1)3}$	0.5813(0.0437)	0.5829(0.0443)	0.5811(0.0448)	0.5819(0.0423)	0.5832(0.0447)
$\alpha_{(1)4}$	0.4707(0.0706)	0.4699(0.0708)	0.4684(0.0739)	0.4704(0.0740)	0.4701(0.0725)
$\alpha_{(1)5}$	0.4972(0.0322)	0.4973(0.0322)	0.4955(0.0320)	0.4955(0.0330)	0.4966(0.0333)
$\alpha_{(1)6}$	0.4772(0.0184)	0.4785(0.0189)	0.4780(0.0186)	0.4784(0.0188)	0.4781(0.0190)
$\alpha_{(1)7}$	0.2195(0.0437)	0.2180(0.0448)	0.2189(0.0456)	0.2177(0.0442)	0.2188(0.0444)
$\alpha_{(1)8}$	0.5937(0.0274)	0.5949(0.0278)	0.5953(0.0273)	0.5945(0.0281)	0.5949(0.0273)
$\alpha_{(1)9}$	0.5495(0.0909)	0.5492(0.0878)	0.5488(0.0916)	0.5473(0.0898)	0.5487(0.0908)
$\alpha_{(1)10}$	0.0107(0.0005)	0.0106(0.0005)	0.0106(0.0005)	0.0106(0.0005)	0.0106(0.0005)
$\alpha_{(2)}$	-0.1508(0.0272)	-0.1506(0.0273)	-0.1516(0.0282)	-0.1590(0.0534)	-0.1525(0.0258)
β_0	4.5970(0.0321)	4.5940(0.0313)	4.5950(0.0318)	4.5960(0.0301)	4.5970(0.0302)
$\beta_{(1)1}$	0.3662(0.0154)	0.3684(0.0152)	0.3677(0.0153)	0.3671(0.0154)	0.3681(0.0155)
$\beta_{(1)2}$	0.9763(0.0321)	0.9765(0.0306)	0.9755(0.0315)	0.9771(0.0312)	0.9770(0.0312)
$\beta_{(1)3}$	1.1920(0.0369)	1.1950(0.0364)	1.1950(0.0366)	1.1950(0.0370)	1.1960(0.0369)
$\beta_{(1)4}$	1.2540(0.0633)	1.2500(0.0637)	1.2520(0.0642)	1.2520(0.0635)	1.2540(0.0631)
$\beta_{(1)5}$	0.7597(0.0287)	0.7581(0.0306)	0.7581(0.0296)	0.7587(0.0297)	0.7581(0.0300)
$\beta_{(1)6}$	0.6655(0.0190)	0.6658(0.0192)	0.6650(0.0192)	0.6657(0.0190)	0.6657(0.0193)
$\beta_{(1)7}$	0.6172(0.0388)	0.6179(0.0393)	0.6171(0.0393)	0.6173(0.0392)	0.6188(0.0396)
$\beta_{(1)8}$	0.7874(0.0239)	0.7884(0.0242)	0.7874(0.0239)	0.7883(0.0246)	0.7884(0.0246)
$\beta_{(1)9}$	1.3280(0.0747)	1.3260(0.0723)	1.3290(0.0727)	1.3300(0.0735)	1.3280(0.0733)
$\beta_{(1)10}$	0.0382(0.0007)	0.0382(0.0007)	0.0382(0.0007)	0.0382(0.0007)	0.0382(0.0007)
$\beta_{(2)}$	-0.1812(0.0380)	-0.1811(0.0365)	-0.1797(0.0365)	-0.1821(0.0348)	-0.1830(0.0344)
v_{11}	0.0174(0.0028)	0.0173(0.0028)	0.0174(0.0029)	0.0172(0.0029)	0.0174(0.0028)
v_{22}	0.0335(0.0053)	0.0334(0.0050)	0.0334(0.0051)	0.0334(0.0051)	0.0336(0.0052)
$\tilde{\rho}$	0.7541(0.0697)	0.7561(0.0665)	0.7481(0.0664)	0.7548(0.0663)	0.7562(0.0649)
σ_e^2	2.3390(0.0161)	2.3430(0.0163)	2.3400(0.0165)	2.3420(0.0162)	2.3420(0.0162)

somewhat more interesting. Table 5 explores the sensitivity of the results to variations on the reference prior described in Section 2.2.2. It is noted that these priors are "informative" in the sense that the prior means and variances of the components of V are all finite. Different parameters for the gamma distributions of η_1 and η_2 (i.e., with larger means and variances) were selected to in order to evaluate the sensitivity of the results reported in Table 1. We observe that the posterior distribution of all model parameters and variance components remains very stable.

