An analytics approach to designing patient centered medical homes

Saeede Ajorlou · Issac Shams · Kai Yang

Received: 16 December 2013 / Accepted: 8 June 2014 / Published online: 19 June 2014
© Springer Science+Business Media New York 2014

Abstract Recently the patient centered medical home (PCMH) model has become a popular team based approach focused on delivering more streamlined care to patients. In current practices of medical homes, a clinical based prediction frame is recommended because it can help match the portfolio capacity of PCMH teams with the actual load generated by a set of patients. Without such balances in clinical supply and demand, issues such as excessive under and over utilization of physicians, long waiting time for receiving the appropriate treatment, and non-continuity of care will eliminate many advantages of the medical home strategy. In this paper, by using the hierarchical generalized linear model with multivariate responses, we develop a clinical workload prediction model for care portfolio demands in a Bayesian framework. The model allows for heterogeneous variances and unstructured covariance matrices for nested random effects that arise through complex hierarchical care systems. We show that using a multivariate approach substantially enhances the precision of workload predictions at both primary and non primary care levels. We also demonstrate that care demands depend not only on patient demographics but also on other utilization factors, such as length of stay. Our analyses of a recent data from Veteran Health Administration further indicate that risk adjustment for patient health conditions can considerably improve the prediction power of the model.

Keywords Healthcare analytics · Patient centered medical home · Bayesian · Generalized linear model · Multivariate · Multilevel · Healthcare workforce

1 Introduction

Health care delivery is a complex multilevel system in which primary care is the base level and acts as a principal point of consultation for patients. The traditional format of primary care is mainly featured by primary care physicians (PCP), in which each PCP has a designated set of patients, called a patient panel. In current practices of most providers, the panel is simply decided by a predetermined maximum size; that is when the quota is reached, no more patients will be added [1, 2]. Typical panel sizes range from 1200 to 1600 patients. However, this number alone cannot reflect the actual healthcare workload generated in the panel. For example, a PCP with 1,200 young and healthy patients might be generally underutilized, while one with 1,200 elderly patients having multiple comorbidities may experience excessive workload, causing long delays in its panel appointment times and forcing patients to switch their PCPs.

It is found that many factors such as patient’s age, gender, and diagnostic codes can influence the required healthcare workload. Ostbye and colleagues [3] find that patients with different chronic diseases regularly have different visiting frequencies to their PCPs. Naessens and colleagues [4] discover that the number of chronic conditions in a patient will significantly affect clinical workload and medical cost. Potts and colleagues [5] propose a risk-standard method to adjust the panel size for each PCP calculating disease burden of each physician panel for six chronic diseases. However, there is no
Balasubramanian and colleagues [6] apply classification and regression trees (CART) to classify approximately 20,000 patients at the Mayo Clinic into 28 categories by using age and gender as the input attributes, so that each category has different workload patterns.

In recent years, the patient-centered medical home (PCMH) has been introduced as a prominent intervention to improving the US primary care systems with better-quality outcomes at lower costs [7]. This model consists of different health professionals grouped together to provide comprehensive, coordinated, accessible and cost effective care while maintaining high levels of quality and stability. Each team consists of a group of medical professionals such as primary care provider, registered nurse, nutritionist, social worker, and medical clerk that are well poised to provide many aspects of primary care. Theoretically, medical homes are composed of “joint principles” that ideally complement one another and feed into a comprehensive vision of appropriate primary care delivery. The principles are consisted of having a personal physician with an ongoing relationship, a whole person orientation care for all stages of life, a physician-directed medical practice taking responsibilities for all of the ongoing care, a coordinated and/or integrated care system across all elements of the care systems, a continuous emphasis on quality and safety, an enhanced access to care through such systems as open scheduling and expanded hours, and finally an appropriate payment system that recognize the added value provided to PCMH patients [8]. Augmented with modern health information technology, the PCMH is crafted to initiate numerous reforms in health care delivery and reimbursement systems [9].

