8-26-2004

A Hierarchical Multivariate Two-Part Model for Profiling Providers' Effects on Healthcare Charges

John W. Robinson
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, jrobinso@jhsph.edu

Scott L. Zeger
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, szeger@jhsph.edu

Christopher B. Forrest
Johns Hopkins Bloomberg School of Public Health, Department of Health Policy & Management, cforrest@jhsph.edu

A Hierarchical Multivariate Two-Part Model for Profiling Providers' Effects on Healthcare Charges

John W. Robinson, Scott L. Zeger, and Christopher B. Forrest[1]

August 26, 2004

[1]John W. Robinson is a doctoral student, Department of Biostatistics, The Johns Hopkins University Bloomberg School of Public Health, 615 N. Wolfe St., Baltimore, MD, 21205 (email: jrobinso@jhsph.edu). Scott L. Zeger is Professor and Chair, Department of Biostatistics, The Johns Hopkins University Bloomberg School of Public Health. Christopher B. Forrest is Associate Professor, Department of Health Policy and Management, The Johns Hopkins University Bloomberg School of Public Health.

Abstract

Procedures for analyzing and comparing healthcare providers' effects on health services delivery and outcomes have been referred to as provider profiling. In a typical profiling procedure, patient-level responses are measured for clusters of patients treated by providers that in turn, can be regarded as statistically exchangeable. Thus, a hierarchical model naturally represents the structure of the data. When provider effects on multiple responses are profiled, a multivariate model rather than a series of univariate models, can capture associations among responses at both the provider and patient levels. When responses are in the form of charges for healthcare services and sampled patients include non-users of services, charge variables are a mix of zeros and highly-skewed positive values that present a modeling challenge. For analysis of regressor effects on charges for a single service, a frequently used approach is a two-part model (Duan, Manning, Morris, and Newhouse 1983) that combines logistic or probit regression on any use of the service and linear regression on the log of positive charges given use of the service. Here, we extend the two-part model to the case of charges for multiple services, using a log-linear model and a general multivariate log-normal model, and employ the resultant multivariate two-part model as the within-provider component of a hierarchical model. The log-linear likelihood is reparameterized as proposed by Fitzmaurice and Laird (1993), so that regressor effects on any use of each service are marginal with respect to any use of other services. The general multivariate log-normal likelihood is constructed in such a way that variances of log of positive charges for each service are provider-specific but correlations between log of positive charges for different services are uniform across providers. A data augmentation step is included in the Gibbs sampler used to fit the hierarchical model, in order to accommodate the fact that values of log of positive charges are undefined for unused service. We apply this hierarchical, multivariate, two-part model to analyze the effects of primary care physicians on their patients' annual charges for two services, primary care and specialty care. Along the way, we also demonstrate an approach for incorporating prior information about the effects of patient morbidity on response variables, to improve the accuracy of provider profiles that are based on patient samples of limited size.

Key Words: Gibbs sampler; Data augmentation; Rejection sampling; Primary care; Referral to specialists; Point-of-service health plan.

1 INTRODUCTION

1.1 Provider Profiling

Procedures for analyzing and comparing providers' effects on healthcare delivery and outcomes have been referred to as provider profiling (DeLong et al. 1997; Normand, Glickman, and Gatsonis 1997; Daniels and Gatsonis 1999). In early profiling studies, provider effects were typically represented as fixed parameters (Blumberg 1988; Hannan, Kilburn, O'Donnell, Lukacik, and Shields 1990; Williams, Nash, and Goldfarb 1991; Salem-Schatz, Moore, Rucker, and Pearson 1994), applying what Goldstein and Spiegelhalter (1996) have referred to as an "independent" model. In more recent studies, statisticians have consistently preferred hierarchical models (Thomas and Longford 1994; Goldstein and Spiegelhalter 1996; Normand, Glickman, and Gatsonis 1997; Daniels and Gatsonis, 1999; Burgess, Christainsen, Michalak, and Morris 2000; Shaihan et al. 2001; Landrum, Normand, and Rosenheck 2003; Liu, Louis, Pan, Ma and Collins 2003). This evolving consensus stems from the design of the typical profiling study, in which a patient-level response is measured for clusters of patients treated by providers that in turn, can be regarded as statistically exchangeable (Goldstein and Spiegelhalter 1996). Patient sample sizes and morbidity burden often vary widely between profiled providers, and although researchers routinely employ structural models to control for variation in morbidity, the effectiveness of such approaches is incomplete and inconsistent (Green and Wintfeld 1995; Iezzoni 1997; Shaihan et al. 2001). Thus, hierarchical models represent a better conceptual fit and can be expected to provide more accurate and precise estimates of provider effects (Goldstein and Spiegelhalter 1996; DeLong et al. 1997; Burgess et al. 2000; Shaihan et al. 2001).

1.2 Multivariate Provider Profiling

The majority of reported hierarchical profiling studies have involved univariate responses (Landrum, Normand, and Rosenheck 2003). However, important research and policy questions often involve multivariate responses and the associations among them. For example, the association between the use of primary care and specialty care is of interest to healthcare policy-makers and payers. There is evidence that medical and surgical specialists provide more resource-intensive services than do primary care physicians (PCPs), even when managing common, uncomplicated problems (Greenfield et al. 1992; Carey et al. 1995). Such evidence has contributed to concerns that patients who self-refer to specialists for routine care might incur unnecessary expense and expose themselves to unnecessary tests or procedures (Franks, Clancy, and Nutting 1992); concerns that in turn, have contributed to decisions by managed care organizations to establish "gatekeeping" rules for health maintenance organization (HMO) members, whereby a visit to a specialist

is paid for only if referral to the specialist was approved by a member's PCP (Kerr, Mittman, Hays, Siu, Leake, and Brook 1995).

Gatekeeping has been unpopular with patients (Bodenheimer 1996, Grumbach et al. 1999) and PCPs (Halm, Causino, and Blumenthal 1997; Peter, Reed, Kemper, and Blumenthal 1999) who regard it as a barrier to care, and in response, managed care organizations have increasingly offered "point-of-service" (POS) plans, that blend HMO, preferred-provider, and traditional indemnity benefits, allowing members to bypass their PCPs and self-refer to specialists in exchange for higher out-of-pocket payments (Bodenheimer 1996). In a study of 3 geographically diverse POS plans, 17% to 30% of members who visited specialists self-referred (exercising their preferred-provider or indemnity benefit) while the remainder obtained referrals from their PCPs (Forrest et al. 2001). Those who self-referred reported less satisfaction with their PCPs and more established relationships with their specialists (Braun et al. 2003), suggesting that PCPs affect patients' utilization of specialists both *directly* by approving or disapproving referrals and *indirectly* through the quality of their work with patients.

1.3 A Multivariate Two-Part Model

In analyses that include users and non-users of a health service, charges for that service are a mixture of zeros and highly-skewed, continuously distributed, positive values that cannot be approximated by any simple parametric form. This mixture can be thought of as arising from a two-part process, the first part determining whether any use of the health service occurs and the second part determining the amount of charges given use, hence, the conceptual basis for the two-part model (Duan, Manning, Morris, and Newhouse 1983). Under the two-part model, charges for a health service are represented by a binary variable, U that equals 1 if any of the service was used and 0 if not, and a continuous variable Y that equals the log of charges if $U = 1$. Effects of covariates on U are modeled using logistic or probit regression and on $Y\,|U = 1$, using a linear model with normal errors or generalized linear model with gamma errors (Diehr, Yanez, Ash, Hornbrook, and Yin 1999).

Here, we extend the two-part model to the case of multiple charge variables, using a log-linear model and a general multivariate log-normal model, and employ the resultant multivariate two-part model as the within-provider, or likelihood component of a hierarchical model. For p charge variables, or services, the multivariate two-part model involves p deterministic and $\sum_{k=1}^{p} 4\,(k-1)$ stochastic associations between elements of variable vectors $\mathbf{U}$ and $\mathbf{Y}$. These within-patient associations and approaches to modeling them are presented schematically for $p = 2$, in Figure 1. Following Fitzmaurice and Laird (1993), we re-parameterize the log-linear model, so that random provider effects on a patient's probability of use of each service are

marginal with respect to use of other services, while within-patient associations between use of different services are represented as conditional log odds ratios.

Modelling the logs of positive charges is complicated by the fact that observed vectors, $\mathbf{y} = (y_1, ..., y_p)'$ contain undefined elements for patients who do not use all services. Thus, we assume the presence of an underlying vector, $\mathbf{Y}^* = \left(Y_1^*, ..., Y_p^*\right)'$ from a general multivariate normal distribution, representing the *potential* logs of positive charges, and let $Y_k = Y_k^*$ if $U_k = 1$ and Y_k be undefined if $U_k = 0$, for $k = 1, ..., p$. To fit the model as a function of $\mathbf{Y}^*$, we include a data augmentation step in the Gibbs sampler used to fit the hierarchical model, that replaces undefined elements of $\mathbf{y}$ with random draws from the full conditional distribution of the corresponding subvector of $\mathbf{Y}^*$ at each Gibbs iteration (Tanner and Wong 1987; Gelfand and Smith 1990). The resultant augmented response vectors, $\mathbf{y}^*$ are free of undefined elements, yet are informed only by the observed data and model assumptions. Hence, the augmentation is true to the fact that undefined elements correspond to observed values of charges that are zero, *not* missing.

Three other points about the specification of the general multivariate log-normal model deserve emphasis: First, as suggested in Figure 1, we regress each element of Y_k^*, on indicators of any use of services other than k, in order to estimate the effects of any use of one service on the log of positive charges for each other service. Second, we allow the variance of log of positive charges to be provider-specific, but moderate that assumption through the use of a prior specification. And third, we assume that correlations between the logs of positive charges for different services are uniform across providers.

1.4 Related Work

Landrum, Normand, and Rosenheck (2003) describe a hierarchical multivariate profiling model that is similar to ours in many respects, but takes an entirely different approach to modeling within-patient associations. They use a pair of two-part models to estimate provider effects on utilization of two services, outpatient and inpatient mental health care. The first part of each of their two models is a probit regression on any use of service and the second part is a multivariate normal regression on 3 measures of level of service use. Substantive considerations led them to represent the within-patient association between utilization of the two services solely through random patient effects that are shared between the two probit regressions but not with the second part of either of the two models. The result is that each patient's effect on the probability of any use of one service is assumed to be the *same* as that patient's effect on the probability of any use of the other service and given any use of the other service, *independent* of the level of use of the other service. Additionally their models assume that if both services are used, the level of use of each service is *independent* of the level of use of the other. None of these assumptions would have been appropriate to

the application that we were considering, as suggested by the above discussion of the association between the use of primary care and specialty care.