TABLE 5. Posterior Estimation for Different Yang and Berger Priors

Parameter	Γ(1,1)	Γ(1,10)	Γ(0.5,50)	$I\Gamma$(0.5,100)
α_0	0.2683(0.0235)	0.2708(0.0244)	0.2698(0.0224)	0.2696(0.0242)
$\alpha_{(1)1}$	0.7829(0.0130)	0.7823(0.0132)	0.7830(0.0130)	0.7832(0.0131)
$\alpha_{(1)2}$	0.4925(0.0351)	0.4920(0.0357)	0.4911(0.0352)	0.4916(0.0341)
$\alpha_{(1)3}$	0.5816(0.0441)	0.5790(0.0454)	0.5800(0.0431)	0.5804(0.0448)
$\alpha_{(1)4}$	0.4690(0.0716)	0.4685(0.0730)	0.4728(0.0724)	0.4719(0.0735)
$\alpha_{(1)5}$	0.4978(0.0321)	0.4980(0.0328)	0.4965(0.0313)	0.4969(0.0324)
$\alpha_{(1)6}$	0.4787(0.0188)	0.4775(0.0187)	0.4784(0.0187)	0.4791(0.0187)
$\alpha_{(1)7}$	0.2188(0.0435)	0.2183(0.0438)	0.2196(0.0436)	0.2190(0.0455)
$\alpha_{(1)8}$	0.5942(0.0282)	0.5939(0.0282)	0.5937(0.0271)	0.5945(0.0283)
$\alpha_{(1)9}$	0.5510(0.0891)	0.5504(0.0906)	0.5494(0.0889)	0.5477(0.0906)
$\alpha_{(1)10}$	0.0106(0.0005)	0.0106(0.0005)	0.0106(0.0005)	0.0106(0.0005)
$\alpha_{(2)}$	-0.1628(0.0547)	-0.1520(0.0279)	-0.1519(0.0273)	-0.1527(0.0290)
β_0	4.5970(0.0312)	4.5960(0.0330)	4.5940(0.0311)	4.5950(0.0310)
$\beta_{(1)1}$	0.3667(0.0153)	0.3672(0.0152)	0.3676(0.0165)	0.3679(0.0156)
$\beta_{(1)2}$	0.9764(0.0317)	0.9777(0.0316)	0.9779(0.0316)	0.9768(0.0315)
$\beta_{(1)3}$	1.1950(0.0373)	1.1950(0.0356)	1.1930(0.0374)	1.1950(0.0383)
$\beta_{(1)4}$	1.2520(0.0629)	1.2550(0.0641)	1.2510(0.0646)	1.2520(0.0646)
$\beta_{(1)5}$	0.7582(0.0296)	0.7583(0.0308)	0.7588(0.0302)	0.7587(0.0300)
$\beta_{(1)6}$	0.6655(0.0187)	0.6672(0.0189)	0.6664(0.0197)	0.6664(0.0199)
$\beta_{(1)7}$	0.6155(0.0388)	0.6150(0.0406)	0.6164(0.0399)	0.6173(0.0403)
$\beta_{(1)8}$	0.7878(0.0248)	0.7887(0.0244)	0.7882(0.0239)	0.7873(0.0250)
$\beta_{(1)9}$	1.3310(0.0727)	1.3280(0.0733)	1.3300(0.0731)	1.3290(0.0721)
$\beta_{(1)10}$	0.0382(0.0007)	0.0381(0.0007)	0.0382(0.0007)	0.0382(0.0007)
$\beta_{(2)}$	-0.1818(0.0376)	-0.1822(0.0383)	-0.1797(0.0359)	-0.1817(0.0372)
v_{11}	0.0191(0.0033)	0.0187(0.0031)	0.0178(0.0030)	0.0175(0.0029)
v_{22}	0.0377(0.0062)	0.0372(0.0059)	0.0348(0.0055)	0.0335(0.0052)
$\tilde{\rho}$	0.7612(0.0665)	0.7549(0.0666)	0.7631(0.0689)	0.7562(0.0698)
σ_e^2	2.3380(0.0161)	2.3410(0.0163)	2.3390(0.0162)	2.3410(0.0166)

A considerably more common choice of covariance matrix prior is the Inverse Wishart distribution. We first evaluated the sensitivity for "noninformative" versions of this prior. This was accomplished by setting the degrees of freedom equal to the matrix dimension (i.e., two), and then changing the scale matrix R. These results are reported in Table 6 and show that the smaller the diagonal element, the smaller the posterior mean and corresponding standard deviation of the variance components v_{11} and v_{22} and the larger the correlation $\tilde{\rho}$ between the random effects. The first column of this table agrees reasonably well with the results of Table 1, consistent with earlier statements made regarding expected posterior behavior under an Inverse Wishart prior with small R. However, the remaining columns demonstrate a disturbing level of sensitivity to the scale matrix. No sampling convergence problems were observed for any of these parametrizations, suggesting that the prior distribution itself is the main source of difficulty.