As of 2007, there was some literature examining the prevalence and effectiveness of medical homes. For instance, Fisher [10] outlined some recommendations for the success of medical homes such as increasing effective communication and sharing of information across health care providers, broadening the medical performance measures to include patients’ experience with care and ordinary assessment of outcomes, and establishment of medical-home payment system that share savings among all providers involved. A survey by Commonwealth Fund of 3,535 US adults found that when they were provided with a medical home, racial and ethnic disparities in care access and quality were substantially reduced [11]. Furthermore, having a medical home was associated with more preventive screenings and better management of chronic conditions. The Centers for Medicare & Medicaid Services (CMS) planned to pursue Medicare pilot projects in 400 practices in 8 regional sites, and by 2009, twenty bills promoting the PCMH concept have been successfully introduced in 10 states [12]. Another study within the Group Health system in Seattle showed that a medical home prototype led to 29 % fewer emergency visits, 6 % fewer hospitalizations, and total savings of $10.30 per patient per month over a twenty-1 month period [13]. Bates and Bitton [14] indicated seven health information technology domains deemed to be critical for the success of the PCMH model including telehealth, measurement of quality and efficiency, care transitions, personal health records, and, most importantly, registries, team care, and clinical decision support for chronic diseases.

Practically, as of December 2009, there were about 26 pilot projects involving medical home being directed in 18 states. These consist of over 14,000 physicians and approximately 5 million patients [15]. Of interest, the Veterans Health Administration (VHA) launched a nationwide 3-year program in April 2010 to create PCMHs in more than 900 primary care clinics. Early results indicated dramatic improvements such as reducing the appointment waiting time from as long as 90 days down to 1 day and decreasing the percentage of inappropriate emergency department visits from 52 to 12 % [16].

However, there are difficulties in fully achieving the benefits of the PCMH in practice. It is found that much work still have to be done on the PCMH model before its potential cost savings can take place [12]. From an operations management point of view, one of the key success factors of any health delivery system is achieving a balance between supply and demand of care services. This issue is even more critical for the PCMH model since the clinical supply and demand is portfolio (vector) in nature. Unlike health demands, the supply of healthcare services can be treated as deterministic and calculated easily based on head counts and available service hours to be offered from all professional lines on an annual basis. Yet, the estimation of clinical workload portfolio based on key patient and provider attributes is a challenging task, and to our knowledge there is no related literature tackling this problem in a team based medical home perspective.

Looking retrospectively at data sets of our US Department of Veteran Affairs (VA) sponsored PCMH project, we observe some structural properties listed as follows:

- Patients (unit of analysis) are grouped within PCMH teams and within VA medical facilities with related patient-level, team-level, and also facility-level predictors. In addition, there exists significant heteroscedasticity within each level of hierarchy.
- The actual health workload (outcome) is captured with two variables, one for primary care and one for non-primary care, and these are correlated at some levels of hierarchy.
- Distributions of both outcomes are not normal at all levels of hierarchy.

In this paper, we develop a multivariate hierarchical based portfolio prediction model that takes into account postulated attributes from different levels such as disease types...
Designing patient centered medical homes

(patient-level), years of experience of the assigned provider (team-level), and zip-code based distance between the patient’s home and his/her assigned facility (facility-level). To the best of our knowledge, our work is the first attempt to develop such a clinical portfolio prediction model for medical homes within the OR/MS community. The approach is multivariate since there are two (possibly dependent) response variables, primary care and non-primary care workloads, modeled as probabilistic functions of some risk factors defined in the three hierarchical levels. One key advantage of our model is that it allows for passing heterogeneous variances and unstructured covariance matrices for the nested random effects as well as their interactions with response variables and covariates simultaneously. We perform a fully Bayesian approach for model fitting and inference that involves setting appropriate priors for input variables and sketching a fast and efficient Markov Chain Monte Carlo algorithm. We apply our proposal to real-world data from the VHA to produce findings that have key public and medical implications. Finally, we seek to screen highly contributing risk factors to primary care and non-primary care demand portfolio variations, since it would inform program analysts on areas more likely affecting the care portfolio balance.

The remainder of this paper is laid out as follows. Section 2 introduces our data sources and study variables. Section 3 describes the main methods including model specification and model fitting strategies. The analyses and results from a recent case study in the VHA are given in section 4 and 5, respectively. Some discussion points and future research directions are presented in Section 5.

