Published in final edited form as:

J Health Econ. 2010 January ; 29(1): 110–123. doi:10.1016/j.jhealeco.2009.11.010.

A Flexible Two-Part Random Effects Model for Correlated Medical Costs

Lei Liu,

Division of Biostatistics and Epidemiology, Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22908-0717, U.S.A

Mark E. Cowen,

Quality Institute, St. Joseph Mercy Health System, Ann Arbor, MI 48105, U.S.A

Robert L. Strawderman, and

Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY 14853, U.S.A

Ya-Chen T. Shih

Section of Health Service Research, Department of Biostatistics, Division of Quantitative Sciences, M. D. Anderson Cancer Center, University of Texas, Houston, TX 77030, U.S.A

Lei Liu: liulei@virginia.edu; Mark E. Cowen: ; Robert L. Strawderman: ; Ya-Chen T. Shih:

Abstract

In this paper, we propose a flexible “two-part” random Effects model (Olsen and Schafer 2001; Tooze, Grunwald, and Jones 2002) for correlated medical cost data. Typically, medical cost data are right-skewed, involve a substantial proportion of zero values, and may exhibit heteroscedasticity. In many cases, such data is also obtained in hierarchical form, e.g., on patients served by the same physician. The proposed model specification therefore consists of two generalized linear mixed models (GLMM), linked together by correlated random Effects. Respectively, and conditionally on the random Effects and covariates, we model the odds of cost being positive (Part I) using a GLMM with a logistic link and the mean cost (Part II) given that costs were actually incurred using a generalized gamma regression model with random Effects and a scale parameter that is allowed to depend on covariates (c.f. Manning, Basu, and Mullahy 2005). The class of generalized gamma distributions is very flexible and includes the lognormal, gamma, inverse gamma and Weibull distributions as special cases. We demonstrate how to carry out estimation using the Gaussian quadrature techniques conveniently implemented in SAS Proc NLMIXED. The proposed model is used to analyze pharmacy cost data on 56,245 adult patients clustered within 239 physicians in a mid-western U.S. managed care organization.

Keywords

Medical cost data; Mixed model; Random Effect; Health econometrics; Zero-inflated data

1. Introduction

The recently rekindled debate on health care reform in the United States and interest in how to pay for an expanded national program of care provides motivation to understand the relative influence of demographics, patient-level illness burden and provider-level variability in practice patterns on these costs. With advanced data collection and management techniques, medical cost data are now recorded routinely by hospitals, disease registries, and health insurance companies. In combination with development and application of state-of-the-art statistical methods of analysis, such widely available data can now be used to enhance our understanding of these various influences.

Medical cost data are frequently right-skewed, involve a substantial proportion of zero values, and may exhibit heteroscedasticity. For example, healthy people may incur no costs in a given year, while otherwise similar patients having a particular disease (e.g., cancer) may incur medical costs that increase tremendously by disease severity. Both observations suggest that no simple parametric distribution is suitable for describing such “semi-continuous” data.

Recognizing the need to account for the semi-continuous nature of medical cost data arguably began with the development of the “two-part” model (Cragg 1971; Manning et al., 1981)¹. Specifically, denote the cost of a given subject by Y ; then, a two-part model for the probability distribution of Y consists of (i) modeling the probability that $Y > 0$ using, say, a logistic or probit model (“Part I”); and, (ii) separately modeling the probability distribution of $Y|Y > 0$ (“Part II”). Numerous options exist for Part II; a convenient and common choice is to assume that $[\log Y|Y > 0]$ follows a linear regression model with normal errors (Manning et al. 1983; Diehr et al. 1999).

More recently, the two-part model has also been extended to deal with correlated outcome data. For example, Lu, Lin and Shi (2004) proposed a “marginal” version of this model, utilizing GEE for the purposes of estimation and inference. Olsen and Schafer (2001) and Tooze, Grunwald and Jones (2002) instead extended the two-part model to account for correlation through the introduction of random Effects; see also Zhang, Strawderman, Cowen and Wells (2006), who considered a related Bayesian formulation of this model. In the present context, such models permit one to account for situations in which costs incurred by subjects may be related to each other, as might be expected for patients served by the same physician or treated within the same hospital, or in cases where longitudinal cost information is available on a given patient. A typical random Effects formulation of the two-part model employs a generalized linear mixed model (GLMM) for the binary outcomes $Y > 0$ in Part I and a linear mixed model for the positive continuous outcomes $[\log Y|Y > 0]$ in Part II. The two parts of the model are then linked together by imposing a correlation structure on the random Effects that appear in Part I and Part II. Given the pair of random Effects and all covariates, the propensity (i.e., $Y > 0$) is usually assumed to be independent of the conditional response level (i.e., $[\log Y|Y > 0]$); averaging over the random Effects thus induces two forms of correlation. In the context of medical cost data with clusters of patients served by different physicians, the first form is standard and captures (i) the marginal correlation between the propensity to incur costs among patients served by the same physician; and, (ii) the marginal correlation between the level of costs actually incurred by patients served by the same physician. The second form is more specific to the two-part model and captures the relationship between the physician-specific propensity to incur cost and physician-specific level of actual costs incurred. Such “cross-part” correlation is of interest; for example, in the case of pharmacy cost data, it is interesting to ask whether a physician that has a higher probability of prescribing medication for his patients also

¹Enami and Mullahy (2008) traced the predecessor of two-part models back to Hald (1949) in the actuarial literature and Aitchison (1955) in statistics.

tends to prescribe more expensive medications for these patients (Zhang et al., 2006). As shown in Albert (2005), the parameters of a marginal two-part model, such as that proposed in Lu et al. (2004), can depend strongly on the underlying degree of cross-part correlation, resulting in a misleading assessment of the regression Effects. The specification of a random Effects model avoids this problem, allowing one to estimate interpretable regression Effects as well as characterize the marginal Effects of each variable on the response variable.

One important disadvantage of the typical formulation of the two-part model relates to the use of a transformation in Part II of the model. The resulting difficulties of re-transformation that arise are only compounded in the presence of random Effects. Duan (1983) proposed the “smearing” method for estimating the mean of the untransformed response Y after fitting a linear regression model to a transformed response (e.g., $\log Y$). However, if the transformation does not stabilize the variance, heteroscedasticity may be present (Duan et al. 1983; Manning 1998; Zhou, Stroup, and Tierney 2001; Zhou, Lin and Johnson, 2008). In this case, Duan’s smearing estimate cannot be employed and the use of ordinary least squares (OLS) leads to a biased estimate of the covariate Effect on the untransformed mean of Y (e.g., Manning, 1998; Mullahy, 1998; Manning and Mullahy, 2001). In response to these and other concerns, Manning, Basu, and Mullahy (2005, hereafter MBM) proposed to use the generalized gamma distribution for modeling the probability distribution of $Y|Y > 0$. This distribution is useful in that it contains the standard gamma, inverse gamma, Weibull and lognormal distributions as special cases. As a result, standard testing methods for nested hypotheses may be used in evaluating the fit of these simpler models; in addition, the model provides greater flexibility in cases where none might permit an adequate description of the data.

In this paper, we propose a new two-part model that incorporates correlated random Effects. Analogously to MBM, our model uses a generalized gamma GLMM for Y in part II. Similarly to MBM and also Zhou et al. (2001), we further permit the scale parameter of this distribution to depend on covariate information, allowing for subject-level heteroscedasticity in the cost data. The proposed model is both flexible and more general than those currently available in the literature, being capable of dealing with clustered data, heteroscedasticity, and cross-part correlation. Maximum likelihood estimation for the proposed model is easily implemented within SAS Proc NLMIXED (Littell et al. 2006). Further contributions of this paper include a clear demonstration of the Effects of heteroscedasticity in Part II of the two-part model on the interpretation and bias of estimated covariate Effects and a detailed discussion of the relationship between the proposed model and a strongly associated class of marginal two-part models.

We illustrate the use of this model by analyzing pharmacy cost data from a mid-western U.S. managed care organization (MCO) on 56,245 patients served by 239 primary care physicians (PCP). Important features of the data include the clustering of patients within PCP, the substantial proportion (i.e., 26%) of zero cost patients, and the highly skewed nature of the cost data among those patients with non-zero cost during the year under study (i.e., respectively, a mean and median cost of \$497.95 and \$87.63). A similar dataset is analyzed in Zhang et al. (2006) using a Bayesian two-part random Effects model, with a focus on profiling the physician contribution to patient pharmacy costs. In the present paper, the primary interest lies more in characterizing the patient-level factors (i.e., covariates) that influence the pharmaceutical expenditures of adult patients, accounting for the possibility of both a physician Effect and heteroscedasticity.

We remark here that the two-part model has been the subject of some controversy in the econometrics literature, where it has been frequently compared and contrasted with the sample selection model of Heckman (1976, 1979). The most relevant comparison is perhaps with the adjusted tobit model of van de Ven and van Praag (1981), a variant of the sample selection

model in which (i) ‘censored’ observations are actually observed as true zeros instead of missing data; and, (ii) a correlation structure is imposed on the two possible counterfactual responses (i.e., potential outcomes) at the level of the subject. Duan, Manning, Morris and Newhouse (1984) demonstrated that these two model classes are in fact distinct; Manning, Duan and Rogers (1987) provided Monte Carlo evidence to show the sensitivity of the adjusted tobit model to exclusion restrictions. However, Leung and Yu (1996) later presented alternative Monte Carlo evidence that demonstrated the results of Manning et al. (1987) were inherently biased against the selection model, particularly so when model parameters are estimated via limited information maximum likelihood. Leung and Yu (1996) concluded that the two-part and sample selection model classes are in general designed to answer distinct inferential questions and that both are useful in their respective contexts.

Despite appearances, the correlation structure induced by the two-part model with correlated random Effects is distinct from that imposed by the adjusted tobit model of van de Ven and van Praag (1981). As indicated above, the former asserts that the correlation structure exists at the level of the physician; it is not imposed at the level of the subject, as is the case with the adjusted tobit model. Moreover, unlike the sample selection model, there is information in the observed data (i.e., the cluster structure) that allows one to identify this correlation structure regardless of whether exclusion restrictions are imposed. For these reasons, the relationship between two-part and sample selection models is not considered further in this paper.

The rest of the paper is organized as follows. In Section 2, we review the generalized gamma distribution and its properties and then introduce the proposed two-part model. In Section 3, we give the relevant likelihood function and describe how estimation can be carried out in SAS; example SAS code is provided in the Appendix. In Section 4, simulation is used to assess the performance of the estimation method. In Section 5, the proposed model is applied to the dataset described above. Concluding remarks are given in Section 6.