A small number of other reported hierarchical profiling studies have also involved multivariate responses. Landrum, Bronskill, and Normand (2000) profiled hospitals that treated patients for myocardial infarction, by estimating a latent quality trait using 4 binary measures of treatment quality and outcome. Burgess, Lourdes, and West (2000) profiled psychiatric hospitals by estimating hospital-specific time-series parameters affecting a binary indicator of appropriate post-hospitalization care, measured in 10 consecutive years. Bronskill, Normand, Landrum, and Rosenheck (2002) profiled cardiac surgeons by estimating parameters that described their longitudinal effects on post-operative mortality over 6 consecutive years and profiled mental health networks by estimating their longitudinal effects on rates of psychiatric re-admission. None of these three studies involved health care charges as a response or employed a two-part model as we do here.

2 DATA SOURCE AND RISK ADJUSTMENT

2.1 The POS Health Plan Study

To evaluate the performance of our profiling model, we drew a sample of 50 primary care physicians (PCPs), all family physicians, participating in a POS plan offered by a not-for-profit insurer in the Northeast. The POS plan was one of three that contributed administrative data to a large study of referral patterns in POS health plans (Forrest et al. 2001). Each member of the northeastern POS plan selected a PCP from among those participating in the plan; female members could optionally select an obstetrician-gynecologist as a second PCP (an "ObGyn-PCP"); and all members could change PCPs and/or ObGyn-PCPs as often as they wished. When a member exercised the HMO benefit, the member's PCP or ObGyn PCP functioned as a gatekeeper, deciding whether to authorize specialist referrals. Alternatively, a member could exercise the preferred-provider or indemnity benefit and self-refer to a specialist, at higher out-of-pocket cost.

2.2 Profiling Study Sample Selection

In order to justify the assumption of provider exchangeability, we chose to profile PCPs from one primary care specialty, family practice, and in order to assure that each PCP's patient sample would include an adequate number of users of specialty care, we selected the 50 family practice PCPs with the largest case-loads of POS plan members. Our patient sample consisted of all adult, male POS plan members who were enrolled for 12 months of 1996, were assigned solely to one of the 50 selected PCPs, received at least one claimed health service, and were not diagnosed with a psychiatric disorder. Reasons for the exclusions were

as follows: females and members that changed PCPs during 1996 because our model could not estimate the combined effects of two or more PCPs on a patient's utilization; members younger than 18 because we planned to later compare family practice PCPs with internal medicine PCPs, who do not generally treat that age group; members 65 and older because the POS plan did not cover seniors; members enrolled fewer than 12 months because we required 12 months of claims data; members with psychiatric diagnoses because psychiatric claims data was not available; and members without claims because at least one claim was necessary for risk adjustment. Applying these criteria, 3,308 patients were included, resulting in within-PCP sample sizes ranging from 30 to 152.

We used annual *allowed* charges per patient by service to measure utilization. We chose allowed charges rather than billed charges, because allowed charges were set by the managed care plan and therefore generally uniform across providers. Although the model presented in Section 3 can be applied to any number of services, the application presented in Section 4 involves just two, primary care and specialty care; where primary care and specialty care refer to outpatient evaluation and management services provided by PCPs and medical and non-ophthalmologic surgical specialists, respectively.

2.3 Risk Adjustment Using ACGs

For the POS study, each patient had been assigned to one of 93 Adjusted Clinical Groups (ACGs), mutually exclusive categories based on age, gender, and 12 months of diagnoses. The ACGs were developed by researchers and practicing physicians to sort patients, solely based on information from health care claims, into face-valid categories predictive of current and future health services utilization (Health Services Research and Development Center 2001). The ability of ACGs to predict utilization can be roughly measured by an ANOVA of annual, per patient charges for ambulatory health services, using ACGs as a one-way classification. Investigators have done so, using data from various public- and private-sector health plans, and have obtained values of R^2 ranging from .34 to .47 (Weiner, Starfield, Steinwachs, and Mumford 1991; Reid, MacWilliam, Verhulst, Roos, and Atkinson 2001).

Each of the 3,308 selected patients was in one of 40 ACGs. (Only 40 of 93 ACGs were represented because many applied only to women or children.) Some of the 40 ACGs were represented so infrequently that we could not validly estimate their effects on utilization solely on the basis of information contained in our patient sample. For example, six ACGs had frequencies of less than 10 among the full 3,308 patients, and five had frequencies of one or two among the 1,004 patients with positive charges for specialty care. However, each of the 40 ACGs was well represented in the overall POS plan membership from which our sample had been drawn. Thus, using the 38,878 adults enrolled in the POS plan for 12 months of 1996, we

ranked the 40 ACGs on each of the two services, primary care and specialty care, on the basis of percent of members using the service and mean charges among users of the service, resulting in four sets of ranks. Hence, for each ACG that was infrequent in our profiled sample, we had identified other ACGs of similar rank that were better represented, allowing us through model specification (described in Section 3.1) to borrow strength from well-represented ACGs in estimating the effects of infrequent ACGs.

3 MODEL SPECIFICATION

3.1 Likelihood

Let $C_{ijk} = \begin{cases} 0 & \text{if} \quad U_{ijk} = 0 \\ \exp\left(Y^*_{ijk}\right) & \text{if} \quad U_{ijk} = 1 \end{cases}$, where C_{ijk} represents annual allowed charges (in dollars) for service k, for patient j, of PCP i, and U_{ijk} and Y^*_{ijk} are distributed as defined in Sections 3.1.1 and 3.1.2.

3.1.1 Part One: Reparameterized Log-Linear Model

For part one of the likelihood, let

$$P\left(\mathbf{U}_{ij} = \mathbf{u}_{ij} | \mathbf{\Psi}_{ij}, \mathbf{\Omega}\right) = \exp\left\{\mathbf{\Psi}'_{ij}\mathbf{u}_{ij} + \mathbf{\Omega}'\mathbf{w}_{ij} - A(\mathbf{\Psi}_{ij}, \mathbf{\Omega})\right\},$$

where $\mathbf{U}_{ij} = (U_{ij1}, \ldots, U_{ijp})'$; $\mathbf{\Psi}_{ij} = \left(\psi_{ij1}, \ldots, \psi_{ijp}\right)'$; $\mathbf{W}_{ij} = (U_{ij1}U_{ij2}, \ldots, U_{ijp-1}U_{ijp}, \ldots, U_{ij1}U_{ij2}\cdots U_{ijp})'$ is a vector of $2^p - p - 1$ two- and higher-way cross products of elements of $\mathbf{U}_{ij}$; $\mathbf{\Omega} = \left(\omega_{12}, .., \omega_{(p-1)p}, \ldots, \omega_{12\ldots p}\right)'$; and $A(\mathbf{\Psi}_{ij}, \mathbf{\Omega}) = \log \sum_{\mathbf{U}_{ij}} \exp(\mathbf{\Psi}'_{ij}u_{ij} + \mathbf{\Omega}'w_{ij})$ is a normalizing constant, with $\sum_{\mathbf{U}_{ij}}$ representing summation over all 2^p possible values of $\mathbf{U}_{ij}$.

Note that ψ_{ijk} represents the log odds of any use of service k, given no use of any other service. However, since we are interested in parameters that represent the marginal, rather than conditional log odds of any use of each service, following Fitzmaurice and Laird (1993), we make the 1:1 transformation $(\mathbf{\Psi}_{ij}, \mathbf{\Omega}) \rightarrow (\boldsymbol{\pi}_{ij}, \mathbf{\Omega})$, where $\boldsymbol{\pi}_{ij} = (\pi_{ij1}, ..., \pi_{ijp})$ and $\pi_{ijk} = \mathrm{E}(U_{ijk})$, $k = 1, ..., p$. We then characterize the effects of ACGs and PCPs on the probability of any use of service k by letting $\pi_{ijk} = \mathrm{logit}^{-1}\left(\eta_{ijk}\right) = \exp\left(\eta_{ijk}\right) / \left\{1 + \exp\left(\eta_{ijk}\right)\right\}$, where

$$\eta_{ijk} = \boldsymbol{\alpha}'_k \mathbf{B}(x^a_{h(ij)k}) + \phi^a_{h(ij)k} + \lambda^a_{ik}.$$

Here, $h = 1, ..., q$ represent ACGs; $h(ij)$ indicates that h is a function of i and j (because each patient is assigned to one and only one ACG); $x^a_{h(ij)k}$ is the prior rank of ACG h with respect to the percentage of members using service k in the larger POS plan sample (as described in Section 2.3); $\mathbf{B}(\cdot) = \{B_0(\cdot), B_1(\cdot), ..., B_5(\cdot)\}'$ is a B-spline sequence for a piecewise cubic polynomial with two equally-spaced interior knots, each with two

continuous derivatives (de Boor 1978); $\boldsymbol{\alpha}_k = (\alpha_{k0}, \alpha_{k1}, \ldots, \alpha_{k5})'$ is a corresponding parameter vector; $\phi^a_{h(ij)k}$ is the "extra-rank" effect of ACG h; and λ^a_{ik} is the effect of PCP i. By "extra-rank" effect we mean that $\phi^a_{h(ij)k}$ is the effect of ACG h that cannot be explained by the B-spline expansion of its prior rank.

The interpretation of the parameter vector, $\boldsymbol{\Omega}$ is not affected by the variable transformation, thus the elements of $\boldsymbol{\Omega}$ represent conditional log odds ratios, log ratios of conditional odds ratios, and so on, as in the untransformed log-linear model (Liang, Zeger, and Qaqish 1992). Note that if $p = 2$, $\boldsymbol{\Omega}$ consists of a single element, representing the log odds ratio for any use of the two services.