TABLE 6. Posterior Estimation for "Noninformative" Inverse Wishart Priors

Parameter	$R = \begin{pmatrix} 0.01 & 0 \\ 0 & 0.01 \end{pmatrix}$	$R = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$	$R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$R = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$	$R = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}$
α_0	0.2688(0.0225)	0.2692(0.0248)	0.2725(0.0276)	0.2778(0.0465)	0.2848(0.1026)
$\alpha_{(1)1}$	0.7832(0.0130)	0.7825(0.0133)	0.7829(0.0128)	0.7852(0.0129)	0.7841(0.0131)
$\alpha_{(1)2}$	0.4919(0.0350)	0.4917(0.0362)	0.4915(0.0361)	0.4920(0.0372)	0.4944(0.0358)
$\alpha_{(1)3}$	0.5809(0.0435)	0.5835(0.0451)	0.5814(0.0439)	0.5832(0.0444)	0.5807(0.0443)
$\alpha_{(1)4}$	0.4687(0.0699)	0.4693(0.0719)	0.4720(0.0720)	0.4689(0.0713)	0.4684(0.0731)
$\alpha_{(1)5}$	0.4969(0.0315)	0.4967(0.0329)	0.4985(0.0314)	0.5006(0.0327)	0.5029(0.0320)
$\alpha_{(1)6}$	0.4785(0.0186)	0.4782(0.0190)	0.4796(0.0189)	0.4792(0.0184)	0.4809(0.0186)
$\alpha_{(1)7}$	0.2193(0.0437)	0.2176(0.0446)	0.2215(0.0448)	0.2191(0.0442)	0.2211(0.0448)
$\alpha_{(1)8}$	0.5951(0.0280)	0.5944(0.0284)	0.5939(0.0280)	0.5968(0.0284)	0.5987(0.0280)
$\alpha_{(1)9}$	0.5489(0.0883)	0.5477(0.0936)	0.5481(0.0923)	0.5485(0.0889)	0.5512(0.0902)
$\alpha_{(1)10}$	0.0106(0.0001)	0.0106(0.0001)	0.0106(0.0001)	0.0107(0.0005)	0.0106(0.0001)
$\alpha_{(2)}$	-0.1508(0.0270)	-0.1510(0.0293)	-0.1532(0.0331)	-0.1598(0.0568)	-0.1649(0.1271)
β_0	4.5940(0.0303)	4.5940(0.0333)	4.5970(0.0340)	4.6010(0.0497)	4.6170(0.1088)
$\beta_{(1)1}$	0.3682(0.0158)	0.3676(0.0154)	0.3667(0.0154)	0.3680(0.0153)	0.3685(0.0157)
$\beta_{(1)2}$	0.9750(0.0315)	0.9762(0.0314)	0.9757(0.0307)	0.9763(0.0314)	0.9754(0.0309)
$\beta_{(1)3}$	1.1950(0.0367)	1.1940(0.0360)	1.1950(0.0364)	1.1950(0.0372)	1.1980(0.0371)
$\beta_{(1)4}$	1.2520(0.0630)	1.2530(0.0643)	1.2510(0.0635)	1.2470(0.0647)	1.2390(0.0623)
$\beta_{(1)5}$	0.7580(0.0301)	0.7592(0.0289)	0.7567(0.0297)	0.7569(0.0299)	0.7561(0.0298)
$\beta_{(1)6}$	0.6660(0.0189)	0.6650(0.0192)	0.6659(0.0192)	0.6655(0.0192)	0.6636(0.0196)
$\beta_{(1)7}$	0.6171(0.0385)	0.6172(0.0390)	0.6173(0.0394)	0.6182(0.0391)	0.6162(0.0404)
$\beta_{(1)8}$	0.7882(0.0243)	0.7888(0.0248)	0.7896(0.0244)	0.7897(0.0243)	0.7916(0.0240)
$\beta_{(1)9}$	1.3270(0.0714)	1.3260(0.0737)	1.3250(0.0740)	1.3230(0.0726)	1.3210(0.0736)
$\beta_{(1)10}$	0.0382(0.0001)	0.0382(0.0001)	0.0381(0.0001)	0.0381(0.0007)	0.0381(0.0001)
$\beta_{(2)}$	-0.1800(0.0371)	-0.1790(0.0380)	-0.1808(0.0405)	-0.1801(0.0602)	-0.1851(0.1337)
v_{11}	0.0182(0.0031)	0.0201(0.0033)	0.0302(0.0042)	0.0984(0.0115)	0.6498(0.0718)
v_{22}	0.0365(0.0060)	0.0383(0.0060)	0.0484(0.0067)	0.1198(0.0142)	0.6775(0.0741)
$\tilde{\rho}$	0.7787(0.0166)	0.6924(0.0660)	0.4566(0.0764)	0.1714(0.0805)	0.0387(0.0762)
σ_ϵ^2	2.3420(0.0166)	2.3410(0.0165)	2.3410(0.0165)	2.3400(0.0162)	2.3400(0.0165)

The moments of the Inverse Wishart priors of Table 6 are all infinite. Using more informative priors (i.e., those having finite first and second moments), the exceptional sensitivity of the variance components to the scale matrix R remained (results not shown), supporting earlier discussion regarding the importance of this parameter. Finally, we also evaluated whether the Inverse Wishart prior family itself had an effect on our results. In Table 7, we report the results for four different informative Inverse Wishart priors whose first and second moments approximately match those of the Yang and Berger prior specifications used in Table 5. It is interesting to observe that the diagonal elements of R are small in every case. Consistent with earlier results, the variance components v_{11} and v_{22} are seen to be rather insensitive to the choice of prior family. However, there continues to be a noticeable effect on the correlation coefficient $\tilde{\rho}$, which implies a strong effect on v_{12}. In our view, these results demonstrate the inferiority of the Inverse Wishart prior, at least for

this problem. Given the populariy of the Inverse Wishart prior, further investigation into the generality of such phenomena would be worthwhile.