2. Modeling Framework

2.1. The Generalized Gamma Distribution

Let $\Gamma(\cdot)$ denote the standard gamma function. Then, the density of the generalized gamma probability distribution depends on the parameters κ , μ and σ and is given by (e.g., MBM)

$$f(y; \kappa, \mu, \sigma) = \frac{\eta^\eta}{\sigma y \Gamma(\eta) \sqrt{\eta}} \exp[u \sqrt{\eta} - \eta \exp(|\kappa|u)] \quad (1)$$

where $\eta = |\kappa|^{-2} > 0$ and $u = \text{sign}(\kappa)(\log(y) - \mu)/\sigma$ depends on the sign of κ . It follows that

$$E(y) = \exp \left\{ \mu + \frac{\sigma \log(\kappa^2)}{\kappa} + \log[\Gamma(1/\kappa^2 + \sigma/\kappa)] - \log[\Gamma(1/\kappa^2)] \right\} \quad (2)$$

and

$$\text{Var}(y) = \{\exp(\mu) \kappa^{2\sigma/\kappa}\}^2 \left\{ \frac{\Gamma(1/\kappa^2 + 2\sigma/\kappa)}{\Gamma(1/\kappa^2)} - \left[\frac{\Gamma(1/\kappa^2 + \sigma/\kappa)}{\Gamma(1/\kappa^2)} \right]^2 \right\}$$

The generalized gamma distribution is very flexible, including the standard gamma, inverse gamma, Weibull and lognormal distributions as special cases. For example, if $\sigma = \kappa$, the generalized gamma distribution density (1) reduces to

$$f(y; \nu, \eta) = \frac{1}{\nu^\eta \Gamma(\eta)} y^{\eta-1} \exp(-y/\nu),$$

a standard gamma density with shape parameter $\eta = \kappa^{-2}$ and scale parameter $\nu = \kappa^2 \exp(\mu)$, i.e., mean $\exp(\mu)$ and variance $\kappa^2 \exp(2\mu)$. The inverse gamma distribution is instead obtained by setting $\kappa = -\sigma$ with $\sigma > 0$; that is, letting $\varepsilon = \eta e^\mu$, (1) reduces to

$$f(y; \varepsilon, \eta) = \frac{\varepsilon^\eta}{\Gamma(\eta)} \left(\frac{1}{y}\right)^{\eta+1} \exp(-\varepsilon/y),$$

which is recognized as a inverse gamma probability density function with shape parameter η and scale parameter ε (e.g., Robert, 2007, p. 520). The inverse gamma distribution plays an important role in Bayesian inference problems for normal regression models as the conjugate prior for the scale parameter (e.g., Robert, 2007, §4.4); the observation that this distribution arises as a special case of the generalized gamma density does not appear to be well known.

In contrast to the gamma and inverse gamma families, the Weibull and lognormal distributions are obtained by fixing κ at a specific value. Specifically, when $\kappa = 1$, density (1) reduces to the probability density function of a Weibull distribution. Alternatively, taking the limit of (1) as $\kappa \rightarrow 0$, one obtains

$$f(y; \mu, \sigma) = \frac{1}{\sigma y \sqrt{2\pi}} \exp \left\{ -\frac{(\log(y) - \mu)^2}{2\sigma^2} \right\},$$

a lognormal probability density function with parameters μ and σ .

2.2. A Flexible Two Part Random Effects Model

We are interested in modeling correlated semi-continuous cost data, characterized by a significant proportion of zeros and highly right-skewed and heteroscedastic continuous positive values. Denote Y_{ij} as the cost for the j -th patient clustered within physician i , where $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, n_i$. Define X_{ij} and Z_{ij} as the covariate vectors for fixed and random Effects, respectively; we assume that X_{ij} contains a column of ones for all subjects and hence that the model contains an intercept. Patients clustered within the same physician are expected to have a more similar prescription pattern in comparison with those treated by a different physician. The two-part model is specified conditionally on the covariates and a set of correlated random Effects $a_1 \dots a_n$ and $b_1 \dots b_n$. Specifically, for patient j treated within physician i , let $H_{ij} = (X_{ij}^T, Z_{ij}^T, a_i^T, b_i^T)^T$, and assume:

- Part I: $\pi_{ij} = P(Y_{ij} > 0 | H_{ij})$ denotes the probability of observing a positive cost, where

$$\text{logit}(\pi_{ij}) = X_{ij}^T \alpha + Z_{ij}^T a_i, \quad (3)$$

and α is a vector of regression coefficients;

- Part II: the probability distribution of $Y_{ij}(Y_{ij} > 0, H_{ij})$ follows the generalized gamma distribution of Section 2.1 with parameters κ , $\mu = \phi_{ij}$ and $\sigma = \tau_{ij}$, where

$$\phi_{ij} = X_{ij}^T \beta + Z_{ij}^T b_i, \quad (4)$$

$$\tau_{ij}^2 = \exp(X_{ij}^T \delta), \quad (5)$$

and β and δ are vectors of regression coefficients.

The random Effects a_i and b_i respectively account for the variation within clusters in parts I and II. It is assumed that $r_i = (a_i^T, b_i^T)^T \sim N(0, D)$, with D being a positive definite matrix. The correlation between a_i and b_i captures the cross-part correlation between the odds of nonzero cost and the level of positive cost for patients treated by the same physician. The dependence of τ_{ij} on X_{ij} allows for the possibility of heteroscedasticity; a homoscedastic model is obtained if $\delta_{(-1)} = 0$, where $\delta_{(-1)}$ denotes the set of regression coefficients excluding the intercept. While other functional forms can also be chosen in (5), the proposed form has the advantage that the estimated scale parameter is guaranteed to be positive, thus avoiding an unbounded likelihood (e.g., Crisp and Burridge, 1994). While it is technically possible for different covariates vectors to be used in (3)–(5) by simply fixing the corresponding fixed or random Effects to be zero, using the same set of covariates likely improves the interpretation of the coefficients and resulting summaries (e.g., expected costs).

Remark

An alternative way to capture within cluster correlation is to employ a common random Effect in parts I and II of the model; equivalently, the model specifies that the correlation of a_i and b_i is equal to one. An important disadvantage of this model specification is the interpretation of the random Effect, which has a different meaning and numerical Effect in Parts I and II of the model.

The proposed model is very comprehensive and includes many others as special cases. For example, by specifying $\kappa \rightarrow 0$, $\delta_{(-1)} = 0$ in (5), and $D = 0$, one obtains the original two-part model of Manning et al. (1981); if $\kappa \rightarrow 0$, $\delta_{(-1)} = 0$ and $D \neq 0$ is a positive definite 2×2 matrix, one instead obtains the model considered in both Olsen and Schafer (2001) and Tooze et al. (2002). As suggested in Section 2.1, allowing κ to be non-zero permits a much wider array of modeling options. For example, the choice $\kappa = 1$ leads to specifying Weibull model in Part II, with $(\delta_{(-1)} = 0)$ or without $(\delta_{(-1)} \neq 0)$ homoscedastic errors. Alternatively, if $\delta_{(-1)} = 0$, $\kappa = \pm\tau$, and $D = 0$ then one obtains a two-part model in which Part II respectively corresponds to a gamma or inverse gamma generalized linear model; a generalized linear mixed model instead obtains if D is assumed to be a positive definite matrix. If $\delta_{(-1)} \neq 0$, then it becomes impossible to enforce the condition that $\kappa = \pm\tau$, and the resulting part II model is no longer a member of either the gamma or inverse gamma families.

Note that while the proposed model specifies Part I as a logistic regression model, one is not restricted to this choice, either in theory or in computation. For example, specifying Part I via a probit model and letting $\kappa = 0$, $\delta_{(-1)} = 0$ and $D \neq 0$ be a positive definite 2×2 matrix leads to a frequentist version of the model considered in Zhang et al. (2006).

2.3. Interpretation of covariate Effects and impact of heteroscedasticity

2.3.1. For continuous covariates—In order to assess the overall impact of a continuous covariate x_s on the expected cost, it is sensible to consider (e.g., Manning et al., 2005, Sec. 2.1)

$$\frac{d \log E[Y|H]}{dx_s} = \frac{d \log P(x_s)}{dx_s} + \frac{d \log M(x_s)}{dx_s}, \quad (6)$$

where $H = \{x, z, a, b\}$, x and z are generic covariate vectors,

$$P(x_s) = \text{expit} \left\{ x_s \alpha_s + x_{(-s)}^T \alpha_{(-s)} + z^T a \right\}$$

for $\text{expit}(t) = e^t / (1 + e^t)$ and, from (2),

$$M(x_s) = \exp \left\{ \mu(x_s) + \frac{\sigma(x_s) \log(\kappa^2)}{\kappa} + \log[\Gamma(1/\kappa^2 + \sigma(x_s)/\kappa)] - \log[\Gamma(1/\kappa^2)] \right\}$$

with $\mu(x_s) = x_s \beta_s + x_{(-s)}^T \beta_{(-s)} + z^T b_i$ and $\sigma(x_s) = \exp \left\{ .5(x_s \delta_s + x_{(-s)}^T \delta_{(-s)}) \right\}$. Suppose that z is independent of x_s ; this is true, for example, in a model that only includes random intercept terms in parts I and II or with any covariate that appears in both parts I and II of the model only as a fixed Effect. Then, a straightforward calculation shows that (6) can be equivalently written

$$\frac{d \log E[Y|H]}{dx_s} = \alpha_s \{1 - P(x_s)\} + \beta_s + \delta_s \frac{\sigma(x_s)}{2\kappa} (\log \kappa^2 + \Psi(\theta(x_s))), \quad (7)$$

where $\theta(x_s) = \kappa^{-2} + \kappa^{-1} \sigma(x_s)$ and $\Psi(\cdot)$ denotes the digamma function. Equation (7) measures the local rate of change in $E[Y|H]$, considered as a function in x_s , relative to $E[Y|H]$. One can view this as an approximation to the percentage change in $E[Y|H]$ for a one unit increase in x_s (e.g., Wooldridge, 2002, p. 17). In fact, multiplying (7) by x_s , the resulting quantity is the cluster-specific “partial elasticity” of the mean cost with respect to x_s . The partial elasticity is a measure commonly used in economics (e.g., Wooldridge, 2002, p. 16).

The result (7), a computation carried out conditionally on the random Effects, shows that the derivative depends on the random Effects $r = (a^T, b^T)^T$ only through a in the term $P(x_s)$. There are several ways to evaluate $\frac{d \log E[Y|H]}{dx_s}$ with respect to random Effect a . For example, one might compute the average derivative by taking the expectation of (7) with respect to the marginal distribution of a . Because the random Effects $r = (a^T, b^T)^T$ appear only through a in the term $P(x_s)$, the average derivative is equal to (7) with $P(x_s)$ replaced by $E_a[P(x_s)]$. In practice, this can be calculated by numerical integration. A simpler but less appealing alternative is to compute $P(x_s)$ at a fixed value of a , such as $a = 0$. When multiplied by x_s , one therefore obtains the average partial elasticity of the mean cost with respect to x_s .