3.1.2 Part Two: General Multivariate Log-Normal Model

Let $\mathbf{Y}^*_{ij} = \left(Y^*_{ij1}, \ldots, Y^*_{ijp}\right)' \backsim \mathrm{N}_p\left(\boldsymbol{\mu}_{ij}, \boldsymbol{\Sigma}_i\right)$, where $\boldsymbol{\mu}_{ij} = (\mu_{ij1} \ldots, \mu_{ijp})'$ and

$$\mu_{ijk} = \boldsymbol{\beta}'_k \mathbf{B}(x^b_{h(ij)k}) + \phi^b_{h(ij)k} + \lambda^b_{ik} + \boldsymbol{\gamma}'_k \mathbf{u}_{ij,-k}.$$

Here, $x^b_{h(ij)k}$ is the prior rank of ACG h with respect to mean charges among users of service k in the larger POS plan sample (as described in Section 2.3); $\mathbf{B}(\cdot) = \{B_0(\cdot), B_1(\cdot), ..., B_5(\cdot)\}'$ is a B-spline sequence for a piecewise cubic polynomial with two equally-spaced interior knots, each with two continuous derivatives; $\boldsymbol{\beta}_k = (\beta_{k0}, \beta_{k1}, \ldots, \beta_{k5})'$ is a corresponding parameter vector; $\phi^b_{h(ij)k}$ is the extra-rank effect of ACG h; λ^b_{ik} is the effect of PCP i; and $\mathbf{u}_{ij,-k} = (u_{ij1}, \ldots, u_{ij,k-1}, u_{ij,k+1}, \ldots, u_{ijp})'$, is the vector, $\mathbf{u}_{ij}$ less element u_{ijk}. The $p-1$ elements of $\boldsymbol{\gamma}_k$ represent the change in mean of log of positive charges for service k given any use of each of the other services.

Covariance matrix, $\boldsymbol{\Sigma}_i$, is modeled using a "separation strategy" (Barnard, McCulloch, and Meng 2000), letting $\boldsymbol{\Sigma}_i = \mathrm{diag}\,(\boldsymbol{\sigma}_i)\,\mathbf{R}\,\mathrm{diag}\,(\boldsymbol{\sigma}_i)$, where $\boldsymbol{\sigma}^2_i = \left(\sigma^2_{i1}, ..., \sigma^2_{ip}\right)'$ is PCP-specific. Because $C_{ijk}>0$ is assumed to have a log-normal distribution, the posterior expectation of C_{ijk} depends on both the expectation and variance of Y^*_{ijk}. Thus, by allowing the variance of Y^*_{ijk} to be PCP-specific, each PCP's effect on charges can be more accurately estimated. Note that correlation matrix, $\mathbf{R}$ contains information about within-patient correlations between log of positive charges for different services and that those correlations are assumed to be uniform across PCPs. Also note that $\mathbf{R}$ is assumed to be independent of $\mathbf{U}$, an assumption that may be inappropriate for some applications involving $p > 2$.

3.2 Prior Distributions

PCP Regression Effects. Let $\boldsymbol{\lambda}_i = \left(\boldsymbol{\lambda}^{a\prime}_i, \boldsymbol{\lambda}^{b\prime}_i\right)' \sim \mathrm{N}\,(\mathbf{0}, \mathbf{D})$, identically and independently for PCPs $i = 1, \ldots, m$, where $\boldsymbol{\lambda}^a_i = (\lambda^a_{i1}, \ldots, \lambda^a_{ip})'$ and $\boldsymbol{\lambda}^b_i = (\lambda^b_{i1}, \ldots, \lambda^b_{ip})'$ are regression effects from parts one and two of the likelihood, respectively.

ACG Extra-Rank Effects. Let $\boldsymbol{\phi}_h^a \backsim \mathrm{N}(\mathbf{0}, \mathbf{T}_a)$ and $\boldsymbol{\phi}_h^b \backsim \mathrm{N}(\mathbf{0}, \mathbf{T}_b)$, identically and independently for ACGs $h = 1, \ldots, q$, where $\boldsymbol{\phi}_h^a = (\phi_{h1}^a, \ldots, \phi_{hp}^a)'$ and $\boldsymbol{\phi}_h^b = (\phi_{h1}^b, \ldots, \phi_{hp}^b)'$ are regression effects from parts one and two of the likelihood, respectively.

PCP-Specific Variances, $\boldsymbol{\sigma}_i^2 = \left(\sigma_{i1}^2, ..., \sigma_{ip}^2\right)'$. For $k = 1, \ldots, p$, let $\sigma_{ik}^2 \sim \mathrm{IG}\left(n_\mathrm{o}/2, n_\mathrm{o}\sigma_{\mathrm{o}k}^2/2\right)$ identically and independently for each $i = 1, \ldots, m$, where $\mathrm{IG}(.,.)$ represents the inverse gamma distribution. Thus, the prior mean of σ_{ik}^2 is $\{n_\mathrm{o}/(n_\mathrm{o}-2)\}\,\sigma_{\mathrm{o}k}^2$, $n_\mathrm{o} > 2$. The value of n_o is discussed in Section 4.1.

Correlation Matrix, $\mathbf{R}$. Assume that the prior distribution of $\mathbf{R}$ is uniform over the space of correlation matrices of dimension p (Barnard, McCulloch, and Meng 2000).

Other likelihood parameters. Assume vague priors: For $k = 1, \ldots, p$, let $\boldsymbol{\alpha}_k \sim \mathrm{N}_6\left(\boldsymbol{\epsilon}_k^\alpha, \mathrm{diag}\left\{10^6\right\}\right)$, $\boldsymbol{\beta}_k \sim \mathrm{N}_6\left(\boldsymbol{\epsilon}_k^\beta, \mathrm{diag}\left\{10^6\right\}\right)$, and $\boldsymbol{\gamma}_k \sim \mathrm{N}_{(p-1)}\left(\boldsymbol{\epsilon}_k^\gamma, \mathrm{diag}\left\{10^6\right\}\right)$, where $\boldsymbol{\epsilon}_k^\alpha = (\epsilon_{k0}^\alpha, \ldots, \epsilon_{k5}^\alpha)'$, $\boldsymbol{\epsilon}_k^\beta = (\epsilon_{k0}^\beta, \ldots, \epsilon_{k5}^\beta)'$, and $\boldsymbol{\epsilon}_k^\gamma = (\epsilon_{k1}^\gamma, \ldots, \epsilon_{k,k-1}^\gamma, \epsilon_{k,k+1}^\gamma, \ldots, \epsilon_{k,p}^\gamma)'$ are approximate estimates of $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$, and $\boldsymbol{\gamma}_k$ from a non-Bayesian analysis of the same data. Let $\boldsymbol{\Omega} \sim \mathrm{N}_{(p^2-p-1)}\left(\mathbf{0}, \mathrm{diag}\left\{10^6\right\}\right)$.

Lastly, assume that the prior distributions are mutually independent:

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \mathbf{R}) = p(\boldsymbol{\alpha})p(\boldsymbol{\beta})p(\boldsymbol{\Omega})p(\boldsymbol{\gamma})p(\boldsymbol{\lambda})p(\boldsymbol{\phi})p(\boldsymbol{\sigma}^2)p(\mathbf{R}).$$

3.3 Hyperprior Distributions

PCP Effects Covariance Matrix. Let $\mathbf{D} \sim \mathrm{IW}(2p\mathbf{D}_\mathrm{o}, 2p)$, where $\mathrm{IW}(.,.)$ represents the inverse Wishart distribution and $\mathbf{D}_\mathrm{o}$ is diagonal, with diagonal elements equal to rough estimates of the variances of the elements of $\boldsymbol{\lambda}$.

ACG Extra-Rank Effects Variances, $\mathbf{T}$. Let $\mathbf{T} = \mathrm{diag}(\mathbf{T}_a, \mathbf{T}_b) = \mathrm{diag}\left\{(\tau_1^a, .., \tau_p^a, \tau_1^b, ..., \tau_p^b)'\right\}$, where $\tau_k^l \sim \mathrm{IG}\left(.5, .5\xi_k^l\right)$, for $l = a, b$ and $k = 1, \ldots, p$, and the ξ_k^l are rough estimates of the τ_k^l. $\mathbf{T}$ is taken as diagonal because there is no conceptual basis for suspecting a within-ACG association between the $2p$ extra-rank effects.

Prior Means of Variances of Log of Positive Charges, $\boldsymbol{\sigma}_\mathrm{o}^2 = \left(\sigma_{\mathrm{o}1}^2, ..., \sigma_{\mathrm{o}p}^2\right)'$. Assume $\sigma_{\mathrm{o}k}^2 \sim \mathrm{G}\left(.5, .5\upsilon_k^{-2}\right)$, for $k = 1, \ldots, p$, where where $\mathrm{G}(.,.)$ represents the gamma distribution and the υ_k^2 are set equal to rough estimates of the $\sigma_{\mathrm{o}k}^2$.

Lastly, assume $p(\mathbf{D}, \mathbf{T}, \boldsymbol{\sigma}_\mathrm{o}^2) = p(\mathbf{D})p(\mathbf{T})p(\boldsymbol{\sigma}_\mathrm{o}^2)$.

4 MODEL ESTIMATION

4.1 Gibbs Sampler: Design

The model presented in Section 3 can be estimated using a Gibbs sampler comprised of the 12 conditional distributions listed in the Appendix. Seven of the conditionals have closed forms and can be directly simulated. One of these, $f\left(\mathbf{Y}^*_{ij(\mathbf{u}=\mathbf{0})} \middle| \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}^b_i, \boldsymbol{\phi}^b, \mathbf{R}, \boldsymbol{\sigma}^2_i, \mathbf{y}_{ij(\mathbf{u}=\mathbf{1})}\right)$, having a multivariate normal form, corresponds to the data augmentation step, where $\mathbf{Y}^*_{ij(\mathbf{u}=\mathbf{1})}$ and $\mathbf{Y}^*_{ij(\mathbf{u}=\mathbf{0})}$ are the subvectors of $\mathbf{Y}^*_{ij} = \left(Y^*_{ij1}, \ldots, Y^*_{ijp}\right)'$ corresponding to services that were used and not used, respectively. Note that no augmentation is necessary for an observation wherein $\mathbf{u}_{ij} = \mathbf{0}$, since such an observation contributes no information to part two of the likelihood.