TABLE 7. Posterior Estimation for Informative Inverse Wishart Priors

Scale and df Parameters	$R = \begin{pmatrix} 1/4 & 0 \\ 0 & 5/18 \end{pmatrix}$ $df = 7$	$R = \begin{pmatrix} 1/40 & 0 \\ 0 & 1/32 \end{pmatrix}$ $df = 7$	$R = \begin{pmatrix} 2/153 & 0 \\ 0 & 5/273 \end{pmatrix}$ $df = 6$	$R = \begin{pmatrix} 10/1503 & 0 \\ 0 & 5/552 \end{pmatrix}$ $df = 6$
α_0	0.2701(0.0246)	0.2684(0.0241)	0.2687(0.0221)	0.2670(0.0220)
$\alpha_{(1)1}$	0.7829(0.0129)	0.7832(0.0131)	0.7831(0.0133)	0.7825(0.0132)
$\alpha_{(1)2}$	0.4923(0.0361)	0.4927(0.0353)	0.4925(0.0353)	0.4918(0.0367)
$\alpha_{(1)3}$	0.5825(0.0451)	0.5828(0.0445)	0.5812(0.0444)	0.5816(0.0444)
$\alpha_{(1)4}$	0.4728(0.0723)	0.4714(0.0713)	0.4676(0.0731)	0.4702(0.0720)
$\alpha_{(1)5}$	0.4976(0.0322)	0.4957(0.0312)	0.4977(0.0307)	0.4970(0.0322)
$\alpha_{(1)6}$	0.4785(0.0188)	0.4782(0.0189)	0.4784(0.0193)	0.4779(0.0186)
$\alpha_{(1)7}$	0.2196(0.0442)	0.2186(0.0453)	0.2168(0.0439)	0.2195(0.0436)
$\alpha_{(1)8}$	0.5941(0.0270)	0.5966(0.0276)	0.5951(0.0287)	0.5947(0.0276)
$\alpha_{(1)9}$	0.5472(0.0897)	0.5529(0.0902)	0.5519(0.0882)	0.5482(0.0898)
$\alpha_{(1)10}$	0.0106(0.0006)	0.0106(0.0005)	0.0106(0.0005)	0.0106(0.0005)
$\alpha_{(2)}$	-0.1514(0.0292)	-0.1510(0.0285)	-0.1513(0.0261)	-0.1684(0.0260)
β_0	4.5970(0.0336)	4.5940(0.0319)	4.5950(0.0318)	4.5930(0.0295)
$\beta_{(1)1}$	0.3678(0.0152)	0.3683(0.0152)	0.3671(0.0161)	0.3679(0.0149)
$\beta_{(1)2}$	0.9762(0.0316)	0.9761(0.0310)	0.9766(0.0314)	0.9765(0.0317)
$\beta_{(1)3}$	1.1950(0.0373)	1.1940(0.0369)	1.1940(0.0379)	1.1950(0.0369)
$\beta_{(1)4}$	1.2540(0.0626)	1.2510(0.0626)	1.2540(0.0637)	1.2520(0.0629)
$\beta_{(1)5}$	0.7570(0.0302)	0.7587(0.0299)	0.7586(0.0295)	0.7595(0.0298)
$\beta_{(1)6}$	0.6661(0.0186)	0.6668(0.0187)	0.6660(0.0196)	0.6669(0.0187)
$\beta_{(1)7}$	0.6170(0.0397)	0.6192(0.0386)	0.6172(0.0387)	0.6152(0.0391)
$\beta_{(1)8}$	0.7887(0.0240)	0.7885(0.0245)	0.7885(0.0233)	0.7877(0.0247)
$\beta_{(1)9}$	1.3290(0.0757)	1.3250(0.0727)	1.3290(0.0740)	1.3280(0.0733)
$\beta_{(1)10}$	0.0382(0.0007)	0.0382(0.0007)	0.0382(0.0007)	0.0382(0.0007)
$\beta_{(2)}$	-0.1824(0.0395)	-0.1813(0.0372)	-0.1759(0.0373)	-0.1790(0.0359)
v_{11}	0.0207(0.0032)	0.0172(0.0029)	0.0172(0.0029)	0.0170(0.0029)
v_{22}	0.0386(0.0060)	0.0348(0.0055)	0.0346(0.0057)	0.0345(0.0058)
$\tilde{\rho}$	0.6221(0.0706)	0.7764(0.0611)	0.7873(0.0612)	0.8006(0.0605)
σ_e^2	2.3410(0.0166)	2.3420(0.0164)	2.3420(0.0164)	2.3420(0.0165)

5. DISCUSSION

Statistically speaking, the main contribution of this paper is a Bayesian version of a two-stage hierarchical model that is able to deal with the correlated, highly skewed and semicontinuous nature of patient-level pharmacy cost data. Based on this constructed model, performance indices relevant to identifying "extreme" physicians (i.e., those whose costs appear significantly above or below the "average") are defined. Physicians are subsequently ranked according to their performance. Useful summary measures ($APEV$ and PEV) that help quantify the physician contribution to the observed variability in patient pharmacy costs are also derived. Relevant posterior distributions of physician-level effects (hence associated levels of uncertainty) are obtained as a by-product of MCMC sampling. With 56,522 observations and using a 2.5GHz Pentium-class PC with 1 gigabyte of RAM, fitting this Bayesian hierarchical model in WinBUGS (using 10,000 iterations for burn in and another 10,000 for posterior computations) took approximately 10 hours if we only monitored the model parameters (i.e., $\underline{\alpha}$, $\underline{\beta}$, V and σ_e^2). Monitoring the performance indices and ranks added an additional 4-5 days to the computation time. Though it may be possible to reduce the amount of time required by writing programs (e.g., in C or FORTRAN) designed specifically for fitting this model, the extent of such a reduction is currently unknown and consequently such an involved endeavor is unlikely to be worthwhile. Importantly, though, our results do show that prior selections in this problem ought to be carried out with significant

care, particularly in the case of the variance components. These observations cannot be trivialized for the simple reasons that the performance measures on which the entire profiling process rests depends directly on these variance components; see, for example, (3.3) and (3.6).