In the absence of heteroscedasticity in x_s (i.e., $\delta_s = 0$), (7) reduces to

$$\frac{d \log E[Y|H]}{dx_s} = \alpha_s \{1 - P(x_s)\} + \beta_s. \quad (8)$$

Here, the logarithm of the mean cost is observed to increase (decrease) if both α_s and β_s are positive (negative). Moreover, if $\text{sign}(\alpha_s) = \text{sign}(\beta_s)$, then the sign of β_s agrees with the sign of (7) and $|\beta_s|$ represents a lower bound on both the subject-specific and average rate of change regardless of x and z . The presence of heteroscedasticity adds the corresponding terms from (7) after the β_s term. This increases the complexity of the elasticity calculation and interpretation of Effects as well as demonstrating that its neglect is expected to create bias. For example, equality of the signs of α_s and β_s is no longer sufficient to determine the overall impact of a covariate x_s on the mean cost because this Effect additionally depends on the sign of $\delta_s \kappa^{-1} (2 \log |\kappa| + \Psi(\theta(x_s)))$, as well as the relative magnitudes of the two terms appearing on the right-hand side of (7). Importantly, such complications arise not because of the use of a two-part model or the inclusion of random Effects but rather due to the dependence of the mean of the generalized gamma distribution on the location and scale parameters; see (2).

2.3.2. For binary covariates—With a binary covariate, the formulas (6) and (7) no longer apply. Without loss of generality, suppose that x_s (e.g. s -th disease indicator) takes on the value 0 or 1. Then, the proper analog of (6), the log ratio of expected cost with and without disease s , is

$$\log \frac{P(1)}{P(0)} + \log \frac{M(1)}{M(0)}, \quad (9)$$

where, using previous notation, we have

$$\log \left(\frac{P(1)}{P(0)} \right) = \alpha_s + \log \left(\frac{1 + \exp\{x_{(-s)}^T \alpha_{(-s)} + z^T a\}}{1 + e^{\alpha_s} \exp\{x_{(-s)}^T \alpha_{(-s)} + z^T a\}} \right).$$

and

$$\log \left(\frac{M(1)}{M(0)} \right) = \beta_s + \frac{\log(\kappa^2)}{\kappa} [\sigma(1) - \sigma(0)] + \log \left(\frac{\Gamma(1/\kappa^2 + \sigma(1)/\kappa)}{\Gamma(1/\kappa^2 + \sigma(0)/\kappa)} \right).$$

In the absence of heteroscedasticity in x_s (i.e., $\sigma(1) = \sigma(0)$), the logarithm of the mean cost is again observed to increase (decrease) if both α_s and β_s are positive (negative). Moreover, and similarly to the case of a continuous covariate, equality of the signs of α_s and β_s in the presence of heteroscedasticity is not sufficient to determine whether costs (either overall or considering only contributions from Part II) increase or decrease. Similarly to (7), the derivative (9) can be evaluated by taking the expectation with respect to the marginal distribution of random Effect a or by evaluating it at a given value of a .

2.4. Marginal versus conditional two-part models

The interpretation of the fixed Effects (e.g., regression parameters) in the proposed nonlinear mixed Effects model is conditional on the level of the random Effects a and b . Such a “subject

specific” interpretation may not always be of interest; hence, following the suggestions of a referee, we introduce a so-called marginal model that does not rely on the specification of a random Effects structure. In contrast to the mixed Effects model, the fixed Effects in a marginal model have a “population averaged” interpretation and are in general distinct from their analogs in the mixed model outside the setting of additive linear models (e.g., Zeger, Liang and Albert, 1988).

Section 2.2 posits a conditional model for $E[Y|H]$, where we recall the notation $H = \{x, z, a, b\}$ for generic covariate vectors x and z . Letting $H^* = \{x, z\}$, a two-part model for the marginal expectation $E[Y|H^*]$ is an obvious alternative to $E[Y|H]$. Specifically, and analogously to Section 2.2, we consider the marginal two-part model $E[Y|H^*] = P_{\alpha_M}(x)M_{\theta_M}(x)$, where $P_{\alpha_M}(x) = P\{Y > 0|H^*\} = \expit\{x^T \alpha_M\}$,

$$M_{\theta_M}(x) = E[Y|H^*, Y > 0] = \exp\left\{x^T \beta_M + \frac{\sigma_{\delta_M} \log(\kappa_M^2)}{\kappa_M} + \log[\Gamma(1/\kappa_M^2 + \sigma_{\delta_M}/\kappa_M)] - \log[\Gamma(1/\kappa_M^2)]\right\},$$

$\theta_M = (\beta_M^T, \delta_M^T, \kappa_M)^T$, and $\sigma_{\delta} = \exp(x^T \delta_M)/2$. In contrast to the model of Section 2.2, this model involves no random Effects and the parameters used to describe the fixed Effects, or α_M and θ_M , therefore have population-averaged interpretations. However, similarly to Section 2.3 and unlike many marginal models, the model parameters α_M and θ_M do not directly describe the marginal Effect of an individual variable x_s on the mean cost or its average partial elasticity; see, for example, (7) and (8).

The indicated marginal two-part model coincides with that of Section 2.2 if the covariance matrix D of $(a^T, b^T)^T$ in the latter random Effects model is set to zero. More generally, if $D \neq 0$, the model of Section 2.2 induces a different formula for the marginal mean cost $E[Y|H^*]$. To see this, note that an easy algebraic argument shows that we may rewrite $E[Y|H]$ as $P_{\alpha}(x)M_{\theta}(x)K_{\alpha}(a, b, x, z)$, where

$$K_{\alpha}(a, b, x, z) = \frac{e^{z^T(a+b)}(1 + e^{x^T \alpha})}{1 + e^{x^T \alpha + z^T a}}$$

and $P_{\alpha}(x)$ and $M_{\theta}(x)$ are defined exactly as above, but with α_M and θ_M replaced by α and $\theta = (\beta^T, \delta^T, \kappa)^T$. Averaging over the random Effects then induces the marginal mean cost model $E[Y|H^*] = P_{\alpha}(x)M_{\theta}(x)K_{\alpha}(x, z)$, where

$$\tilde{K}_{\alpha}(x, z) = (1 + e^{x^T \alpha}) \times E\left[\frac{e^{z^T(a+b)}}{1 + e^{x^T \alpha + z^T a}} \middle| H^*\right]. \quad (10)$$

Using the fact that the joint distribution $(a^T, b^T)^T$ is multivariate normal with mean zero and covariance matrix D , it is easy to show that $\tilde{K}_{\alpha}(x, z) = 1$ if and only if $D = 0$. Thus, in the presence of random Effects, the conditional and marginal two-part models do not lead to equivalent representations for $E[Y|H^*]$.

Consistent with the calculations of Albert (2005), the results above further suggest that the interpretation of α_M and θ_M may be impaired in cases where a cluster-specific relationship

exists between the probability of positive response (here, cost) and the associated level, given any. Specifically, in order for equality to hold between $P_{\alpha}(x)M_{\theta}(x)\tilde{K}_{\alpha}(x, z)$ and $P_{\alpha_M}(x)M_{\theta_M}(x)$ at a given x , the parameters α_M and θ_M must necessarily depend on D . This reduces the appeal of the marginal model for characterizing the impact of covariates on both the propensity for incurring costs and the average level of costs, given any. However, similarly to Section 2.3, the Effect of a given covariate x_s on $E[Y|H^*]$ or $\log E[Y|H^*]$ can also be characterized by taking the derivative with respect to x_s of the marginal two-part model mean prediction $P_{\alpha_M}(x)M_{\theta_M}(x)$ or its logarithm. An expected disadvantage of this approach is the relative inefficiency of the estimates derived under the marginal model. Comparisons between the corresponding average partial elasticities that would be obtained under both models are provided in our simulation study (Section 4) and data analysis (Section 5).

Remark—In the case where $D \neq 0$, $\tilde{K}_{\alpha}(x, z)$ in (10) has an interesting interpretation as the relative error in the mean cost prediction that would be incurred by using the subject-specific parameter values α and θ in the marginal two-part model mean prediction function $P_{\alpha}(x)M_{\theta}(x)$. It is difficult to characterize the behavior of $\tilde{K}_{\alpha}(x, z)$ in general. However, in the setting where a and b denote random intercepts (i.e., where z is a vector of ones), it is possible to calculate bounds on $\tilde{K}_{\alpha}(x, z)$ that depend only on the components of D . Specifically, let ρ denote the correlation between a and b and note that $\sigma_a^2 = d_{11}$ and $\sigma_b^2 = d_{22}$ denote the variances of a and b . Then, tedious but straightforward calculations show that

$$\tilde{K}_{\alpha}(x, z) \geq e^{\frac{1}{2}(\sigma_b^2(1-\rho^2))} \times \left[e^{\frac{1}{2}(\rho\sigma_b + \sigma_a)^2} \{1 - \Phi(\rho\sigma_b + \sigma_a)\} + e^{\frac{1}{2}(\rho\sigma_b)^2} \Phi(\rho\sigma_b) \right]$$

and

$$\tilde{K}_{\alpha}(x, z) \leq e^{\frac{1}{2}(\sigma_b^2(1-\rho^2))} \times \left[e^{\frac{1}{2}(\rho\sigma_b + \sigma_a)^2} \Phi(\rho\sigma_b + \sigma_a) + e^{\frac{1}{2}(\rho\sigma_b)^2} \{1 - \Phi(\rho\sigma_b)\} \right],$$

where $\Phi(\cdot)$ denotes the cumulative distribution function for the standard normal distribution. Of importance here are the facts that these bounds depend only on the components of D and are each equal to one when $D = 0$. Hence, subtracting the lower bound from the upper bound and rearranging terms, these bounds demonstrate directly that $\tilde{K}_{\alpha}(x, z)$ must be close to 1 regardless of x when both σ_a and $\rho\sigma_b$ are small. In other words, in cases where cluster-specific contribution to variation is small, the mean predictions and estimates of fixed Effects obtained under the two-part random Effect model should be reasonably close to those obtained under the marginal two-part model. In this case, the proposed methodology is primarily expected to result in efficiency gains.