Five of the conditionals have non-conjugate priors and thus, do not have closed forms, but can be simulated using the rejection sampling approach described by Zeger and Karim (1991). That approach uses a multivariate normal envelope, with mean equal to the conditional posterior mode of the simulated parameter and variance equal to the product of a constant (typically, 2) times the inverse Fisher information of the log conditional posterior density. The mode and Fisher information for three of the five conditionals without closed forms, $f\left(\boldsymbol{\alpha}, \boldsymbol{\Omega} \middle| \boldsymbol{\phi}^a, \boldsymbol{\lambda}^a, \boldsymbol{\epsilon}^\alpha, \boldsymbol{\epsilon}^\Omega, \mathbf{u}\right)$, $f\left(\boldsymbol{\phi}^a_h \middle| \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{\lambda}^a, \mathbf{T}_a, \mathbf{u}\right)$, and $f\left(\boldsymbol{\lambda}^a_i \middle| \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{\phi}^a, \boldsymbol{\lambda}^b_i, \mathbf{D}, \mathbf{u}_i\right)$, can be readily obtained using a slight modification of the approach described by Fitzmaurice and Laird (1993) for deriving the maximum likelihood estimates and Fisher information for $\boldsymbol{\alpha}, \boldsymbol{\Omega}$. The modification involves adding the first and second derivatives of the log of the prior distribution to the respective derivatives of the log likelihood before solving the score equations and taking the expectation of the Hessian matrix. Since the priors of all three distributions are multivariate normal, the necessary computations are straightforward.

The fourth conditional without a closed form is $f\left(\boldsymbol{\sigma}^2_i \middle| \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}^b, \boldsymbol{\lambda}^b_i, \mathbf{R}, n_\mathrm{o}, \boldsymbol{\sigma}^2_\mathrm{o}, \mathbf{y}^*_i\right)$, for $i = 1, ..., m$, which can be expressed as the product of a constant that does not depend on $\boldsymbol{\sigma}^2_i$ and

$$
\begin{aligned}
p\left(\boldsymbol{\sigma}^2_i\right) &= \left\{\prod_{k=1}^{p}\left(\sigma^2_{ik}\right)\right\}^{-\left(n_\mathrm{o}+n^+_i-p+3\right)/2} \\
&\quad \times \exp\left[-\frac{1}{2}\,\mathrm{tr}\left\{n_\mathrm{o}\,\mathrm{diag}\left(\boldsymbol{\sigma}^2_\mathrm{o} \odot \boldsymbol{\sigma}_i^{-2}\right) + \mathbf{S}_i\,\mathrm{diag}\left(\boldsymbol{\sigma}_i^{-1}\right)\mathbf{R}^{-1}\,\mathrm{diag}\left(\boldsymbol{\sigma}_i^{-1}\right)\right\}\right],
\end{aligned}
$$

where $\mathbf{S}_i = \sum_{j=1}^{n_i}\left\{I\left(\mathbf{1}'\mathbf{u}_{ij} > 0\right)\left(\mathbf{y}^*_{ij} - \boldsymbol{\mu}_{ij}\right)\left(\mathbf{y}^*_{ij} - \boldsymbol{\mu}_{ij}\right)'\right\}$, $\mathbf{y}^*_{ij}$ represents the augmented vector of log of positive charges, $n^+_i = \sum_{j=1}^{n_i} I\left(\mathbf{1}'\mathbf{u}_{ij} > 0\right)$, and $I\left(\mathbf{.}\right)$ represents the indicator function. Thus, n^+_i is the number of patients of PCP i with use of at least one service (and thus, a vector $\mathbf{y}^*_{ij}$ with no undefined elements). In deriving $p\left(\boldsymbol{\sigma}^2_i\right)$ we employed the Jacobian, $\left\{\prod_{k=1}^{p}\left(\sigma_{ik}\right)\right\}^{p-1}$ for the transformation $\boldsymbol{\Sigma}_i \longrightarrow \left(\boldsymbol{\sigma}^2_i, \mathbf{R}\right)$. The first

derivative and Hessian matrix of $\log p\left(\boldsymbol{\sigma}_i^2\right)$ are

$$\partial \log p\left(\boldsymbol{\sigma}_i^2\right)/\partial\left(\boldsymbol{\sigma}_i^2\right)=\frac{1}{2}\left\{-\left(n_{\mathrm{o}}+n_i^{+}-p+3\right)\boldsymbol{\sigma}_i^{-2}+n_{\mathrm{o}}\boldsymbol{\sigma}_{\mathrm{o}}^2\odot\boldsymbol{\sigma}_i^{-4}+\operatorname{diag}\left(\boldsymbol{\sigma}_i^{-3}\right)\left(\mathbf{S}_i\odot\mathbf{R}^{-1}\right)\boldsymbol{\sigma}_i^{-1}\right\},$$

and

$$\begin{aligned}\partial^2 \log p\left(\boldsymbol{\sigma}_i^2\right)/\partial\left(\boldsymbol{\sigma}_i^2\right)^2=&\frac{1}{2}\left(n_{\mathrm{o}}+n_i^{+}-p+3\right)\operatorname{diag}\left(\boldsymbol{\sigma}_i^{-4}\right)-n_{\mathrm{o}}\operatorname{diag}\left(\boldsymbol{\sigma}_{\mathrm{o}}^2\odot\boldsymbol{\sigma}_i^{-6}\right)\\&-\frac{3}{4}\operatorname{diag}\left\{\operatorname{diag}\left(\boldsymbol{\sigma}_i^{-5}\right)\left(\mathbf{S}_i\odot\mathbf{R}^{-1}\right)\boldsymbol{\sigma}_i^{-1}\right\}-\frac{1}{4}\operatorname{diag}\left(\boldsymbol{\sigma}_i^{-3}\right)\left(\mathbf{S}_i\odot\mathbf{R}^{-1}\right)\operatorname{diag}\left(\boldsymbol{\sigma}_i^{-3}\right).\end{aligned}$$

To implement rejection sampling, the conditional posterior mode of $\boldsymbol{\sigma}_i^2$ can be located using Newton's method and its asymptotic variance can be approximated as minus the inverse Hessian of $\log p\left(\boldsymbol{\sigma}_i^2\right)$ evaluated at the posterior mode.

From $p\left(\boldsymbol{\sigma}_i^2\right)$, it is apparent that the shrinkage of $\boldsymbol{\sigma}_i^2$ toward the prior mean, $\left\{n_{\mathrm{o}}/\left(n_{\mathrm{o}}-2\right)\right\}\boldsymbol{\sigma}_{\mathrm{o}}^2$, $n_{\mathrm{o}}>2$, increases with n_{o}/n_i^{+}. For the model estimation described in Section 4.2, we set $n_{\mathrm{o}}=50$, implying that the prior mean carried a weight roughly equivalent to 50 observations and that for a PCP with $n_i^{+}=50$, equal weight was given to $\mathbf{S}_i/n_i^{+}$ and $\boldsymbol{\sigma}_{\mathrm{o}}^2$ in estimating $\boldsymbol{\sigma}_i^2$.

The last conditional without a closed form, $f\left(\mathbf{R}\left|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\phi}^b,\boldsymbol{\lambda}^b,\boldsymbol{\sigma}_1^2,...,\boldsymbol{\sigma}_m^2,\mathbf{y}^*\right.\right)$ can be expressed as the product of a constant that does not depend on $\mathbf{R}$ and $p\left(\mathbf{R}\right)=\left|\mathbf{R}\right|^{-\frac{n^{+}}{2}}\exp\left\{-\frac{1}{2}\operatorname{tr}\left(\mathbf{R}^{-1}\mathbf{A}\right)\right\}$,where $\mathbf{A}=\sum_{i=1}^{m}\left\{\operatorname{diag}\left(\boldsymbol{\sigma}_i^{-1}\right)\mathbf{S}_i\operatorname{diag}\left(\boldsymbol{\sigma}_i^{-1}\right)\right\}$ and $n^{+}=\sum_{i=1}^{m}n_i^{+}$. The correlation coefficients in $\mathbf{R}$ can be sampled one at a time, conditioning on the most recent sampled values of the other correlation coefficients and assuring the positive definiteness of $\mathbf{R}$, as suggested by Barnard, McCulloch, and Meng (2000). The first derivative of and expected Hessian matrix of $\log p\left(\mathbf{R}\right)$ are

$$\partial \log p\left(\mathbf{R}\right)/\partial \operatorname{vech}\mathbf{R}=\frac{1}{2}\mathbf{H}'\left(\mathbf{R}^{-1}\otimes\mathbf{R}^{-1}\right)\operatorname{vec}\left(\mathbf{A}-n^{+}\mathbf{R}\right),$$

and

$$\mathrm{E}\left\{\partial^2 \log p\left(\mathbf{R}\right)/\partial\left(\operatorname{vech}\mathbf{R}\right)^2\right\}=-\frac{n^{+}}{2}\mathbf{H}'\left(\mathbf{R}^{-1}\otimes\mathbf{R}^{-1}\right)\mathbf{H},$$

where $\operatorname{vech}\mathbf{R}$ denotes the vector that is obtained from $\operatorname{vec}\mathbf{R}$ by eliminating the supradiagonal (redundant) elements of $\mathbf{R}$ and $\mathbf{H}$ is defined by $\operatorname{vec}\mathbf{R}=\mathbf{H}\operatorname{vech}\mathbf{R}$ (Magnus and Neudecker 1999, pp. 316-318).

4.2 Gibbs Sampler: Implementation

A program for the Gibbs sampler was written in SAS Interactive Matrix Language (SAS Institute 1999) and implemented for the case of two services ($p = 2$), primary care and specialty care, where primary care and specialty care refer to outpatient evaluation and management provided by PCPs and medical and non-ophthalmologic surgical specialists, respectively. To select starting points, the posterior distribution of the model parameters was approximated using non-Bayesian methods. Then for each model fitted, three parallel chains were initiated from systematically selected, over-dispersed locations in this approximate target distribution, as suggested by Carlin and Louis (1996, p.196). Convergence was monitored using potential scale reductions (PSRs) as proposed by Gelman and Rubin (1992). PSRs for all model parameters fell below 1.1 within the first 1,000 iterations of all chains, each of which was then run for an additional 5,000 iterations. The initial 1,000 iterations from each chain were discarded and the final 5,000 retained. Posterior estimates for all models were based on 15,000 retained draws combined from three chains. Each chain of 6,000 iterations required approximately 40 hours of computing time on a Dell Precision 340 workstation with a 2.4 Gigahertz Pentium 4 processor.