From a profiling perspective, the main goal of this study is to gain some useful insight into the extent to which physician practices contribute to pharmacy costs. Ideally, this information should then be used to encourage cost-effective prescribing habits. Incentivizing physicians is a common method of encouragement, and this paper proposes a novel incentive scheme for physicians working under such an arrangement. This scheme combines the uncertainty in a physician's rank with objectively pre-defined criteria for determining financial rewards or penalties. Such features represent a significant improvement over the industry status quo, which in extreme cases have proposed to base incentives directly on OLS residuals obtained from linear regression models with minimal case mix adjustment. The proposed scheme requires computing the posterior probabilities that a physician's rank falls in a pre-determined percentile range. Taking a practical view, given that the MCO will continue to force physicians to accept some component of financial risk, the utilization of posterior rank probabilities as described here has certain undeniable benefits. For example, it provides an effective way to the limit financial liability at the physician level while maintaining some level of fairness in the risk/reward process, recognizing both a physician's apparent (relative) ranking and its associated uncertainty. It also avoids certain ethical issues, such as kickbacks or rebates that might be linked to prescribing for specific patients.

The results summarized in Table 2 suggest that physician behavior explains a comparatively small component of escalating pharmacy expenditures. This conclusion was also reached by Cowen and Strawderman (2002). However, the current paper significantly improves upon the methodology of Cowen and Strawderman (2002), whose conclusions were largely based on point estimates derived from the fits of simple linear mixed models. Though the tiny magnitudes of $APEV$ and PEV do not necessarily negate the benefits of incentivizing physicians to reduce costs, they do suggest that some proportionality is in order with regard to the financial burden imposed by the incentive scheme and the level of attention currently being paid to individual physicians.

Normand *et al.* (1997) comment that the use of ranks in profiling are of limited value, a statement requiring some qualification. The work of Goldstein and Spiegelhalter (1996) to which these authors refer is based on ranking posterior means. Louis (1996) points out that computing the posterior probability associated with each potential rank is far more preferable, particularly when posterior variances are unequal. The comparatively small physician effect might be interpreted as suggesting that one should not focus on physician rankings unless there are important and measurable differences in the performance index between, say, rank quartiles. Such a concern seems valid if rankings are based on point estimates alone (e.g., posterior means), but is less worrisome if the entire posterior distribution of each physician's ranking is used. In fact, the computation that a physician's rank lies in a prespecified percentile group represents a typical example of the type of performance index proposed in Normand *et al.* (1997, eqn. 5). That is, these probabilities

could themselves be used as the basis for constructing additional performance measures, or for other actions a managed care organization may wish to pursue with such information. For example, rather than attaching financial compensation directly to a physician's rank distribution, the MCO could use this information as a way to identify physicians for more detailed study. "Academically detailing" a physician is known to be one of the more useful interventions for changing prescribing habits. However, it is also an expensive and labor intensive process. The rank distribution of a physician, possibly combined with additional information, could be used to as a way to identify physicians that could potentially benefit from this practice.

In the current dataset, patient pharmacy costs are actually documented according to a predefined set of 28 pharmacy classes. These classes are distinct from the *diagnosis* categories being used for case-mix adjustment. In particular, the pharmacy cost "response" variable used in this study represents an aggregated patient-level cost totalled over all 28 pharmacy classes for the period of one year. Consequently, given the limited set of covariates ultimately used, it is not possible to discern whether a high cost patient arises as the result of a few costly medications or from a chronic disease requiring prescription medication for its management. Incorporating the number of prescriptions written for a patient into the analysis, either by using the mean cost or by adjusting for it as an additional covariate, might be helpful in this regard. From a profiling perspective, this is likely to be as or more useful than using longitudinal information at the patient level. On the other hand, using longitudinal information on the performance indices at the physician level could prove to be very valuable indeed. Further insight could potentially be gained by breaking these aggregated costs down into their respective pharmacy classes, an approach explored in Cowen and Strawderman (2002). All of these suggestions significantly increase the modeling and computational burden, and the "added value" of doing such an in-depth analysis must certainly be weighed against these costs.

Further attempts to improve the validity of the profiling process must focus on the inherently multivariate nature of the problem. Though the focus on pharmacy costs alone is perhaps justified for the limited goals of this analysis, it is clearly insufficient for profiling providers or the process of care more generally. For example, if the physicians costing the organization significantly less money on pharmacy are also those providing a relatively low quality of care, both the patients and organization may be adversely affected. The adverse effect on patients is clear and requires no elaboration. The effect on the organization is primarily economic, since patients receiving a distinctly lower quality of care may look elsewhere for care and/or file suit against the organization. Consequently the process of profiling providers should involve several dimensions of care, and not focus on a single performance measure. Challenges here include both determining the most appropriate indicators of care and constructing useful, interpretable multivariate models. Some progress in this area in the case of simple Bayesian hierarchical models has been made. For example, Landrum, Bronskill, and Normand (2000) employ latent variable models for this purpose assuming that (1) data are aggregated at the provider level and (2) the multiple dimensions of care all consist of binomial counts (i.e., the components of the multivariate response are all of the same basic data type). Extending this methodology to mixed response data (available either at the provider or patient level) would be worthwhile, provided suitable dimensions of care can be identified and lead to interpretable summaries of provider performance.