3. Maximum Likelihood Estimation for the Two Part Random Effects Model

Denote $\mathbf{O}_i = \{Y_i, X_i, Z_i\}$ as the observed data for cluster i , where $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{i,n_i})^T$ and X_i and Z_i are defined accordingly. The observed data likelihood function is

$$L(\alpha, \beta, \delta, D) = \prod_{i=1}^n L_i(\alpha, \beta, \delta, D),$$

where $L_i(\alpha, \beta, \delta, D)$ denotes the contribution of cluster i (i.e., for the data $\mathbf{O}_i$). The parameter δ is a vector; under a homoscedasticity assumption, $\delta_{(-1)}$ is fixed at zero and the parameter is Effectively scalar. Under the proposed model, we have

$$L_i(\alpha, \beta, \delta, D) \propto \int \exp(l_1(r_i) + l_2(r_i)) p(r_i) dr_i, \quad (11)$$

where $r_i = (a_i^T, b_i^T)^T$,

$$l_1(r_i) = \sum_{j=1}^{n_i} [I(Y_{ij} > 0) \text{logit}(\pi_{ij}) + \log(1 - \pi_{ij})], \quad (12)$$

π_{ij} is defined as in (3),

$$l_2(r_i) = \sum_{j=1}^{k_i} [(\eta - 0.5) \log(\eta) - \log \tau_{ij} - \log Y_{ij} - \log \Gamma(\eta) + u_{ij} \sqrt{\eta} - \eta \exp(-|k|Y_{ij})], \quad (13)$$

where $k_i = \sum_{j=1}^{n_i} I(Y_{ij} > 0)$, $u_{ij} = \text{sign}(\kappa)(\log(Y_{ij}) - \phi_{ij}) = \tau_{ij}$, τ_{ij} is defined as in (5) and $p(r_i)$ denotes the probability density function for a $N(0, D)$ random variable. The specifications (12) and (13) respectively correspond to parts I and II of the proposed model as defined in Section 2.2.

As can be seen from (11), the likelihood contributions $L_i(\alpha, \beta, \delta, D)$, $i = 1 \dots n$ each involve the integral of a complicated nonlinear function of the vector r_i with respect to the multivariate normal probability density function. Various computational techniques, such as (adaptive) Gaussian quadrature (e.g., Anderson and Aitkin, 1985; Hedeker and Gibbons, 1994; Pinheiro and Bates, 1995; Liu and Pierce 1994), Laplace approximation (Tierney and Kadane, 1986; Wolfinger 1993), partial quasi-likelihood (PQL; Breslow and Clayton, 1993; Goldstein and Rasbash, 1996), and Monte Carlo techniques (McCulloch 1997) have been proposed to deal with such analytically intractable integrals frequently arising in nonlinear mixed Effects models. A recent summary of these methods is available in Molenberghs and Verbeke (2005, Chapter 14).

In the case where $\{Y|Y > 0, b\}$ follows a lognormal distribution, Olsen and Schafer (2001) suggested a sixth order Laplace approximation (Raudenbush, Yang, and Yosef 2000) for estimation, whereas Tooze et al. (2002) utilized adaptive Gaussian quadrature. In our experience, both methods generally perform well, with the sixth order Laplace approximation being considerably faster in high dimensional settings (i.e., ≥ 4 random Effects). However, this approximation is not yet available in commercial software packages for fitting nonlinear mixed Effects models; by contrast, adaptive Gaussian quadrature is conveniently implemented in SAS Proc NLMIXED (Littell et al. 2006). An overview of the adaptive Gaussian quadrature approximation method is given in the appendix, along with sample SAS code for fitting the proposed model.

4. Simulation

In this section we report results from a simulation study done to evaluate the performance of the proposed estimation procedure. We assume that there is a single fixed Effects covariate

X_1 , simulated here from a Uniform(0, 1) distribution. Fixed Effects coefficients are set to be $\alpha = (\alpha_0, \alpha_1)^T = (-1, 0.5)^T$ and $\beta = (\beta_0, \beta_1)^T = (1, 1)^T$, implying that each part of the model has an intercept term. We further assume that the random Effects a_i and b_i are each scalar quantities, hence that the model intercepts vary by cluster; moreover, $r_i = (a_i, b_i)^T$ follows a bivariate normal distribution with covariance matrix

$$D = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

We set $\kappa = 2$ in (1) and permit heteroscedasticity, assuming specifically that $\delta = (\delta_0, \delta_1)^T = (1, 2)^T$ in (5). In this model, approximately 65% of the simulated costs are zero; furthermore, the mean of the positive costs is \$29.6 with a standard deviation of \$232.9. That is, the distribution of the simulated costs is semi-continuous, with a large point mass at zero and highly skewed to the right.

The simulation results summarized in Table 1 are based on 600 independent samples, each consisting of $n = 200$ physicians and $n_i = 12$ patients per physician. All models were fit using SAS Proc NLMIXED and adaptive 5-point Gaussian quadrature. Increasing the number of quadrature points to 10 substantially increased computation time with negligible changes to the results reported in Table 1.

The panel of Table 1 labeled “Proposed Model” demonstrates that the empirical biases are negligible and the coverage probabilities are acceptably close to the nominal level 0.95 for each of the model parameters; in addition, we observe small differences between the simulated and average estimated standard error estimates.

For comparison, we additionally fit two reduced (i.e., misspecified) models. The reduced model A assumes a lognormal distribution, allowing for heteroscedasticity; the reduced model B assumes a generalized gamma distribution but assumes homoscedasticity (i.e., $\tau_{ij} = \tau$ is constant for all subjects). The results for reduced models A and B are shown in the middle and right panels of Table 1. We observe that the estimates obtained for Part I are virtually identical. However, large discrepancies are found between the β coefficients estimated from either of reduced models and those estimated using the proposed model. We additionally find biased estimates for δ_0 and d_{22} in reduced model A, and for κ and d_{22} in reduced model B.

To study the overall Effect of the continuous covariate X_1 on mean cost, we calculate the average derivative of $\frac{d \log E[Y|H]}{dX_1}$ using (7), where $E_a[P(X_1)]$ in (7) is obtained by numerical integration. As explained earlier, multiplying the resulting point estimates by the value of X_1 gives the estimated average partial elasticity. One may also calculate $\frac{d \log E[Y|H]}{dX_1}$ under reduced models A and B. Under reduced model A,

$$\frac{d \log E[Y|H]}{dX_1} = \alpha_1 \{1 - P(X_1)\} + \beta_1 + \delta_1 \frac{\sigma^2(X_1)}{2}; \quad (14)$$

under reduced model B, we obtain (8), i.e.,

$$\frac{d \log E[Y|H]}{dX_1} = \alpha_1 \{1 - P(X_1)\} + \beta_1. \quad (15)$$

As described in Section 2.4 and per the suggestion of a referee, we also include the analogous results obtained from fitting a marginal two-part model to the simulated data.

Table 2 summarizes the results of these calculations for X_1 in all settings, which we refer to as the impact of X_1 on the logarithm of the mean cost (or log mean cost). Because the marginal Effect differs with the value of X_1 , we show these results at the quartiles $X_1 = 0.25, 0.5$, and 0.75 . With a correctly specified model, we find that the estimated Effect for X_1 is very close to its theoretical value at all three covariate values. However, under reduced models A and B, the estimates are substantially biased. Interestingly, we further observe that the bias under reduced model A is even more substantial than that under reduced model B, an indication of the importance of selecting an appropriate parametric model in Part II. The use of a marginal model yields point estimates in Table 2 that are close to those obtained under the random Effects model. However, these estimates exhibit greater spread, as measured by the distance between the upper and lower 2.5% percentiles of the corresponding sampling distributions. In particular, and in each case, the distance between these percentiles obtained under the marginal model is on average 13% longer than those obtained under the random Effects model.

Finally, we run a simulation where reduced model A holds. The parameters are identical to the first setting described above, except that Part II of the data generating model is assumed to follow a lognormal distribution (i.e., a generalized gamma distribution with $\kappa \rightarrow 0$). Fitting the data assuming a generalized gamma distribution in Part II, the parameter estimates are observed to have very small biases (results not shown); for example, the simulated mean of the MLE $\hat{\kappa}$ is 0.003 (SE=0.093). In addition, the corresponding mean estimates derived under (7) exhibit little, if any, bias. These results demonstrate that the proposed estimation procedure behaves well in this limiting special case.

5. Analysis of Pharmacy Cost Data

The pharmacy cost data are taken from a mid-western US MCO. The original dataset includes 98,382 patients clustered within 252 PCPs. Our study sample is limited to non-pediatric physicians serving only adult patients (i.e., aged 18 years and over) and consists of 56,245 patients served by 239 PCPs. The outcome variable is total pharmacy expenditure, defined to be the total consumption of pharmaceuticals evaluated at the contracted wholesale price minus a network discount for all drugs prescribed during a one year period. Each physician serves a different number of patients, ranging from 2 to 1,820. Of these patients, 14,721 (26%) had zero pharmacy cost; the remaining patients respectively had a mean and median cost of \$497.95 and \$87.63, with a maximum \$68,442. Further details regarding the study site and data may be found in Cowen et al. (1998); see also Cowen and Strawderman (2002) and Zhang et al. (2006) for related analyses. It should be noted that the dataset utilized in Zhang et al. (2006) differs slightly from our study sample due to refinements in the definition of “non-pediatric” physicians.

The original dataset contains information on 37 diagnostic groups (i.e., covariates), along with limited patient and physician demographic information; see, for example, Cowen and Strawderman (2002). For computational purpose, we have regrouped the data into 11 major diagnosis categories, including: infection, diabetes, affective disorders, neuropsychiatry, epilepsy, HIV, anxiety, hypertension/hyperlipidemia (HTN_LIPD), heart, asthma and chronic obstructive pulmonary disease (Asthcopd), and collagen vascular diseases. Other possible data

reductions were explored, for example, further combining some of the disease groups. However, we determined that the proposed reduction led to the best results in terms of AIC criteria. In addition to these eleven diagnosis groups and similarly to Zhang et al. (2006), we include patient age (centered at its mean, 43 years old, and measured in units of 10 years) and physician gender as covariates. In this study, we additionally consider all two-way interactions between diabetes, heart, HTN_LIPD and also the interaction between anxiety and affective disorders. The same set of covariates are used in the specification of parts I and II of the model, including the scale parameter (5). Similarly to Zhang et al. (2006), we include random intercepts in the Part I and Part II model, allowing for the possibility that these intercepts are correlated.