Model fit was assessed by comparing observed annual rates of any use and mean annual charges for each service to posterior predictions, for groups of patients conditional on their PCP and ACG assignments, and marginally, for all patients taken together. For the assessment of marginal fit, 300 samples of the parameters were drawn systematically from the 15,000 retained iterations from the final model and used to simulate 300 samples of annual charges for each of the 3,308 patients, for each service. Medians and 95% credible intervals were computed for each of the quantiles of the simulated distributions. The medians of simulated quantiles were found to closely track the quantiles of the observed distributions, for both services. For the full patient sample, the observed and simulated annual rates of service use were 80.6% and 80.7%, respectively for primary care, and 30.4% and 30.4%, respectively for specialty care; and the observed and simulated mean annual per patient charges were $122 and $125, respectively for primary care, and $52 and $52, respectively for specialty care. (Note that mean annual charges include non-users of a service.)

4.3 Estimation of Posterior Deviations

The risk-adjusted effects of individual PCPs on measures of service utilization can be represented by functions of the model parameters averaged over each PCP's actual patient sample, here referred to as "deviations". To define a deviation, we introduce terms similar to those proposed by Normand, Glickman, and Gatsonis (1997). Each term is with respect to a PCP's patient sample: A "standardized" mean value refers to the mean of patients' expected values given the effects of ACGs *only*, while a "predicted" mean

value refers to the mean of patients' expected values given the effects of ACGs and the PCP. A "predicted deviation" is the difference between the predicted and standardized mean values. An "observed deviation" is the difference between the mean of patients' observed values and the standardized mean value. In each of these definitions, "value" refers the value of a utilization measure, such as the annual probability of any use of, the log of positive annual charges for, or annual charges for a service type. The observed deviations are essentially risk-adjusted, fixed effects, while the predicted deviations incorporate both risk adjustment and the shrinkage due to the hierarchical model assumptions. Because the deviations and standardized and predicted means are functions of the model parameters, their posterior distributions can be accurately estimated using the Gibbs sampler output.

For example, for service type k, for PCP i, the predicted mean annual charge is

$$\begin{aligned}\theta^P_{ik} &= \frac{1}{n_i}\sum_{j=1}^{n_i}\Big[\text{logit}^{-1}\left\{\boldsymbol{\alpha}'_k\mathbf{B}(x^a_{h(ij)k}) + \phi^a_{h(ij)k} + \lambda^a_{ik}\right\} \\ &\quad \times \exp\left\{\boldsymbol{\beta}'_k\mathbf{B}(x^b_{h(ij)k}) + \boldsymbol{\gamma}'_k\mathbf{u}_{ij,-k} + \phi^b_{h(ij)k} + \lambda^b_{ik} + \sigma^2_{ik}/2\right\}\Big],\end{aligned}$$

standardized mean annual charge is

$$\begin{aligned}\theta^S_{ik} &= \frac{1}{n_i}\sum_{j=1}^{n_i}\Big[\text{logit}^{-1}\left\{\boldsymbol{\alpha}'_k\mathbf{B}(x^a_{h(ij)k}) + \phi^a_{h(ij)k}\right\} \\ &\quad \times \exp\left\{\boldsymbol{\beta}'_k\mathbf{B}(x^b_{h(ij)k}) + \boldsymbol{\gamma}'_k\mathbf{u}_{ij,-k} + \phi^b_{h(ij)k} + \sigma^2_{ok}/2\right\}\Big],\end{aligned}$$

predicted deviation is $\delta^P_{ik} = \theta^P_{ik} - \theta^S_{ik}$, and observed deviation is $\delta^O_{ik} = \frac{1}{n_i}\sum_{j=1}^{n_i} C_{ijk} - \theta^S_{ik}$. The posterior distribution of each of these quantities can be accurately estimated using the Gibbs output. Note that θ^P_{ik} is a function of the PCP-specific variance, σ^2_{ik}, while θ^S_{ik} is a function the prior mean of PCP-specific variances, $\{n_o/(n_o-2)\}\sigma^2_{ok}$.

4.4 Influence of the Prior on D

Simulative studies have suggested that the inverse Wishart prior, even with fully minimized degrees of freedom, can significantly influence the posterior distribution of a variance matrix, $\mathbf{D}$, especially if the prior scale matrix divided by the prior degrees of freedom is far from the true mean of $\mathbf{D}$ (Natarajan and Kass, 2000). We were particularly interested in whether the inverse Wishart prior had substantially influenced the shape and orientation of our posterior estimate of $\mathbf{D}$, since correlations between PCP effects were a focus of study. To this end, we decomposed the posterior mean of $\mathbf{D}$ into its eigenvalues and normalized eigenvector matrix, $\widehat{\mathbf{E}}$, and then compared the distributions of the diagonal elements of $\widehat{\mathbf{E}}'\mathbf{D}\widehat{\mathbf{E}}$ for samples from the prior

and posterior distributions. The diagonal elements of $\widehat{\mathbf{E}}'\mathbf{D}\widehat{\mathbf{E}}$ represent the size of $\mathbf{D}$ along axes oriented by the eigenvectors of its posterior mean. To simulate the prior distribution of $\mathbf{D}$, we drew 15,000 samples from $\mathbf{D}^* \backsim \text{IW}\,(4, 4\mathbf{D}_\text{o})$, letting $\mathbf{D}_\text{o} = \text{diag}\left(\{.12, .18, .029, .080\}'\right)$, our prior rough approximation $\mathbf{D}$.

5 RESULTS

5.1 Within-Patient Associations

Table 1 displays posterior means and credible intervals for within-patient associations between the four response variables, adjusted for ACG and PCP effects. Only the association between U_1 and U_2, any use of primary care and any use of specialty care is statistically significant. The odds ratio, $\exp(\omega)$ is estimated to be .57(95% CI: .45; .73), indicating that the probability that a patient had visited a specialist was significantly reduced if that patient had visited his or her PCP at least once during the year, regardless of who that PCP was.

With regression parameter $\boldsymbol{\gamma}_k$ included in the likelihood, predictions regarding $Y_k\,|U_k = 1$ are conditioned on the observed value of $\mathbf{U}_{-k}$. However, we would like to make predictions of $Y_k\,|U_k = 1$ that are marginal with respect to $\mathbf{U}_{-k}$. Thus because neither γ_1 or γ_2 was statistically significant, we re-estimated the model excluding these parameters. This change did not significantly affect the posterior estimate of any other model parameter hence below, we only present results for the model without γ_1 and γ_2.

5.2 Risk Adjustment Using ACGs

Figure 2 shows posterior estimates of $P(U_2 = 1)$, the probability of any use of specialty care, and $E\,(C_2\,|U_2 = 1\,)$, the expected charges for specialty care given any use of it, adjusted for PCP effects, plotted against ACG. ACGs are ordered by their *prior* ranks to demonstrate the fit of the regression splines. (Note that prior ranks in the two plots correspond to different ACGs, since ranks on percentage of users and charges among users were assigned separately.) The estimated posterior means and 95% credible intervals incorporate both the prior rank effects, $\boldsymbol{\alpha}_2'\mathbf{B}(x_{h2}^a)$ and $\boldsymbol{\beta}_2'\mathbf{B}(x_{h2}^b)$, and the extra-rank effects, ϕ_{h2}^a and ϕ_{h2}^b, and thus, represent a compromise between the prior rank effects and the observed data. As expected, shrinkage toward the fitted splines is greater among ACGs with smaller sample sizes. Analogous plots for $P(U_1 = 1)$ and $E\,(C_1\,|U_1 = 1\,)$ are not shown, but demonstrate similar phenomena.

5.3 PCP Regression and Variance Effects

Table 2 represents the posterior estimate of $\mathbf{D}$, the covariance matrix of PCP risk-adjusted regression effects in terms of standard deviations and correlations. The correlation matrix reveals three important findings: First, PCPs that were more likely to see each of their patients at least once during the year had a lower rate of specialist use by their patients (estimated correlation: $-.40$; 95% CI: $-.71, -.008$). Second, PCPs that provided more services to patients that they saw also had a lower rate of specialist use (estimated correlation: $-.53$; 95% CI: $-.77, -.21$). And third, PCPs that were more likely to see their patients at least once during the year provided more services to patients that they saw (estimated correlation: .45; 95% CI: $.086, .72$).

Figure 3 displays estimated means of individual PCPs' deviations on two measures, the annual probability of any use of service and the log of positive annual charges given use, for primary care and specialty care. The distributions of the deviations reflect the statistically significant positive correlation between λ_1^a and λ_1^b and lack of a substantial correlation between λ_2^a and λ_2^b, shown in Table 2. The shrinkage from observed to predicted deviations is not precisely toward $\mathbf{0}$ (the prior mean of $\boldsymbol{\lambda}$) because $\mathbf{D}$ is not diagonal. (Analogous deviations from a model estimated with off-diagonal elements of $\mathbf{D}$ fixed at zero did shrink precisely toward $\mathbf{0}$ [results not shown].) The greatest overall shrinkage occurs among deviations of log of positive charges for specialty care, evidently because this measure was informed by the smallest number of observations (as noted in section 4.2.)

Figure 4 compares estimated posterior means of PCP-specific variances of log of positive annual charges, σ_{ik}^2, $i = 1, ..., 50$, $k = 1, 2$, against PCP-specific means of squared residuals from the regression of log of positive annual charges on ACG and PCP effects. (Only residuals for observed $\mathbf{y}$, not augmented $\mathbf{y}^*$ were used in computing the latter quantities.) The number of observations on log of positive charges per PCP ranged from 20 to 124 for primary care and from 5 to 53 for specialty care hence, the greater overall shrinkage in Figure 4b compared to 4a.