In any event, the major challenge that currently lies before the managed care industry is to identify real differences arising as a result of modifiable behavior at the patient, physician, and organizational levels, and to devise solutions to these problems that encourage positive and effective change.

6. Appendix: Proof of Theorem 3.1

We first establish a more general result that implies (3.11). To simplify notation, let $\delta_{1ij} = \alpha_0 + \alpha_{(1)}^T X_{1ij}$ for $j = 1 \ldots n_i$. Then, from the arguments leading up to (3.3), we find

$$\begin{aligned} E\left[\mu_{ij}^A \mu_{ik}^A\right] &= E\left[\exp\left(2b_i^{(2)} + 2\beta_0 + \beta_{(1)}^T X_{2ij} + \beta_{(1)}^T X_{2ik} + \sigma_e^2\right) \times \Phi(b_i^{(1)} + \delta_{1ij}) \times \Phi(b_i^{(1)} + \delta_{1ik})\right] \\ &= \exp\left(2\beta_0 + \beta_{(1)}^T X_{2ij} + \beta_{(1)}^T X_{2ik} + \sigma_e^2\right) \times E\left[e^{2b_i^{(2)}} \Phi(b_i^{(1)} + \delta_{1ij})\Phi(b_i^{(1)} + \delta_{1ik})\right] \end{aligned}$$

Let $A_1 = \exp\left(2\beta_0 + \beta_{(1)}^T X_{2ij} + \beta_{(1)}^T X_{2ik} + \sigma_e^2\right)$. Then,

$$\begin{aligned} E\left[\mu_{ij}^A \mu_{ik}^A\right] &= A_1 \times E\left[E\left[e^{2b_i^{(2)}} \Phi(b_i^{(1)} + \delta_{1ij})\Phi(b_i^{(1)} + \delta_{1ik}) \middle| b_i^{(1)}\right]\right] \\ &= A_1 \times E\left[\Phi(b_i^{(1)} + \delta_{1ij})\Phi(b_i^{(1)} + \delta_{1ik}) E\left[e^{2b_i^{(2)}} \middle| b_i^{(1)}\right]\right]. \end{aligned}$$

By assumption,

$$\begin{pmatrix} b_i^{(1)} \\ b_i^{(2)} \end{pmatrix} \sim N\left(\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}\right),$$

where $m_1 = \alpha_{(2)}^T W_{1i}$ and $m_2 = \beta_{(2)}^T W_{2i}$. Thus,

$$b_i^{(2)} | b_i^{(1)} \sim N\left(m_2 + \frac{v_{12}(b_i^{(1)} - m_1)}{v_{11}}, v_{22} - \frac{v_{12}^2}{v_{11}}\right)$$

and consequently

$$\begin{aligned} E\left[\mu_{ij}^A \mu_{ik}^A\right] &= A_1 \times E\left[\Phi(b_i^{(1)} + \delta_{1ij})\Phi(b_i^{(1)} + \delta_{1ik}) \exp\left(2\left[m_2 + \frac{v_{12}(b_i^{(1)} - m_1)}{v_{11}} + v_{22} - \frac{v_{12}^2}{v_{11}}\right]\right)\right] \\ &= A_1 \times \exp\left(2\left[m_2 - \frac{m_1 v_{12}}{v_{11}} + v_{22} - \frac{v_{12}^2}{v_{11}}\right]\right) \times E\left[\Phi(b_i^{(1)} + \delta_{1ij})\Phi(b_i^{(1)} + \delta_{1ik}) \exp\left(\frac{2v_{12}}{v_{11}} b_i^{(1)}\right)\right]. \end{aligned}$$

Define $A_2 = A_1 \times \exp\left(2\left[m_2 - \frac{m_1 v_{12}}{v_{11}} + v_{22} - \frac{v_{12}^2}{v_{11}}\right]\right)$. Then,

$$E\left[\mu_{ij}^A \mu_{ik}^A\right] = A_2 \times \int_{-\infty}^{+\infty} \Phi(b_i^{(1)} + \delta_{1ij})\Phi(b_i^{(1)} + \delta_{1ik}) \frac{\exp\left(\frac{2v_{12}}{v_{11}} b_i^{(1)} - \frac{(b_i^{(1)} - m_1)^2}{2v_{11}}\right)}{\sqrt{2\pi v_{11}}} db_i^{(1)}$$

Now, define $z = \frac{b_i^{(1)} - m_1}{\sqrt{v_{11}}} - \frac{2v_{12}}{\sqrt{v_{11}}}$; then, $b_i^{(1)} = z\sqrt{v_{11}} + 2v_{12} + m_1$. Defining $\Psi_{ij}(z) = \Phi(z\sqrt{v_{11}} + 2v_{12} + m_1 + \delta_{1ij})$, we have

$$\begin{aligned} E\left[\mu_{ij}^A \mu_{ik}^A\right] &= A_2 \times \int_{-\infty}^{+\infty} \Psi_{ij}(z)\Psi_{ik}(z) \exp\left(\frac{2v_{12}}{v_{11}}(z\sqrt{v_{11}} + 2v_{12} + m_1)\right) \frac{1}{\sqrt{2\pi v_{11}}} \exp\left(-\frac{(z + \frac{2v_{12}}{\sqrt{v_{11}}})^2}{2}\right) \sqrt{v_{11}} dz \\ &= A_2 \times \exp\left(\frac{2v_{12}}{v_{11}}(v_{12} + m_1)\right) \times \int_{-\infty}^{+\infty} \Psi_{ij}(z)\Psi_{ik}(z)\phi(z) dz \end{aligned}$$