The results of fitting the heteroscedastic generalized gamma model are summarized in Tables 3 and 4. Table 3 respectively reports the MLEs, estimated standard errors and p-values associated with the regression parameters α and β in parts I and II of the model; Table 4 respectively reports the same information for δ in (5), the variance components d_{11} , d_{12} , and d_{22} (i.e., the unique elements of the 2×2 matrix D), and the generalized gamma distribution parameter κ . For comparison, we also fit models that correspond to the reduced models A and B considered in Section 4; that is, reduced model A corresponds to using a heteroscedastic lognormal distribution in Part II, while reduced model B corresponds to using a homoscedastic generalized gamma distribution in Part II. All models are fitted using SAS Proc NLMIXED with adaptive 5-point Gaussian quadrature. All assessments of statistical significance for the fixed effects throughout the analyses summarized below are carried out conditionally on the estimated variance components.

5.1. Interpretation of regression effects and variance components

Table 3 demonstrates that most covariates have a statistically significant impact on cost. For example, all disease conditions, alone or in combinations, raise the probability of experiencing costs (Part I). In addition, the odds of having a prescription increase for patients served by female physicians and as patients age. We further observe that the signs of the coefficients in Part II of the model agree with those in Part I. However, as pointed in Section 5.2 below, the results of Table 4 demonstrate the presence of statistically significant heteroscedasticity. For reasons discussed earlier, it implies that one cannot directly interpret the signs of the coefficients in the Part II model. For example, the positive coefficient observed for Epilepsy in Part II of the model is by itself insufficient to conclude that the mean cost for a patient with epilepsy is higher than the mean cost for a patient without epilepsy, all else being equal. Importantly, the above issue arises as a direct result of the presence of heteroscedasticity and is not a consequence of using a two-part model nor the inclusion of random effects. We reserve further comment on the interpretation of the fixed effects results in Section 5.3, where we evaluate the impact of disease and demographic covariates on the mean cost through (7) and (9).

The estimated variance components $\hat{d}_{11} = 0.045$ and $\hat{d}_{22} = 0.024$ in Table 4 are observed to be small in magnitude. However, each is highly significant ($p < 0.0001$), suggesting that unexplained heterogeneity may still persist, for example, due to differences in physician prescribing styles and other reasons. The estimated covariance $\hat{d}_{12}$ is positive ($p < 0.0001$), further suggesting a strong positive correlation ($\hat{\rho} = 0.024/[(0.045)(0.030)]^{1/2} = 0.653$) between the two model components. As indicated in Zhang et al. (2006), one possible interpretation of this result is that physicians who are more likely to issue prescriptions also tend to prescribe more intensively (e.g., multiple medications or more expensive medications) for their patients. The MLEs of the variance components and correlation reported here are comparable in magnitude to the posterior means of these parameters reported in Zhang et al. (2006, Table 1).

5.2. Heteroscedasticity and model comparisons

Appropriate versions of reduced models A and B described in Section 4 were also fit to these data. The regression parameters for model B are not provided since we assume homoscedasticity. Similarly to Table 1, Table 3 demonstrates that (a) there exist negligible differences in parameter estimates in Part I among the three models; and, (b) more noticeable differences occur for the estimated regression coefficients under Part II. For example, covariate effects for both Infection and Asthcopd differ by more than 10% between the proposed model and the reduced model A; in reduced model B, we find that the estimates for Neuropsychiatry, Epilepsy, HIV, and the interaction between affective disorder and anxiety differ substantially from those obtained under the proposed model. The majority of the Part II standard error estimates under reduced model B exceed those for the proposed model; this result is likely a consequence of reduced bias, though may potentially reflect improved efficiency of the proposed model.

The likelihood ratio statistic comparing the proposed model to reduced model A evaluates whether specifying a heteroscedastic lognormal distribution for Part II might be appropriate; we find $p < 0.0001$, suggesting that the generalized gamma distribution provides a better fit to these data. Similarly, the likelihood ratio test statistic for the global test $H_0: \delta_{(-1)} = 0$ (i.e., reduced model B) vs. the alternative that at least one of these parameters is non-zero (i.e., the proposed model) is also statistically significant ($p < 0.0001$), providing a strong evidence of heteroscedasticity. Inspection of the individual δ 's and corresponding test results in Table 4 further shows that each disease condition, except neuropsychiatry and HIV, has a negative, statistically significant coefficient. One interpretation of this result is that patients having a given disease status will exhibit less heteroscedasticity in comparison to those without this condition; such a result might be expected in cases where most of the patients under study have only a few of indicated disease conditions. We also find that increasing patient age and female physician gender are associated with increased homoscedasticity.

5.3. Impact of covariates on mean costs

In order to characterize the overall effect of the disease covariates on cost under the proposed model, we first calculate (9) for each binary disease category, holding all other covariates at 0 (or mean value for centered age). The resulting estimated conditional partial effects are then averaged over the estimated distribution of the random effects in order to produce average partial effects. A similar calculation is done for the continuous covariate "patient age" using (7). These results, as well as 95% confidence intervals, are summarized in Table 5. Completely analogous calculations, but carried out under the reduced models A and B, are also summarized in Table 5. Finally, we also summarize results obtained using the marginal two-part model of Section 2.4 that disregards the random effects; the corresponding 95% confidence interval is computed using 600 cluster bootstrap samples.

The resulting average partial effect estimates are summarized in Table 5 and provide a different but equivalent formulation of the effect estimates summarized in Tables 3 and 4. Variables that appear only as main effects in Tables 3 and 4 are summarized in the top half of Table 5 and describe the effect on cost for the indicated condition for a patient of mean age in the absence of all other conditions. For example, the "Infection" category evaluates the effect of infections on cost, relative to those who have not experienced infection, for patients of average age having no other conditions. Variables appearing as both main effects and in interaction terms in Tables 3 and 4 are summarized in the bottom half of Table 5. The interpretations of these effects are similar to those variables appearing only as main effects. For example, the "Diabetes" category evaluates the effect of diabetes on cost in the absence of all other conditions. Similarly, "Diabetes + Heart" characterizes the combined impact of diabetes and heart disease on cost in the absence of all other conditions. In short, Table 5 now describes the impact of several

mutually exclusive disease categories on cost. With 11 disease categories, there are 2048 possible disease combinations; for the purposes of illustration, we therefore consider only the individual conditions represented by the main effect and interaction terms appearing in Tables 3 and 4. In these and all other cases, the estimated effects are easily obtained using “estimate” statement in SAS Proc NLMIXED.

Consider first the proposed model. In the top half of Table 5 and also for the main effects appearing in the bottom half, we observe that the signs of the change in the logarithm of the mean cost in Table 5 agree with the corresponding main effects reported in Table 3. For example, the results suggest that older patients have higher costs overall (Table 5), consistent with receiving prescriptions more often and tending to spend more for their prescriptions (Table 3). In addition, patients seeing female physicians have higher overall costs (Table 5), consistent with being likely to get prescriptions and spending more on their medications (Table 3). In general, all disease groups are estimated to significantly increase the odds of medication prescription, with presence of any individual disease increasing the average pharmacy cost, results that are also reflected in Table 5.

The interaction terms in Tables 3 and 4 have significantly negative coefficients in parts I and II. These results suggest that the impact of the specified disease conditions on each component of the mean pharmacy cost is not additive for patients with two (or more) of these conditions, a result also reflected in the effect estimates summarized in Table 5. For example, we see that the sum of the estimated effects for Heart (i.e., 0.890) and for HTN_LIPID (i.e., 0.732), or 1.622, exceeds that for the combination of these two conditions (i.e., 1.224). This result is not unexpected; because the diagnosis groups are broadly defined, each category may include patients that share common conditions but do not necessarily require additional medication. Either diagnosis alone might require medication, but two diagnoses together may be adequately treated with a single medication. For example, beta blockers can be used to lower blood pressure (i.e., used for the covariate HTN_LIPID), as well as to reduce the mortality risk of myocardial infarctions (Heart). However, because cardiac patients often have hypertension, a single medication may be sufficient to capture the costs associated with both diagnoses.

The proposed model and corresponding marginal model lead to reasonably similar point estimates; as discussed in the next subsection, this similarity can be attributed to the small magnitude of the estimated variance components. However, on average, the 95% confidence intervals obtained under the marginal model are also 10% longer, reflecting decreased efficiency in comparison with the random effects model. Comparisons between the results of the proposed model and those obtained under the reduced models A and B provide further information. For example, while the results of Tables 3 and 4 indicate that the lognormal assumption imposed under model A is not appropriate, the results of Table 5 indicate that the overall effect estimates are not particularly misleading, with model A underestimating some effects and overestimating others. Comparisons between the proposed model and those under model B help illustrate the persistent biases that arise when one ignores heteroscedasticity. For example, from Table 5, the overall covariate effect of Asthcopd on the change in the logarithm of the mean cost is estimated to be 0.795 under the proposed model and 1.014 under reduced model B. This difference is remarkable because the estimated impact of this variable in Part II are observed to be rather close (0.673 and 0.706 respectively). In other words, failure to adjust for heteroscedasticity ($\delta = -0.299$ in Table 4) tends to inflate the estimated impact of this variable. A similar pattern is observed for many other disease variables, an interesting exception being neuropsychiatry. For neuropsychiatry patients, the estimated impact in Table 5 is somewhat higher under the proposed model than that under model B (i.e., 2.272 versus 2.169), with estimated Part II effects for this variable respectively being 1.530 and 1.907 and the estimated effect on heteroscedasticity being $\delta = 0.512$ (Table 4). That is, neuropsychiatry patients tend to exhibit greater heteroscedasticity in costs; this results in an inflated assessment

of its effect in Part II but a slight decrease in its estimated impact on the overall mean cost. More generally, we see that the 95% confidence intervals for the log mean cost obtained under the proposed model do not overlap with those obtained under model B for 7 risk factors (and all but one of the included interactions): infection, diabetes, affective, anxiety, HTN_LIPD, Asthcopd, and age. For these same effects, interval estimates obtained under the proposed model lie completely below those obtained under model B, indicating that a failure to adjust for heteroscedasticity results in overestimating the impact of each variable.

5.4. Physician-specific vs. population-averaged interpretations

The interpretations of the regression effects summarized in Tables 3 and 4 for the proposed nonlinear mixed effects model are physician-specific, i.e., conditional on the physician-level random effects. Such conditional summaries may be appropriate and of interest when the primary goal is to characterize the impact of disease on both the propensity to prescribe and subsequent average level of pharmacy costs among those receiving prescriptions after accounting for physician-specific differences (e.g., prescribing styles, risk arrangements).

However, as explained in Section 2.3 and regardless of the magnitude of the variance components, the interpretation of the individual model components in parts I and II are also adversely impacted by the presence of heteroscedasticity. One way of deflecting the impact that heteroscedasticity has on the interpretation of physician-specific regression coefficients is to instead compute physician-specific measures of partial effects and elasticity measures; see, in particular, (7) and (9). A disadvantage here is that the random effects appearing in the corresponding physician-specific quantity may lack an interpretable unit of measurement; for example, the correlated random effects used in the proposed two-part model are measured on very different scales. Thus, while the random effects structure remains a useful device for capturing sources of dependence (see Section 6 for further elaboration on its utility), it may be challenging to specify meaningful pairs of random effects for the purposes of comparison and reporting conditional partial effects.