5.4 Deviations of Charges

Figure 5 displays estimated means and 95% credible intervals for predicted deviations, δ_{ik}^P, $k = 1, 2$, of annual charges for (a) primary care and (b) specialty care for all PCPs, ordered by mean of δ_{ik}^P. Examining the credible intervals in Figure 5a reveals that four PCPs provided significantly less and seven provided significantly more primary care than expected after risk-adjustment. In Figure 6, the estimated means of PCPs' predicted and observed deviations of annual charges for primary care and specialty care are shown to have an inverse relationship. Comparison of Figures 5a and 6 reveals that the estimated means of predicted

deviations of charges for specialty care are positive for all four PCPs identified in Figure 5a to have low primary care charges and negative for six of seven noted to have high primary care charges.

In Figure 5b, none of the 95% credible intervals for predicted deviations of charges for specialty care excludes zero. This appears to be due less to a lack of precision of the estimates, and more to the narrowness of their range, from −$12 to $18. In contrast, the range of the estimated means of predicted deviations of charges for primary care is from −$37 to $40.

5.5 Influence of the Prior on D

Figure 7 compares the empirical distributions of the diagonal elements of $\widehat{\mathbf{E}}'\mathbf{D}^*\widehat{\mathbf{E}}$ and $\widehat{\mathbf{E}}'\mathbf{D}\widehat{\mathbf{E}}$, based on 15,000 draws, as described in Section 4.4. The prior densities of the four diagonal elements are sufficiently flat that it does not appear that the prior distribution of $\mathbf{D}$ was substantially informative with respect to the eigenvalues and eigenvectors, or shape and orientation of the posterior mean of $\mathbf{D}$.

6 DISCUSSION

When profiling providers' effects on multiple responses, fitting a multivariate model rather than a series of univariate models can yield informational gains in the form of insights about provider- and patient-level associations between responses. For instance, we found that PCPs who were more likely to see their patients at least once during the year or provided more services to patients that they saw had a lower rate of specialist use by their patients. This suggests that some PCPs substituted their services for those of specialists while others may have intentionally or unintentionally encouraged the substitution of specialists' services for their own. Thus, the overall efficiency of a PCP's practice could not have been fairly assessed by examining utilization of primary care or specialty care in isolation.

In addition to informational gains, a multivariate model can yield efficiency gains, specifically, by increasing the precision of estimated regression coefficients (Zellner 1962). To evaluate this possibility, we repeated the estimation described in Section 4.2 excluding patient-level associations, by setting $\mathbf{\Omega} = \mathbf{0}$, $\mathbf{R} = \mathbf{I}$, and $\boldsymbol{\gamma} = \mathbf{0}$. Resultant estimates of regression parameters and their standard errors (not shown) were virtually identical to those reported in Section 5 for the model that included patient-level associations. A likely explanation for the absence of an efficiency gain was that the design matrices for the regression equations for primary care and specialty care were highly collinear (Zellner 1962). Each of the design matrices consisted of indicator variables for 50 PCPs and 40 ACGs and 6 covariates resulting from the B-spline expansions of prior ACG ranks. For a given patient, the indicators for ACG and PCP were of course, identical across the regressions while the B-spline covariates were similar (because the prior ACG ranks on utilization of the two

services were similar).

The distinction drawn by the multivariate two-part model between the probability of any use of a service and the amount of charges given use facilitated findings of substantive importance. For instance, we found a statistically significant within-patient association between any use but not between the amount of use of the two services, and not between any use of one service and the amount of use of the other.

The risk-adjustment approach employed here allowed providers to be compared to an internal, as opposed to external standard, while taking advantage of prior information about the effects of risk categories on response variables. We consider the standardized mean values introduced in Section 4.3, to represent an internal standard because they are derived from the profiled sample. Alternatively, we might have developed an external standard by estimating effects of ACGs on responses using a separate, larger and more diverse patient sample (DeLong et al. 1997). We chose an internal standard so that PCPs would be compared to their peer group of family physicians with moderate-to-large case-loads of POS plan members. An advantage of this approach was that the marginal distribution of annual per-patient charges simulated using the fitted model closely matched the observed distribution, a result that would have been unlikely had an external standard been used. The challenge of estimating an internal standard in this application was that some of the ACGs were very infrequently represented in the profiled sample. Our remedy was to incorporate external information about the effects of ACGs in the form of prior ranks, and then to use those prior ranks to borrow strength across ACGs within the profiled sample. The result was an internal standard that was far more robust than could have been developed solely on the basis of information contained in the profiled sample.

7 APPENDIX

The model presented in Section 3 can be estimated using a Gibbs sampler comprised of the following 12 conditional distributions:

1. $f\left(\boldsymbol{\alpha}, \boldsymbol{\Omega} \left| \boldsymbol{\phi}^a, \boldsymbol{\lambda}^a, \boldsymbol{\epsilon}^\alpha, \boldsymbol{\epsilon}^\Omega, \mathbf{u} \right.\right)$.
2. $f\left(\boldsymbol{\phi}_h^a \left| \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{\lambda}^a, \mathbf{T}_a, \mathbf{u} \right.\right)$ for ACGs $h = 1, ..., q$.
3. $f\left(\boldsymbol{\lambda}_i^a \left| \boldsymbol{\alpha}, \boldsymbol{\Omega}, \boldsymbol{\phi}^a, \boldsymbol{\lambda}_i^b, \mathbf{D}, \mathbf{u}_i \right.\right)$ for PCPs $i = 1, ..., m$.
4. $f\left(\mathbf{Y}^*_{ij(\mathbf{u}=0)} \left| \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}_i^b, \boldsymbol{\phi}^b, \mathbf{R}, \boldsymbol{\sigma}_i^2, \mathbf{y}_{ij(\mathbf{u}=1)} \right.\right)$ for patients $j = 1, ..., n_i$ of PCPs $i = 1, ..., m$.
5. $f\left(\boldsymbol{\beta}, \boldsymbol{\gamma} \left| \boldsymbol{\phi}^b, \boldsymbol{\lambda}^b, \mathbf{R}, \boldsymbol{\sigma}_1^2, ..., \boldsymbol{\sigma}_m^2, \boldsymbol{\epsilon}^\beta, \boldsymbol{\epsilon}^\gamma, \mathbf{y}^* \right.\right)$.
6. $f\left(\boldsymbol{\phi}_h^b \left| \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}^b, \mathbf{R}, \boldsymbol{\sigma}_1^2, ..., \boldsymbol{\sigma}_m^2, \mathbf{T}_b, \mathbf{y}^* \right.\right)$ for ACGs $h = 1, ..., q$.
7. $f\left(\boldsymbol{\lambda}_i^b \left| \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}^b, \mathbf{R}, \boldsymbol{\sigma}_i^2, \boldsymbol{\lambda}_i^a, \mathbf{D}, \mathbf{y}_i^* \right.\right)$ for PCPs $i = 1, ..., m$.
8. $f\left(\boldsymbol{\sigma}_i^2 \left| \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}^b, \boldsymbol{\lambda}_i^b, \mathbf{R}, n_\mathrm{o}, \boldsymbol{\sigma}_\mathrm{o}^2, \mathbf{y}_i^* \right.\right)$ for PCPs $i = 1, ..., m$.
9. $f\left(\mathbf{R} \left| \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\phi}^b, \boldsymbol{\lambda}^b, \boldsymbol{\sigma}_1^2, ..., \boldsymbol{\sigma}_m^2, \mathbf{y}^* \right.\right)$.

10. $f\left(\sigma_{ok}^2 \left| \sigma_{1k}^2, ..., \sigma_{mk}^2, n_o, \upsilon_k^2 \right.\right)$, for services $k = 1, ..., p$.

11. $f\left(\mathbf{D} \left| \boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_m, \mathbf{D}_o \right.\right)$.

12. $f\left(\tau_k^l \left| \phi_{1k}^l, ..., \phi_{qk}^l, \xi_k^l \right.\right)$ for $l = a, b$ and $k = 1, \ldots, p$.

Seven of the distributions involve conjugate priors and take closed forms that can be directly simulated using multivariate normal (#4, 5, 6 and 7), gamma (#10), inverse Wishart (#11), and inverse gamma (#12) distributions. Inverse Wishart draws can be simulated using the method proposed by Odell and Feiveson (1966). The remaining five distributions have non-conjugate priors and thus, do not have closed forms, but can be simulated using rejection sampling, as outlined in Section 4.1

8 REFERENCES

Barnard, J., McCulloch, R., and Meng, X-L. (2000), "Modelling Covariance Matrices in Terms of Standard Deviations and Correlations, With Applications to Shrinkage", *Statistica Sinica*, 10, 1281-1311.

Blumberg, M. S. (1988), "Measuring Surgical Quality in Maryland: A Model," *Health Affairs*, Spring, 62-78.

Bodenheimer, T. (1996), "The HMO Backlash – Righteous or Reactionary," *Journal of the American Medical Association*, 335, 1601-1604.

de Boor, C. (1978), *A Practical Guide to Splines*, New York: Springer.

Braun, B. L., Fowles, J. B., Forrest, C. B., Kind, E. A., Foldes, S. S., and Weiner, J. P. (2003), "Which Enrollees Bypass Their Gatekeepers in a Point-of-Service Plan?," *Medical Care*, 41, 836-841.

Bronskill, S. E., Normand, S-L. T., Landrum, M. B., and Rosenheck, R. A. (2002), "Longitudinal Profiles of Health Care Providers," *Statistics in Medicine*, 21, 1067-1088.

Burgess, J. F., Christiansen, C. L., Michalak, S. E., and Morris, C. N. (2000) "Medical Profiling: Improving Standards and Risk Adjustments Using Hierarchical Models," *Journal of Health Economics*, 19, 291-309.

Burgess, J. F., Lourdes, V., and West, M. (2000), "Profiling Substance Abuse Provider Trends in Health Care Delivery Systems," *Health Services and Outcomes Research Methodology*, 1, 253-276.

Carey, T. S., Garrett, J., Jackman, A., McLaughlin, C., Fryer, J., Smucker, D. R., and the North Carolina Back Pain Project (1995), "The Outcomes and Costs of Care for Acute Back Pain Among Patients Seen by Primary Care Practitioners, Chiropractors, and Orthopedic Surgeons," *The New England Journal of Medicine*, 333, 913-917.

Carlin, B. P., and Louis, T. A. (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, London: Chapman & Hall.