Upon rewriting this expression in our original notation and simplifying the result, we obtain

$$E\left[\mu_{ij}^A \mu_{ik}^A\right] = \exp\left(2\beta_0 + \beta_{(1)}^T X_{2ij} + \beta_{(1)}^T X_{2ik} + 2\beta_{(2)}^T W_{2i} + \sigma_e^2 + 2v_{22}\right) \times \int_{-\infty}^{+\infty} \Psi_{ij}(z)\Psi_{ik}(z)\ \phi(z)\ dz$$

where $\Psi_{ij}(z) = \Phi(z\sqrt{v_{11}} + 2v_{12} + \alpha_{(2)}^T W_{1i} + \alpha_0 + \alpha_{(1)}^T X_{1ij})$. This proves (3.11), as $E[(\mu_{ij}^A)^2]$ is a special case of this result with $j = k$.

To establish (3.12), it is helpful to recall that $Z_{ij} = I(Y_{ij}^{(1)} > 0)\exp(Y_{ij}^{(2)})$. Thus,

$$\begin{aligned} E\left[Z_{ij}^2\right] &= E\left[I\left(Y_{ij}^{(1)} > 0\right)\exp\left(2Y_{ij}^{(2)}\right)\right] \\ &= E\left[E\left[\exp\left(2Y_{ij}^{(2)}\right)I\left(Y_{ij}^{(1)} > 0\right)\middle| Y_{ij}^{(1)}\right]\right] \\ &= E\left[E\left[\exp\left(2Y_{ij}^{(2)}\right)\middle| Y_{ij}^{(1)}\right] I\left(Y_{ij}^{(1)} > 0\right)\right] \end{aligned}$$

Using the simplified notation of (3.6), we know

$$Y_{ij}^{(2)}|Y_{ij}^{(1)} \sim N\left(\mu_2 + \frac{\rho\sigma_2(Y_{ij}^{(1)} - \mu_1)}{\sigma_1}, \sigma_2^2(1-\rho^2)\right).$$

Thus

$$\begin{aligned} E\left[Z_{ij}^2\right] &= E\left[\exp\left(2\left[\mu_2 + \frac{\rho\sigma_2(Y_{ij}^{(1)} - \mu_1)}{\sigma_1} + \sigma_2^2(1-\rho^2)\right]\right) I(Y_{ij}^{(1)} > 0)\right] \\ &= \int_0^\infty \exp\left(2\left[\mu_2 + \frac{\rho\sigma_2(Y_{ij}^{(1)} - \mu_1)}{\sigma_1} + \sigma_2^2(1-\rho^2)\right]\right) \frac{\exp\left(\frac{(y_{ij}^{(1)} - \mu_1)^2}{-2\sigma_1^2}\right)}{\sqrt{2\pi\sigma_1^2}} dy_{ij}^{(1)} \end{aligned}$$

Letting $z = \frac{y_{ij}^{(1)} - \mu_1}{\sigma_1} - 2\rho\sigma_2$, we obtain after simplification the simple expression

$$E\left[Z_{ij}^2\right] = e^{2(\mu_2 + \sigma_2^2)}\Phi\left(\frac{\mu_1}{\sigma_1} + 2\rho\sigma_2\right).$$

Placing this back in our original notation,

$$E\left[Z_{ij}^2\right] = \exp\left\{2(\beta_0 + \beta_{(1)}^T X_{2ij} + \beta_{(2)}^T W_{2i} + v_{22} + \sigma_e^2)\right\} \Phi\left(\frac{\alpha_0 + \alpha_{(1)}^T X_{1ij} + \alpha_{(2)}^T W_{1i} + 2v_{12}}{\sqrt{v_{11} + 1}}\right),$$