Averaging the physician-specific estimates of partial effect in (7) and (9) over the distribution of random effects, as was done for Table 5, leads to partial effect estimates with a population-averaged interpretation. This point of view is frequently preferred in economic evaluation (c.f. Wooldridge, 2002, Sec. 2.2.5) and is appropriate for settings where the goal is to identify the overall influence of disease on cost from a policy or societal perspective. The marginal two-part model of Section 2.4 provides an alternative route for obtaining such effect estimates. As suggested by Tables 2 and 5, the proposed model can be expected to lead to more precise estimates of effect, especially so in settings where the cluster-level contributions to variability are not small. In the present example, the similarity of the point estimates is a consequence of the fact that the overall magnitude of the variance components is small. Specifically, in the data analysis, we estimated $\sigma_a^2 = d_{11}$ to be 0.045, $\sigma_b^2 = d_{22}$ to be 0.03, and $\rho = d_{12}/(\sigma_a\sigma_b)$ to be 0.653. Evaluating the bounds for (10) in Section 2.4, the corresponding relative error is between 0.96 and 1.14. In other words, were we to substitute the parameter estimates of Tables 3 and 4 into the marginal two-part mean model described in Section 2.4, the resulting predicted mean cost would lie within -4% and 14% of the actual mean prediction that would have been obtained by marginalizing the random effects model, no matter what the covariates. Such insensitivity helps to explain the similarities observed between the point estimates in Table 5 obtained under the proposed and marginal two-part models.

6. Discussion

In this paper, we propose a random effects two-part model for clustered semi-continuous response data, the primary novelties being the use of the very flexible generalized gamma

regression model and the incorporation of heteroscedasticity into Part II of the model. Our proposed model simultaneously takes into account the presence of true zeros, right-skewness and heteroscedasticity of positive response values. It further permits cluster-level correlation between the odds of observing a positive response and the actual level of this response. The resulting model encompasses a substantial subset of the parametric models for semicontinuous data previously proposed in the literature, providing a useful framework in which competing models can be evaluated. Estimation is conveniently implemented in SAS Proc NLMIXED, with Gaussian quadrature techniques used for estimation. Our simulation results demonstrate the importance of using a Part II model capable of handling a variety of distributional shapes as well as the impact that unrecognized heterogeneity can have on the quality of estimation.

As suggested by the discussions of Sections 2.4 and 5.4, the inclusion of random effects in regression models may lead to regression coefficients whose subject-specific interpretation may not always be desirable from a policy perspective. However, it is always possible to derive estimates from a random effects model that have a population-averaged interpretation. In addition, the inclusion of random effects in regression models can still provide helpful insights when developing and evaluating health care policy. Random effects can be used to characterize the relative amount of variability attributable to the cluster unit; while this unit is the physician in our study, the clustering unit could conceivably exist at a higher level, such as managed care organizations or pharmacy benefit managers. Methods for managing pharmacy costs might be emphasized or expanded depending on this information. For example, a relatively large cluster effect might encourage policy makers to focus interventions on the clustering unit, such as expanding pay-for-performance programs or changing the pharmacy benefit design to encourage patients to either switch from high- to low-cost providers or leave low-performing providers for high-performing providers. Smaller cluster effects, as observed in the present example, might instead suggest that these interventions have only a modest impact on costs, and therefore more emphasis should be given to patient-level interventions such as disease management programs, further formulary restrictions or higher levels of cost-sharing. Another motivation for estimating the random effects explicitly is to generate rankings of the clustering units that could be used to distribute cluster-level incentives (e.g., Cowen and Strawderman, 2002; Zhang et al. 2006), or to conduct qualitative investigations to identify unknown cluster-level characteristics that might explain the differences between higher versus lower tiered performers. In the context of two-part models, the inclusion of random effects not only allows for modeling clustering effect, but introduces a sensible mechanism for modeling the cross-part correlation. For example, if the correlation is positive in a two-part model for the costs of pharmaceuticals, this suggests that an intervention targeted at reducing the likelihood to use prescription drugs will also tend to reduce the consumption of drugs among the users. In contrast, a model that ignores the cross-part correlation and estimates Parts I and II separately (e.g., a marginal two-part model) may also underestimate the corresponding impact of the intervention.

The model considered in this paper is motivated by a pharmacy cost dataset, with physician representing the unit of clustering. It is easily transferable to other pharmacy-related economic analyses, where clustering occurs within different managed care organizations, pharmacy benefit managers or Medicare Part D plans. The model can also be used in other utilization-based analyses, such as in analyses of elective surgery costs that reflect both the variability among a surgeon's threshold for performing a given procedure and the variability in costs associated with operative and post-operative care. In principle, it is straightforward to extend the proposed model to the multi-level setting, e.g., patients clustered within physicians who are further clustered within hospitals or physician networks; see, for example, Liu, Ma and Johnson (2008), who proposed a multi-level two-part random effects model using a log-normal distribution for Part II in the analysis of data from an alcohol dependence study. Another useful

direction for extension includes the analysis of longitudinal cost data, in which costs are accumulated at different time intervals that are correlated for the same subject. For example, Cooper et al. (2007) studied the prediction of costs due to disease accrued over time, using Bayesian Markov Chain Monte Carlo methods. They applied their method to repeated measures of annual medical cost of early inflammatory polyarthritis. Liu, Conaway, Knaus, and Bergin (2008) proposed a four-part random effects model for analyzing longitudinal monthly medical costs of heart failure patients, further separating outpatient and inpatient costs, as in Duan et al. (1983).

Acknowledgments

The authors thank Drs. Min Zhang, Anirban Basu, and Willard Manning for helpful discussions, and Mr. Bob Brundage for data preparation. This research was supported by AHRQ grant R03 HS016543 and NIAAA grant RC1 AA019274.

References

- Abramowitz, M.; Stegun, I. Handbook of Mathematical Functions. New York: Dover; 1972.
- Albert PS. Letter to the editor. *Biometrics* 2005;47:879–881. [PubMed: 16135045]
- Anderson DA, Aitkin M. Variance component models with binary response: interviewer variability. *Journal of the Royal Statistical Society Series B* 1985;47:203–210.
- Aitchison J. On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association* 1955;50:901–908.
- Basu A, Rathouz PJ. Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics* 2005;6:93–109. [PubMed: 15618530]
- Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993;88:9–25.
- Cooper NJ, Lambert PC, Abrams KR, Sutton AJ. Predicting costs over time using Bayesian Markov Chain Monte Carlo methods: an application to early inflammatory polyarthritis. *Health Economics* 2007;16:37–56. [PubMed: 16981192]
- Cowen ME, Duseau DJ, Toth BG, Guisinger C, Zodet MW. Case mix adjustment of managed care claims data using the Clinical Classification for Health Policy Research (CCHPR) method. *Medical Care* 1998;36:1108–1113. [PubMed: 9674627]
- Cowen ME, Strawderman RL. Quantifying the physician contribution to managed care pharmacy expenses: a mixed effects approach. *Medical Care* 2002;40:650–661. [PubMed: 12187179]
- Crisp A, Burrige J. A note on nonregular likelihood functions in heteroscedastic likelihood functions. *Biometrika* 1994;81:585–587.
- Duan N. Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association* 1983;78:605–610.
- Duan N, Manning WG, Morris C, Newhouse JP. A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics* 1983;1:115–126.
- Duan N, Manning WG, Morris CN, Newhouse JP. Choosing between the sample-selection model and the multi-part model. *Journal of Business and Economic Statistics* 1984;2:283–289.
- Enami, K.; Mullahy, J. Tobit at fifty: a brief history of Tobin's remarkable estimator, of related empirical methods, and of limited dependent variable econometrics in health economics. 2008. Working Paper 14512 <http://www.nber.org/papers/w14512>
- Goldstein H, Rasbash J. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A* 1996;159:505–513.
- Golub GH, Welsch JH. Calculation of Gaussian quadrature rules. *Mathematics of Computation* 1969;23:221–230.
- Hald A. Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Skandinavisk Aktuarietidskrift* 1949;32:119–134.
- Heckman JJ. The common structure of statistical models of truncation, sample selection, and limited dependent variables. *Annals of Economic and Social Measurement* 1976;5:120–137.

- Heckman JJ. Sample selection bias as a specification error. *Econometrica* 1979;47:153–161.
- Hedeker D, Gibbons RD. A random-effects ordinal regression model for multilevel analysis. *Biometrics* 1994;50:933–944. [PubMed: 7787006]
- Lange KL, Little RJA, Taylor JMG. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association* 1989;84:881–896.
- Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD.; Schabenberger, O. *SAS for Mixed Model*. 2. Cary, NC: SAS Institute Inc; 2006.
- Liu L, Conaway M, Knaus W, Bergin J. A random effects four-part model for longitudinal medical costs. *Computational Statistics and Data Analysis* 2008;52:4458–4473.
- Liu L, Ma JZ, Johnson BA. A multi-level two-part random effects model, with application to an alcohol dependence study. *Statistics in Medicine* 2008;27:3528–3539. [PubMed: 18219701]
- Liu Q, Pierce DA. A note on Gaussian-Hermite quadrature. *Biometrika* 1994;81:624–629.
- Lu SE, Lin Y, Shih WJ. Analyzing excessive no changes in clinical trials with clustered data. *Biometrics* 2004;60:257–267. [PubMed: 15032797]
- Manning WG. The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics* 1998;17:283–295. [PubMed: 10180919]
- Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics* 2005;20:465–488. [PubMed: 15811539]
- Manning WG, Duan N, Rogers WH. Monte-Carlo evidence on the choice between sample selection and 2-part models. *Journal of Econometrics* 1987;35:59–82.
- Manning, WG.; Morris, CN.; Newhouse, JP.; Orr, LL.; Duan, N.; Keeler, EB.; Leibowitz, A.; Marquis, KH.; Marquis, MS.; Phelps, CE. A two-part model of the demand for medical care: preliminary results from the health insurance study. In: van der Gaag, J.; Perlman, M., editors. *Health, Economics, and Health Economics*. North-Holland: 1981. p. 103-123.
- McCulloch CE. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* 1997;92:162–170.
- Molenberghs, G.; Verbeke, G. *Models for Discrete Longitudinal Data*. New York: Springer; 2005.
- Mullahy J. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics* 1998;17:247–281. [PubMed: 10180918]
- Nelson KP, Lipsitz SR, Fitzmaurice GM, Ibrahim J, Parzen M, Strawderman R. Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. *Journal of Computational and Graphical Statistics* 2006;15:39–57.
- Olsen MK, Schafer JL. A two-part random effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association* 2001;96:730–745.
- Pinheiro JC, Bates DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* 1995;4:12–35.
- Raudenbush SW, Yang M, Yosef M. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics* 2000;9:141–157.
- Robert, CP. *From Decision-Theoretic Foundations to Computational Implementation*. 2. New York: Springer; 2007. *The Bayesian Choice*. (paperback)
- Tierney L, Kadane JB. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* 1986;81:82–86.
- Tooez JA, Grunwald GK, Jones RH. Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research* 2002;11:341–355. [PubMed: 12197301]
- van de Ven, WP.; van Praag, BM. Risk aversions of deductibles in private health insurance: application of an adjusted tobit model to family health care expenditures. In: van der Gaag, J.; Perlman, M., editors. *Health, Economics, and Health Economics*. Amsterdam: North Holland: 1981.
- Wolfinger R. Laplace's approximation for nonlinear mixed models. *Biometrika* 1993;80:791–795.
- Wooldridge, JM. *Econometric Analysis of Cross-Sectional and Panel Data*. Boston: MIT Press; 2002.
- Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988;44:1049–1060. [PubMed: 3233245]