Daniels, M. J., and Gatsonis, C. (1999), "Hierarchical Generalized Linear Models in the Analysis of Variations in Health Care Utilization", *Journal of the American Statistical Association*, 94, 29-42.

DeLong, E. R., Peterson, E. D., DeLong, D. M., Muhlbaier, L. H., Hackett, S., and Mark, D. B. (1997), "Comparing Risk-Adjustment Methods for Provider Profiling," *Statistics in Medicine*, 16, 2645-2664.

Diehr, P., Yanez, D., Ash, A., Hornbrook, M., and Yin, D. Y. (1999), "Methods for Analyzing Health Care Utilization and Costs," *Annual Review of Public Health*, 20, 125-144.

Duan, N., Manning, W. G., Morris, C. N., and Newhouse, J. P. (1983), "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Business and Economic Statistics*, 1, 115-126.

Fitzmaurice, G. M., and Laird, N. M. (1993), "A Likelihood-Based Method for Analyzing Longitudinal Binary Responses," *Biometrika*, 80, 141-151.

Forrest, C. B., Weiner, J. P., Fowles, J., Vogeli, C., Frick, K. D., Lemke, K. W., Starfield, B. (2001), "Self-Referral in Point-of-Service Health Plans," *Journal of the American Medical Association*, 285, 2223-2231.

Franks, P., Clancy, C. M., and Nutting, P. A. (1992), "Gatekeeping Revisited – Protecting Patients From Overtreatment," *The New England Journal of Medicine*, 327, 424-429.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Margianl Densities," *Journal of the American Statistical Association*, 85, 398-409.

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457-472.

Goldstein, H., and Spiegelhalter, D. (1996), "League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance," *Journal of the Royal Statistical Society*, Ser. A, 159, 385-444.

Green, J. , and Wintfeld, N. (1995) "Report Cards on Cardiac Surgeons: Assessing New York State's Approach," *The New England Journal of Medicine*, 332, 1229-1232.

Greenfield, S., Nelson, E. C., Zubkoff, M., Manning W., Rogers W., Karvitz, R. L., Keller, A., Tarlov, A. R., and Ware, J. (1992), "Variations in Resource Utilization Among Medical Specialties and Systems of Care: Results From the Medical Outcomes Study," *Journal of the American Medical Association*, 267, 1624-1630.

Grumbach, K., Selby, J. V., Damberg, C., Bindman A. B., Quesenberry, C., Truman, A., and Uratsu, C. (1999), "Resolving the Gatekeeper Conundrum: What Patients Value in Primary Care and Referrals to Specialists," *Journal of the American Medical Association*, 281, 261-266.

Halm E. A., Causino N., and Blumenthal D. (1997), "Is Gatekeeping Better Than Traditional Care?" *Journal of the American Medical Association*, 278, 1677-1681.

Hannan, E. L., Kilburn, H., O'Donnell, J. F., Lukacik, G., Shields, E. P. (1990), "Adult Open Heart

Surgery in New York State," *Journal of the American Medical Association*, 264, 2768-2774.

Health Services Research and Development Center at The Johns Hopkins University Bloomberg School of Public Health (2001), *The Johns Hopkins ACG Case-Mix System – Version 5.0*, Baltimore: The Johns Hopkins University.

Iezzoni, L. I. (1997), "The Risks of Risk Adjustment," *Journal of the American Medical Association*, 278, 1600-1607.

Kerr, E. A., Mittman, B. S., Hays, R. D., Siu, A. L., Leake, B., and Brook, R. H. (1995), "Managed Care and Capitation in California: How Do Physicians at Financial Risk Control Their Own Utilization?" *Annals of Internal Medicine*,123, 500-504.

Landrum, M. B., Bronskill, S. E., and Normand, S-L. T. (2000), "Analytical Methods for Constructing Cross-Sectional Profiles of Health Care Providers," *Health Services and Outcomes Research Methodology*, 1, 23-47.

Landrum, M. B., Normand, S-L. T., and Rosenheck, R. A. (2003), "Selection of Related Multivariate Means: Monitoring Psychiatric Care in the Department of Veterans Affairs," *Journal of the American Statistical Association*, 98, 7-16.

Liang, K-Y., Zeger, S. L., and Qaqish, B. (1992), "Multivariate Regression Analyses for Categorical Data," *Journal of the Royal Statistical Society*, Series B, 54, 3-40.

Liu, J., Louis, T. A., Pan, W., Ma, J. Z., Colliins, A. J. (2003), "Methods for Estimating and Interpreting Provider-Specific Standardized Mortality Ratios," *Health Services and Outcomes Research Methodology*, 4, 135-149.

Magnus, J. R., and Neudecker, H. (1999), *Matrix Differential Calculus With Applications in Statistics and Econometrics*, Revised Edition, New York: John Wiley & Sons.

Natarajan, R., and Kass, R. (2000), "Reference Bayesian Methods for Generalized Linear Models," *Journal of the American Statistical Association*, 95, 227-237.

Normand, S-L. T., Glickman, M.E., and Gatsonis, C.A. (1997), "Statistical Methods for Profiling Providers of Medical Care: Issues and Applications," *Journal of the American Statistical Association*, 92, 803-814.

Odell, P. L., and Feiveson, A. H. (1966), "A Numerical Procedure to Generate a Sample Covariance Matrix," *Journal of the American Statistical Association*, 61, 199-203.

Reid, R. J., MacWilliam, L., Verhulst, L., Roos, N., and Atkinson, M. (2001),"Performance of the ACG Case-Mix System in Two Canadian Provinces," *Medical Care*, 39, 86-99.

St. Peter, R. F., Reed, M. C., Kemper, P., Blumenthal, D. (1999), "Changes in the Scope of Care Provided by Primary Care Physicians", *The New England Journal of Medicine*, 341, 1980-1985.

Salem-Schatz, S., Moore, G., Rucker, M., and Pearson, S. D. (1994), "The Case for Case-Mix Adjustment in Practice Profiling: When Good Apples Look Bad," *Journal of the American Medical Association*, 272, 871-874.

SAS Institute Inc. (1999), *SAS/IML User's Guide*, Version 8, Cary, NC: SAS Institute, Inc.

Shahian, D. M., Normand, S. T., Torchiana, D. F., Lewis, S. M., Pastore, J. O., Kuntz, R. E., and Dreyer, P. I. (2001), "Cardiac Surgery Report Cards: Comprehensive Review and Statistical Critique," *Annals of Thoracic Surgery*, 72, 2155-2168.

Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528-540.

Weiner, J. P., Starfield, B. H., Steinwachs, D. M., and Mumford, L. M. (1991), "Development and Application of a Population-Oriented Measure of Ambulatory Care Case-Mix," *Medical Care*, 29, 452-472.

Williams, S. V., Nash, D. B., and Goldfarb, N., "Differences in Mortality From Coronary Artery Bypass Graft Surgery at Five Teaching Hospitals," *Journal of the American Medical Association*, 266, 810-815.

Zeger, S.L., and , Karim, M.R. (1991), "Generalized Linear Models With Random Effects: A Gibbs Sampling Approach," *Journal of the American Statistical Association*, 86, 79-86.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57, 348-368.

9 TABLES

Table 1. Within-Patient Associations Between Utilization of Primary Care (PC) and Specialty Care (Spc): Estimated Posterior Means and 95% Credible Intervals

Measure of Association	Mean	95% CI
$\exp(\omega)$: odds ratio for any use of PC given any use of Spc	.57	$(.45, .73)$
r_{21}: correlation of log of positive charges between PC and Spc	.033	$(-.035, .10)$
γ_1: change in log of positive charges for PC given any use of Spc	$-.033$	$(-.092, .024)$
γ_2: change in log of positive charges for Spc given any use of PC	$-.036$	$(-.18, .11)$

Table 2. Standard Deviations and Correlations of PCP Regression Effects: Estimated Posterior Means and 95% Credible Intervals

	Std Dev:	Correlation: Mean (CI)				
PCP Effect	Mean (CI)		λ_1^a	λ_2^a	λ_1^b	λ_2^b
λ_1^a: log odds any PC	$.40(.28, .53)$	λ_1^a	1	$(-.71, -.0083)$	$(.086, .72)$	$(-.60, .48)$
λ_2^a: log odds any Spc	$.39(.28, .52)$	λ_2^a	$-.40$	1	$(-.77, -.21)$	$(-.50, .53)$
λ_1^b: log(\$>0) for PC	$.17(.13, .21)$	λ_1^b	.45	$-.53$	1	$(-.33, .64)$
λ_2^b: log(\$>0) for Spc	$.11(.063, .16)$	λ_2^b	$-.078$	$-.002$	.19	1

Note: PC = primary care, Spc = specialty care

10 FIGURE CAPTIONS

Figure 1. Within-patient associations between response variables in a multivariate two-part model. The association between binary variables, U_1 and U_2 is represented as an odds ratio, between continuous variables $Y_1\,|U_1 = 1$ and $Y_2\,|U_2 = 1$ as a correlation, and between binary and continuous variables, U_k and $Y_l\,|U_l = 1$, $k \neq l$ as a regression of $Y_l\,|U_l = 1$ on U_k.

Figure 2. Posterior estimates of (a) $P(U_2 = 1)$, probability of any use of specialty care, and (b) $E\,(C_2\,|U_2 = 1\,)$, expected annual charges for specialty care given any use of it, adjusted for PCP effects, for ACGs ordered by their prior ranks. Line represents posterior mean of spline function of prior rank. Scored bars represent posterior means and 95% credible intervals of spline function of prior rank *plus* extra-rank effects, $\boldsymbol{\phi}$. Dots represent observed (a) proportion of patients with any use of specialty care and (b) mean charges among users of specialty care in the profiled sample. Selected ACGs demonstrate relationship between sample size, n, and shrinkage. (In {b}, n represents number of patient-users of specialty care.)

Figure 3. Deviation of log of positive annual charges v. deviation of annual probability of any use of service for (a) primary care and (b) specialty care, for 50 PCPs, each represented by a line. Dots represent estimated mean of observed deviation and squares represent estimated mean of predicted deviation.