proving the result and the theorem. □

REFERENCES

1. Berger JO (1985). *Statistical Decision Theory and Bayesian Analysis.* New York: Springer-Verlag.
2. Bernardo JM (1979). Reference posterior distributions for Bayesian inference (with discussion). Journal of the Royal Statistical Society Ser. B, 41:113-147.
3. Box GEP, Tiao G (1973). *Bayesian Inference in Statistical Analysis.* London: Addison-Wesley.
4. Burgess JF, Christiansen CL, Michalak SE, Morris CN (2000). Medical profiling: improving standards and risk adjustments using hierarchical models. *Journal of Health Economics*, 19:291-309.
5. Chang W, McCracken SB (1996). Applying case mix adjustment in profiling primary care physician performance. *Journal of Health Care Finance*, 22:1-9.
6. Chaudhuri P, Marron JS (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association*, 94:807-823.
7. Chernew M, Cowen ME, Kirking DM, Smith DG, Valenstein P, Fendrick AM (2000). Pharmaceutical cost growth under capitation: A case study. *Health Affairs*, 19:266-276.
8. Cowen ME, Dusseau DJ, Toth BG, Guisinger C, Zodet MW (1998). Case mix adjustment of managed care claims data using the clinical classification for Health Policy Research (CCHPR) method. *Medical Care*, 36:1108-1113.
9. Cowen ME, Strawderman RL (2002). Quantifying the physician contribution to managed care pharmacy expenses: a mixed effects approach. *Medical Care*, 40:650-661.
10. Daniels MJ, Pourahmadi M (2002). Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika*, 89:553-566.
11. Davis PJ, Rabinowitz P (1984). *Methods of Numerical Integration, 2nd ed.* San Diego: Academic Press.
12. Dunteman GH (1989). *Principal Components Analysis.* (eds Lewis-Beck M.S.) SAGE Publications, Inc.
13. Elixhauser A, Andrews RM, Fox S (1993). Clinical classification for health policy research: Discharge statistics by principal diagnosis and procedure. *Division of provider studies research note 17, agency for health policy and research* Rockville, MD: Public Health Service.
14. Elixhauser A (1996). Clinical classifications for health policy research, version 2: software and user's guide. *Healthcare cost and utilization project, research note 2.*
15. Fowles JB, Weiner JP, Knutson D (1994). A comparison of alternative approaches to risk management. Final report to the Physician Payment Review Commission, November 30.
16. Gatsonis CA, Epstein AM, Newhouse JP, Normand SL, McNeil BJ (1995). Variations in the utilization of coronary angiography for elderly patients with an acute myocardial infarction. *Medical Care*, 33:625-642.
17. Gelfand AE, Sahu SK, Carlin BP (1995). Efficient parametrisations for normal linear mixed models. *Biometrika*, 82: 479-88.
18. Goldstein H, Spiegelhalter D (1996). League tables and their limitations: statistical issues in comparisons of institutionsl performance. *Journal of the Royal Statistical Society, Series A*, 159:385-443.
19. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG (1999). The unreliability of individual physician "Report Cards" for assessing the causes and quality of care of chronic disease. *Journal of the American Medical Association,* 281:2098-2105.
20. Kass RE, Wasserman L (1996). The selection of prior distribution by formal rules. *Journal of the American Statistical Association*, 91:1343-1370.
21. Kassirer J (1994). The use and abuse of practice profiles. *New England Journal of Medicine,* 330:634-635.
22. Kerr EA, Mittman BS, Hays RD, Siu AL, Leake B, Brook RH (1995). Managed care and capitation in California: how to physicians at financial risk control their own utilization? *Annals of Internal Medicine,* 123:500-504.
23. Landrum MB, Bronskill SE, Normand SL (2000). Analytic methods for constructing cross-sectional profiles of health care providers. *Health Services and Outcomes Research*, 1, 23-47.
24. Louis TA (1996). Discussion of Goldstein H, Spiegelhalter D (1996) "League tables and their limitations: statistical issues in comparisons of institutionsl performance", in *Journal of the Royal Statistical Society, Series A*, 159:423-424.
25. Marshall EC (1999). Statistical Methods for Institutional Comparisons. PhD thesis, Emmanuel College, Cambridge.
26. Miller ME, Hui SL, Tierney WM, McDonald CJ (1993). Estimating physician costliness: an empirical Bayes approach. *Medical Care,* 31:YS16-28.
27. New York State Department of Health (1993). *Coronary Artery Bypass Surgery in new York State, 1990-1992,* Albany: New York. New York State Department of Health.
28. New York State Department of Health (1995). *Coronary Artery Bypass Surgery in new York State, 1991-1993,* Albany: New York. New York State Department of Health.
29. New York State Department of Health (1996). *Coronary Artery Bypass Surgery in new York State, 1992-1994,* Albany: New York. New York State Department of Health.
30. Normand SL, Glickman ME, Gatsonis CA (1997). Statistical methods for profiling providers of medical care: Issues and applications. *Journal of the American Statistical Association*, 92:803-814.
31. Olsen M, Schafer JL (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statisitical Association,* 96:730-745.
32. Palmer RH, Wright EA, Orav EJ, Hargraves JL, Louis TA (1996). Consistency in performance among primary care practitioners. *Medical Care*, 34:SS52-SS66.
33. Pennsylvania Health Care Cost Containment Council (1992). *A Consumer's Guide to Coronary Artery Bypass Surgery,* Harrisburg, PA:PH4C.
34. Pennsylvania Health Care Cost Containment Council (1994a). *A Consumer's Guide to Coronary Artery Bypass Surgery, Vol II,* Harrisburg, PA:PH4C.

35. Pennsylvania Health Care Cost Containment Council (1994b). *A Consumer's Guide to Coronary Artery Bypass Surgery, Vol III,* Harrisburg, PA:PH4C.
36. Silber JH, Rosenbaum PR, Ross RN (1995). Comparing the contribution of groups of predictors: which ourcomes vary with hospital rather thant patient characteristics. *Journal of the American Statistical Association,* 90:7-18.
37. Spiegelhalter D, Thomas A, Best N (2000). WinBUGs Version 1.3 User Manual.
38. Tucker AM, Weiner JP, Honigfeld S, Parton RA (1996). Profiling primary care physician resource use: examining the application of case mix adjustment. *Journal of Ambulatory Care Management,* 19:60-80.
39. Verbeke G, Lesaffre E (1996). A linear mixed-effects model with heterogeneity in the random-effects populations. *Journal of the American Statistical Society,* 91:217-21.
40. Yang R, Berger JO (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, 22:1195-1211.

[1] Biological Statistics and Computational Biology, Cornell University, Ithaca NY 14853

[2] Quality Institute, St. Joseph Mercy Health System, Ann Arbor MI 48105

Correspondence to: Min Zhang, Biological Statistics and Computational Biology, Cornell University, Ithaca NY 14853

E-mail address: mz35@cornell.edu