- Zhang M, Strawderman RL, Cowen ME, Wells MT. Bayesian inference for a two-part hierarchical model: An application to profiling providers in managed health care. *Journal of the American Statistical Association* 2006;101:934–945.
- Zhou X-H, Stroupe KT, Tierney WM. Regression analysis of health care charges with heteroscedasticity. *Applied Statistics* 2001;50:303–312.
- Zhou X-H, Lin H, Johnson E. Non-parametric heteroscedastic transformation regression models for skewed data with an application to health care costs. *Journal of the Royal Statistical Society, Series B. Statistical Methodology* 2008;70:1029–1047.

Appendix

Adaptive Gaussian quadrature

$$L(\alpha, \beta, \delta, D) = \prod_{i=1}^n L_i(\alpha, \beta, \delta, D),$$

where $L_i(\alpha, \beta, \delta, D)$ denotes the contribution of cluster i (i.e., for the data $\mathbf{O}_i$). This computation requires that we be able to compute the following functions of α , β , δ , and D :

$$L_i(\alpha, \beta, \delta, D) \propto \int \exp(l_1(r_i) + l_2(r_i)) p(r_i) dr_i.$$

Suppose we want to maximize the following likelihood

$$L_i = \int \prod_{j=1}^{n_i} f(y_{ij} | X_{ij}, a_i) p(a_i) da_i$$

where a_i is the random effect with density $p(\cdot)$. L_i can be approximated by a weighted average of the integrand assessed at Q pre-specified quadrature points u_q ($q = 1, 2, \dots, Q$) over the random effects a_i (Pinheiro and Bates 1995, Liu and Pierce 1994), i.e.,

$$L_i \approx \sum_{q=1}^Q \prod_{j=1}^{n_i} f(y_{ij} | X_{ij}, u_q) p(u_q) w_q$$

with $u_q = \sqrt{2} z_q$ and $w_q = \sqrt{2} \eta_q \exp(z_q^2)$, where η_q and z_q can be obtained from tables (Abramowitz and Stegun 1972) or algorithms (Golub and Welsch 1969).

The implementation of Gaussian quadrature techniques has been incorporated in SAS Proc NLMIXED (Littell et al. 2006). For our model, a general log likelihood function composed by SAS programming statements can be conveniently specified and maximized, by option “general” in the “Model” statement in Proc NLMIXED. A sample SAS code is shown below.

In this dataset, “cost” is the response variable, “anycost” is the indicator of cost being positive, “z” is the only covariate, “id” is the physician ID. Random effects “a” and “b” are present in parts I and II, respectively.

Example SAS code

```

/* Invoke proc nlmixed using adaptive GQ with qpoints=5 nodes */
proc nlmixed data=cost qpoints=5;

/* define initial values and bounds */

parms alpha0=-1 alpha1=1 beta0=1 beta1=1 delta0=1 delta1=2 var1=1
var2=1 cov12=.5 k=2;

bounds var1 var2 >=0;

/* Part I log-likelihood */

teta=alpha0 + a + alpha1 * z;

expteta=exp(teta);

p=expteta/(1+expteta);

if anycost=0 then loglik=log(1-p);

/* Part II log-likelihood */

if anycost=1 then do;

mu=beta0 + b + beta1 * z; /* Mean of gen. gamma dist. */

sigma=exp((delta0 + delta1 *z)/2); /* Scale of gen. gamma dist. */

eta=abs(k) ** (-2);

u=sign(k)*(log(cost)-mu)/sigma;

value1=eta *log (eta) - log(sigma) -.5 * log(eta) - lgamma(eta);

loglik=log(p) + value1 + u *sqrt(eta) - eta * exp(abs(k)* u);

end;

/* fit the model above */

model cost ~ general(loglik);

random a b ~ normal([0, 0], [var1, cov12, var2]) subject=id;

/* generate empirical Bayes estimates for random effects
and store them in SAS datasets data1 and data2; */

predict a out=data1;

predict b out=data2;

run;

```

Table 1

Simulation Results

Parameter	Proposed model					Reduced model A					Reduced model B				
	Est	SE	SEM	CP	Est	SE	SEM	CP	Est	SE	SEM	SE	SEM	CP	CP
$\alpha_0 = -1$	-1.005	0.123	0.122	95.0%	-1.007	0.123	0.122	95.4%	-1.004	0.122	0.122	0.122	0.122	95.0%	95.0%
$\alpha_1 = 0.5$	0.502	0.163	0.168	95.1%	0.501	0.164	0.168	95.2%	0.502	0.163	0.168	0.163	0.168	95.1%	95.1%
$\beta_0 = 1$	1.010	0.230	0.245	96.1%	-1.168	0.379	0.357	0.0%	0.672	0.277	0.313	0.277	0.313	85.4%	85.4%
$\beta_1 = 1$	0.998	0.464	0.471	94.3%	-2.669	0.700	0.691	0.0%	2.116	0.398	0.410	0.398	0.410	21.6%	21.6%
$\delta_0 = 1$	0.980	0.205	0.211	96.3%	2.448	0.196	0.113	0.0%							
$\delta_1 = 2$	1.993	0.233	0.233	95.6%	1.991	0.324	0.181	72.0%							
$\kappa = 2$	2.026	0.212	0.208	93.8%					2.224	0.257	0.255	0.257	0.255	93.6%	93.6%
$d_{11} = 1$	1.007	0.168	0.168	93.6%	1.010	0.171	0.167	94.0%	1.007	0.168	0.168	0.168	0.168	93.5%	93.5%
$d_{12} = .5$	0.494	0.187	0.182	94.5%	0.498	0.303	0.284	95.9%	0.505	0.197	0.195	0.197	0.195	95.8%	95.8%
$d_{22} = 1$	0.987	0.334	0.326	92.6%	1.094	0.717	0.774	97.9%	1.162	0.369	0.379	0.369	0.379	96.5%	96.5%

The proposed model corresponds to the logistic-heteroscedastic generalized gamma model described in Section 2.2. The reduced models correspond to submodels of the proposed model. Specifically, reduced model A assumes a lognormal distribution with heteroscedasticity (i.e., $\kappa \rightarrow 0$) and reduced model B assumes a generalized gamma distribution without heteroscedasticity (i.e., $\delta(-1) = 0$). Est denotes the mean of the parameter estimates (600 replicates); SE denotes the standard error of the simulated parameter estimates; SEM denotes the sampling mean of the standard error estimate; CP denotes the coverage probability of the corresponding 95% confidence interval.

Table 2

Overall covariate Effect on the logarithm of the mean cost

	$X_1 = 0.25$	$X_1 = 0.5$	$X_1 = 0.75$
True Value	2.636	3.391	4.465
Proposed Model	2.610 (1.834, 3.362)	3.359 (2.490, 4.140)	4.432 (3.281, 5.636)
Reduced Model A	16.500 (12.550, 20.813)	28.920 (19.018, 40.120)	49.892 (28.449, 79.201)
Reduced Model B	2.456 (1.670, 3.289)	2.443 (1.652, 3.259)	2.429 (1.634, 3.247)
Marginal Model	2.567 (1.657, 3.383)	3.438 (2.506, 4.367)	4.629 (3.409, 6.093)

The estimates above are averaged over the random Effects and computed using numerical integration. The numbers in parentheses are the simulated 2.5% and 97.5% percentiles of the corresponding sampling distribution obtained using 600 simulated datasets.