Figure 4. PCP-specific variances of log of positive annual charges for (a) primary care and (b) specialty care for 50 PCPs, each represented by a line. Line intersects bottom axis at mean of estimated posterior variance and top axis at mean of squared residuals from regression of log of positive charges on ACG and PCP effects, using PCP's patient sample. (Scale of top axis is same as bottom axis.)

Figure 5. Deviations of annual charges in dollars, for (a) primary care and (b) specialty care for 50 PCPs ranked by mean predicted deviation. Dots represent estimated mean of observed deviation. Scored bars represent estimated mean and 95% credible interval of predicted deviation.

Figure 6. Deviations of annual charges in dollars, for primary care v. specialty care for 50 PCPs, each represented by a line. Dots represent estimated mean of observed deviation and squares represent estimated mean of predicted deviation.

Figure 7. Prior and posterior distributions of 15,000 draws of the four diagonal elements of $\widehat{\mathbf{E}}'\mathbf{D}\widehat{\mathbf{E}}$ ("variance components" of $\mathbf{D}$) where $\widehat{\mathbf{E}}$ is the matrix of normalized eigenvectors of the posterior mean of $\mathbf{D}$.

Figure 1

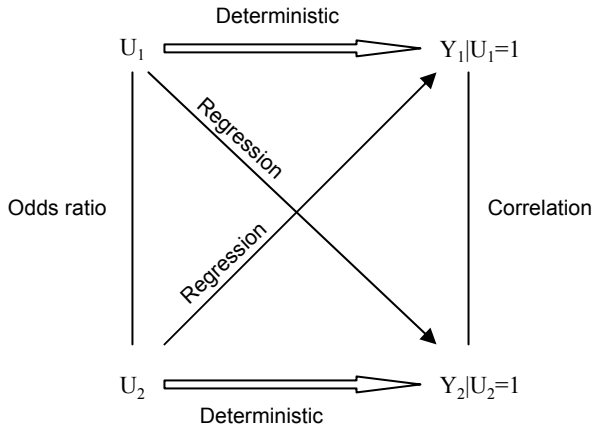

Figure 2

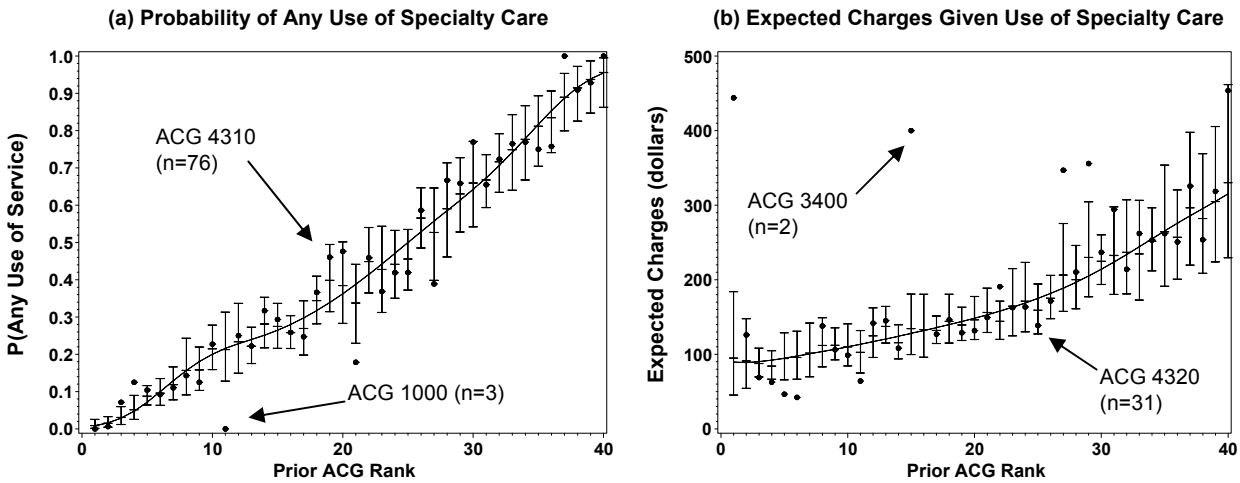

Figure 3

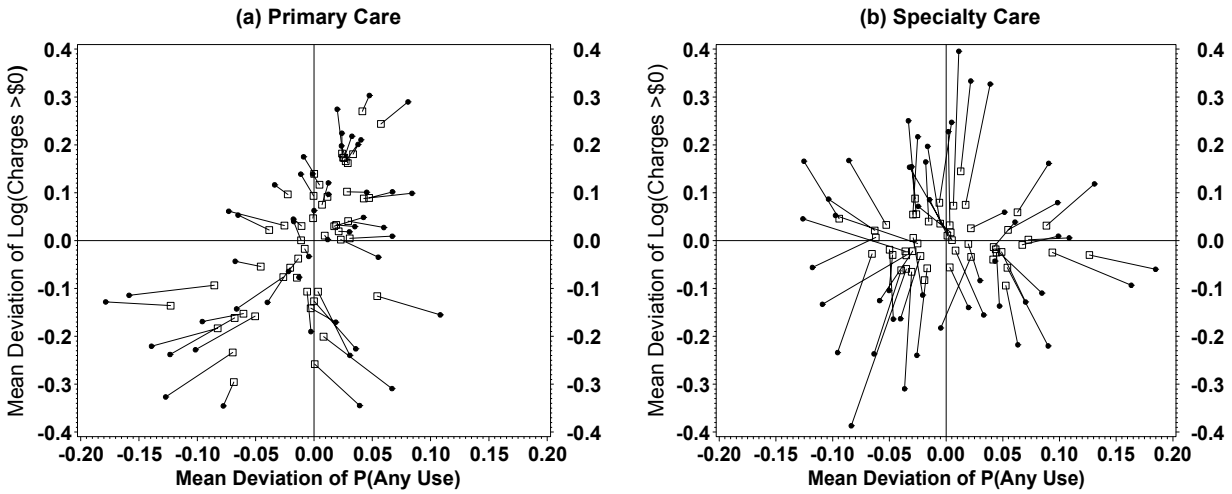

Figure 4

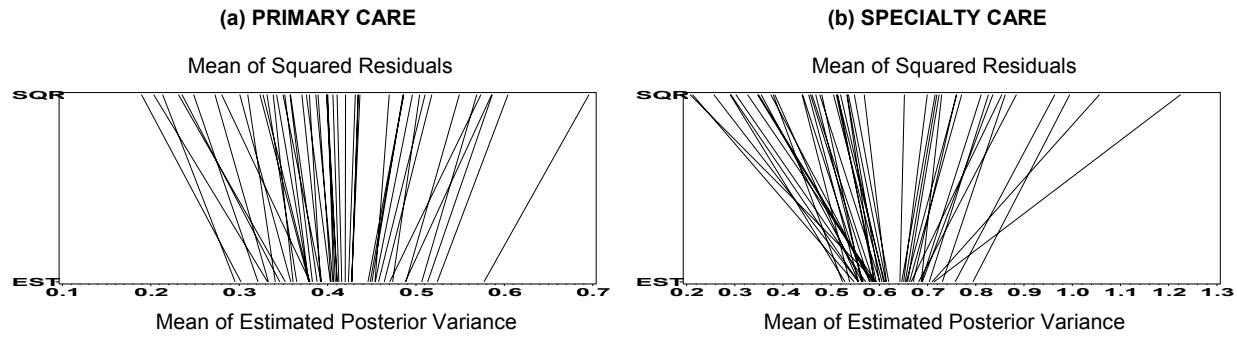

Figure 5

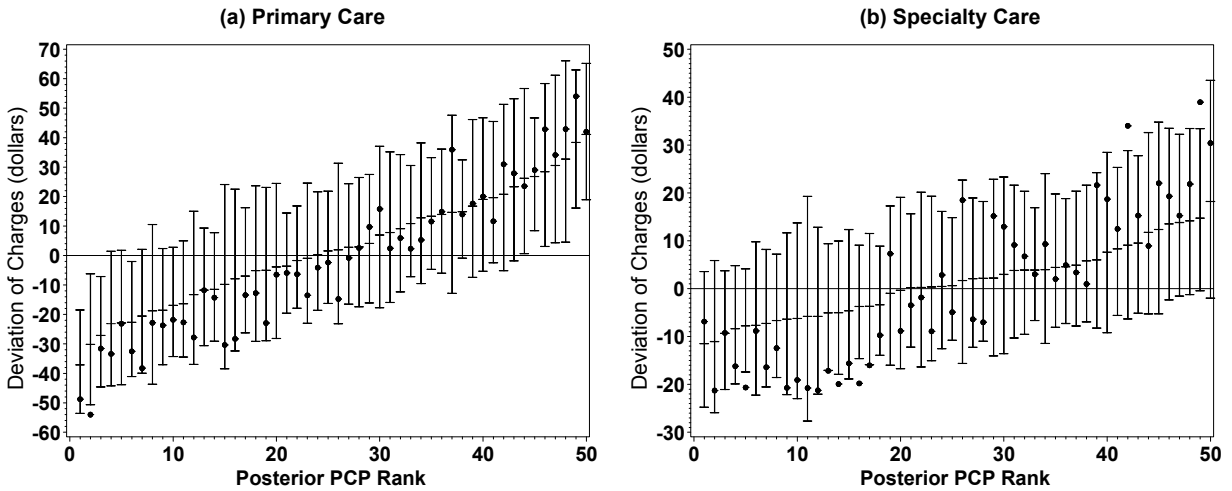

Figure 6

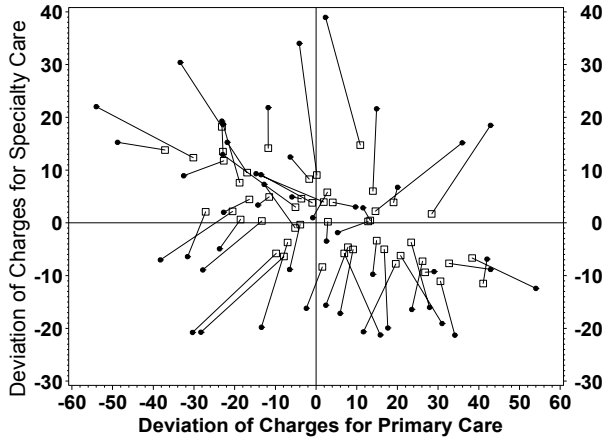

Figure 7