Table 3

Pharmacy Cost Data: Regression Parameters

	Proposed model				Reduced model A				Reduced model B			
	Est	SE	P-value	Est	SE	P-value	Est	SE	Est	SE	P-value	P-value
Part I												
Intercept	0.455	0.036	< 0.0001	0.459	0.036	< 0.0001	0.462	0.036	0.462	0.036	< 0.0001	< 0.0001
Infection	1.383	0.024	< 0.0001	1.383	0.024	< 0.0001	1.383	0.024	1.383	0.024	< 0.0001	< 0.0001
Diabetes	1.156	0.085	< 0.0001	1.158	0.085	< 0.0001	1.157	0.085	1.157	0.085	< 0.0001	< 0.0001
Affective	1.294	0.107	< 0.0001	1.296	0.107	< 0.0001	1.295	0.107	1.295	0.107	< 0.0001	< 0.0001
Neuropsychiatry	0.911	0.247	0.0003	0.900	0.246	0.0003	0.900	0.246	0.900	0.246	0.0003	0.0003
Epilepsy	1.022	0.188	< 0.0001	1.009	0.187	< 0.0001	1.007	0.187	1.007	0.187	< 0.0001	< 0.0001
HIV	0.434	0.464	0.35	0.432	0.463	< 0.0001	0.428	0.464	0.428	0.464	0.36	0.36
Anxiety	1.009	0.066	< 0.0001	1.009	0.066	< 0.0001	1.009	0.066	1.009	0.066	< 0.0001	< 0.0001
HTN_LIPD	0.935	0.037	< 0.0001	0.934	0.037	< 0.0001	0.933	0.037	0.933	0.037	< 0.0001	< 0.0001
Heart	0.784	0.115	< 0.0001	0.776	0.115	< 0.0001	0.776	0.115	0.776	0.115	< 0.0001	< 0.0001
Asthcopd	1.145	0.055	< 0.0001	1.147	0.055	< 0.0001	1.146	0.055	1.146	0.055	< 0.0001	< 0.0001
Collagen	1.175	0.193	< 0.0001	1.179	0.193	< 0.0001	1.178	0.193	1.178	0.193	< 0.0001	< 0.0001
Patient Age in 10 yrs	0.202	0.008	< 0.0001	0.202	0.008	< 0.0001	0.202	0.008	0.202	0.008	< 0.0001	< 0.0001
Male Physician	-0.286	0.043	< 0.0001	-0.290	0.043	< 0.0001	-0.293	0.043	-0.293	0.043	< 0.0001	< 0.0001
Diabetes * HTN_LIPD	-0.521	0.153	0.0008	-0.524	0.153	0.0007	-0.524	0.153	-0.524	0.153	< 0.0001	< 0.0001
Heart * HTN_LIPD	-0.263	0.172	0.13	-0.253	0.172	0.14	-0.251	0.172	-0.251	0.172	0.15	0.15
Heart * Diabetes	-0.603	0.250	0.02	-0.595	0.251	0.02	-0.594	0.251	-0.594	0.251	0.02	0.02
Affective * Anxiety	-0.673	0.213	0.002	-0.675	0.213	0.002	-0.676	0.213	-0.676	0.213	0.002	0.002
Part II												
Intercept	5.153	0.029	< 0.0001	4.627	0.031	< 0.0001	5.113	0.030	5.113	0.030	< 0.0001	< 0.0001
Infection	0.324	0.014	< 0.0001	0.371	0.015	< 0.0001	0.306	0.014	0.306	0.014	< 0.0001	< 0.0001
Diabetes	1.097	0.031	< 0.0001	1.192	0.035	< 0.0001	1.085	0.037	1.085	0.037	< 0.0001	< 0.0001
Affective	1.284	0.035	< 0.0001	1.370	0.040	< 0.0001	1.334	0.042	1.334	0.042	< 0.0001	< 0.0001
Neuropsychiatry	1.530	0.116	< 0.0001	1.488	0.124	< 0.0001	1.907	0.096	1.907	0.096	< 0.0001	< 0.0001

	Proposed model				Reduced model A				Reduced model B			
	Est	SE	P-value	Est	SE	P-value	Est	SE	Est	SE	P-value	P-value
Epilepsy	0.930	0.064	< 0.0001	0.997	0.071	< 0.0001	1.044	0.075	1.044	0.075	< 0.0001	< 0.0001
HIV	2.096	0.218	< 0.0001	2.094	0.239	< 0.0001	2.310	0.209	2.310	0.209	< 0.0001	< 0.0001
Anxiety	0.791	0.029	< 0.0001	0.843	0.032	< 0.0001	0.833	0.031	0.833	0.031	< 0.0001	< 0.0001
HTN_LIPD	0.658	0.017	< 0.0001	0.717	0.019	< 0.0001	0.638	0.019	0.638	0.019	< 0.0001	< 0.0001
Heart	0.751	0.046	< 0.0001	0.799	0.050	< 0.0001	0.776	0.051	0.776	0.051	< 0.0001	< 0.0001
Asthcopd	0.673	0.019	< 0.0001	0.764	0.021	< 0.0001	0.706	0.022	0.706	0.022	< 0.0001	< 0.0001
Collagen	1.234	0.060	< 0.0001	1.287	0.066	< 0.0001	1.336	0.067	1.336	0.067	< 0.0001	< 0.0001
Patient Age in 10 yrs	0.364	0.006	< 0.0001	0.376	0.006	< 0.0001	0.366	0.006	0.366	0.006	< 0.0001	< 0.0001
Male Physician	-0.178	0.032	< 0.0001	-0.197	0.036	< 0.0001	-0.186	0.033	-0.186	0.033	< 0.0001	< 0.0001
Diabetes * HTN_LIPD	-0.382	0.045	< 0.0001	-0.415	0.051	< 0.0001	-0.364	0.057	-0.364	0.057	< 0.0001	< 0.0001
Heart * HTN_LIPD	-0.253	0.055	< 0.0001	-0.262	0.061	< 0.0001	-0.270	0.066	-0.270	0.066	< 0.0001	< 0.0001
Heart * Diabetes	-0.176	0.062	0.005	-0.187	0.071	0.009	-0.165	0.083	-0.165	0.083	0.05	0.05
Affective * Anxiety	-0.607	0.064	< 0.0001	-0.622	0.070	< 0.0001	-0.677	0.077	-0.677	0.077	< 0.0001	< 0.0001

Table 4

Pharmacy Cost Data: Heteroscedasticity and Variance Components

	Proposed model			Reduced model A			Reduced model B		
	Est	SE	P-value	Est	SE	P-value	Est	SE	P-value
Heteroscedasticity									
Intercept	0.896	0.016	< 0.0001	1.060	0.015	< 0.0001			
Infection	-0.163	0.015	< 0.0001	-0.180	0.014	< 0.0001			
Diabetes	-0.452	0.039	< 0.0001	-0.332	0.037	< 0.0001			
Affective	-0.329	0.044	< 0.0001	-0.231	0.042	< 0.0001			
Neuropsychiatry	0.512	0.101	< 0.0001	0.490	0.096	< 0.0001			
Epilepsy	-0.213	0.078	0.007	-0.137	0.075	0.07			
HIV	0.185	0.215	0.39	0.256	0.207	0.22			
Anxiety	-0.169	0.032	< 0.0001	-0.071	0.031	0.02			
HTN_LIPD	-0.316	0.020	< 0.0001	-0.196	0.018	< 0.0001			
Heart	-0.147	0.052	0.005	-0.158	0.049	0.002			
Asthcopd	-0.299	0.023	< 0.0001	-0.253	0.022	< 0.0001			
Collagen	-0.116	0.071	0.10	-0.081	0.067	0.23			
Patient Age in 10 yrs	-0.064	0.006	< 0.0001	-0.018	0.006	0.002			
Male Physician	0.039	0.016	0.01	0.016	0.015	0.28			
Diabetes * HTN_LIPD	0.156	0.060	0.01	0.094	0.057	0.10			
Heart * HTN_LIPD	-0.039	0.068	0.57	-0.018	0.065	0.78			
Heart * Diabetes	0.062	0.087	0.47	0.031	0.083	0.71			
Affective * Anxiety	0.031	0.080	0.70	-0.137	0.077	0.08			
Variance Components									
d_{11}	0.045	0.008	< 0.0001	0.045	0.008	< 0.0001	0.045	0.008	< 0.0001
d_{22}	0.030	0.005	< 0.0001	0.038	0.006	< 0.0001	0.032	0.005	< 0.0001
d_{12}	0.024	0.005	< 0.0001	0.029	0.005	< 0.0001	0.025	0.005	< 0.0001
κ	0.634	0.012	< 0.0001				0.569	0.012	< 0.0001

The estimate of σ obtained under reduced model B is 1.427 (SE=0.006). The regression coefficients in the heteroscedasticity model are all zero under reduced model B; hence, no values are reported. Reduced model A, which allows for heteroscedasticity, is obtained as a limiting case of the proposed model (i.e., at $\kappa = 0$); hence, no value is reported.

Table 5
Pharmacy Cost Data: Overall Covariate Effect (95% CI) on Logarithm of Mean Cost

	Proposed model	Reduced model A	Reduced model B	Marginal model
Infection	0.561 (0.525, 0.597)	0.479 (0.427, 0.531)	0.651 (0.616, 0.686)	0.570 (0.524, 0.608)
Neuropsychiatry	2.272 (1.972, 2.572)	2.663 (2.152, 3.174)	2.169 (1.956, 2.382)	2.258 (2.035, 2.545)
Epilepsy	1.079 (0.925, 1.233)	1.097 (0.857, 1.337)	1.327 (1.165, 1.489)	1.089 (0.923, 1.249)
HIV	2.390 (1.827, 2.953)	2.662 (1.736, 3.588)	2.455 (1.964, 2.946)	2.488 (1.693, 3.127)
Asthcopd	0.795 (0.747, 0.843)	0.750 (0.680, 0.820)	1.014 (0.964, 1.064)	0.784 (0.729, 0.833)
Collagen	1.469 (1.323, 1.615)	1.489 (1.263, 1.715)	1.650 (1.504, 1.796)	1.461 (1.325, 1.623)
Male Physician	-0.270 (-0.355, -0.185)	-0.296 (-0.397, -0.195)	-0.310 (-0.396, -0.224)	-0.244 (-0.325, -0.181)
Patient Age in 10 yrs	0.390 (0.377, 0.403)	0.420 (0.400, 0.440)	0.435 (0.422, 0.448)	0.395 (0.377, 0.410)
Diabetes	1.145 (1.073, 1.217)	1.095 (0.988, 1.202)	1.395 (1.315, 1.475)	1.149 (1.069, 1.236)
Affective	1.414 (1.331, 1.497)	1.405 (1.278, 1.532)	1.666 (1.577, 1.755)	1.409 (1.307, 1.515)
Anxiety	0.964 (0.896, 1.032)	1.027 (0.921, 1.133)	1.117 (1.050, 1.184)	0.971 (0.903, 1.028)
HTN_LIPD	0.732 (0.690, 0.774)	0.730 (0.667, 0.793)	0.907 (0.864, 0.950)	0.734 (0.684, 0.780)
Heart	0.890 (0.780, 1.000)	0.824 (0.662, 0.986)	1.011 (0.899, 1.123)	0.907 (0.764, 1.072)
Diabetes + HTN_LIPD	1.410 (1.329, 1.491)	1.352 (1.228, 1.475)	1.723 (1.626, 1.819)	1.414 (1.336, 1.500)
Heart + HTN_LIPD	1.224 (1.135, 1.313)	1.158 (1.022, 1.293)	1.497 (1.394, 1.599)	1.229 (1.131, 1.323)
Diabetes + Heart	1.709 (1.557, 1.860)	1.609 (1.386, 1.832)	2.032 (1.854, 2.210)	1.751 (1.561, 1.937)
Affect + Anxiety	1.571 (1.466, 1.676)	1.451 (1.299, 1.603)	1.860 (1.737, 1.982)	1.565 (1.457, 1.676)

The parameter estimates for the proposed model and reduced models involve averaging over the random effects and are computed using numerical integration; this is not required for the marginal model of Section 2.4. The numbers in parentheses are 95% confidence intervals. In light of the small magnitude of the variance component $d[1] = 0.045$, the point estimates derived from the random effects models are close to those obtained by conditioning on the random effect $a = 0$; thus, for computational simplicity, the corresponding confidence intervals are calculated conditionally on $a = 0$ and obtained using the `estimat` statement in SAS Proc NLMIXED. The confidence intervals for the marginal model are obtained using a cluster bootstrap sampling scheme with 600 resampled datasets.