2 Data source and study variables

According to National Center for Veterans Analysis and Statistics (NCVAS), VA operates the largest health care system in the USA with 23 geographically different regions (known as VISNs, or Veterans Integrated Service Networks) separated hierarchically within each VISN by level of care or type into different facilities such as VA medical centers (VAMC), Community Based Outpatient Clinic (CBOC), Vet Center (VC), and so forth. Within each facility, every VA primary care enrollee was assigned to an independent physician or non-physician PCP by a standard process-VA Primary Care Management Module. To ensure sufficient staffing and quality of care, each PCP was appointed a target panel size, taking into account the intensity of primary care visits and availability of resources such as supporting staff and capital.

In this study we collected outpatient data from a random sample of 888 different facilities (which corresponds to 130 VAMCs of all 23 VISNs) during FY11 quarter 3 to FY12 quarter 2. The period of 1 year is appropriate; according to the VA program professionals, the primary care population at each practice site is not subject to drastic change from 1 year to the next. The Decision Support System (DSS) and National Patient Care Database (NPCD) files of the VA Corporate Data Warehouse (CDW) were employed to extract demographic, socioeconomic, and other types of variables. In addition, due to its rigorous data validity and availability, we chose DRG (Diagnosis Related Group, 29th version) and its ACC (Aggregated Condition Category) codes for patient case-mix and risk adjustment measures in our predictive analytics [17].

Initially there were 82,000 randomly selected patients with 48 independent attributes coded. All patient visits to primary care and women’s health are assembled for a total capture period of 1 year. Visits from other primary care related clinics, such as Internal Medicine or Geriatric Primary Care, are excluded from the analysis. The two dependent variables are total primary care (PC) and non-primary care (Non-PC) Relative Value Units (or RVUs), and for each unique SSN, they are calculated by converting the primary care and non-primary care Current Procedural Terminology (or CPT) codes from all patient visits during the fiscal year (according to the Centers for Medicare and Medicaid Services model). Simply, the Non-PCRVU refers to all of the non-primary care workload during the year, which could be from one or many visits to outpatient specialty care, and the PCRVU is the primary care workload during the year from outpatient primary care. One advantage of using RVUs in our approach, as opposed to simple face-to-face visit counts, lies in its ability to further accommodate workloads generated by telephone encounters at the VHA. It is noted that the RVU can be seen as a comparable measure of value for care services used in the US Medicare reimbursement and is determined by assigning weight to factors such as personnel time, level of skill, and sophistication of equipment required to render patient services. The predictor variables include baseline demographic and socioeconomic attributes along with some medical factors such as whether the patient has insurance, to which VA facility the patient has been admitted, and so on.

To achieve a better picture of the data environment, we tentatively arranged all independent attributes into five groups: (1) Demographic: patient age, gender, and marital status; (2) Socioeconomic: whether the patient has any type of insurance plans such as Medicare or Medicaid, and his/her employment status; (3) Enrollment: priority levels that represent a combined score of patient service connection and income; (4) Utilization: VISN identifier, facility in which the patient got admitted, distance, PCMH team code, assigned provider position, assigned provider experience (years), number of times during the year that the patient changed his/her
 assigned provider (as denoted by ‘Changed provider count’), provider full time equivalent, and length of stay; and (5) Comorbidity and severity: care assessment need score or CAN score, and aggregated condition category (or ACC) codes. It should be noted that these variables remain the same for a patient during the fiscal year.

The enrollment priority levels range from 1-8 and are assigned based on the veteran’s severity of service-connected disabilities and VA income means test (VHA Handbook 1601A.03). Distance is calculated in miles between patient’s home zip code and the zip code of the facility he/she admitted, considering the latitude and longitude of the two locations. Records with a calculated distance greater than 240 miles were excluded and the remaining were converted into three levels. As defined earlier, variable ‘changed provider count’ can be a marker of unbalanced workload among PCPs and discontinuity of care received by patients. Length of stay (or LOS) displays the number of days spent admitted at a VA hospital. CAN score is the care assessment need score, which reflects the likelihood of admission or death within a specified time period. This score is commonly expressed as a percentile ranging from 0 (lowest risk) to 99 (highest risk) and it indicates how a VA patient is compared with other patients in terms of the likelihood of hospitalization or death. Each PCMH team has a unique 10-digit code throughout all VA medical systems nationwide. Currently all teams have the same number of professions within all VA centers. The number of PCMH teams and VA facilities in our data set are 6,051 and 287 respectively. ACC codes are determined based on the various ICD-9-CM (International Classification of Disease, ninth version, Clinical Modification) codes assigned to the patient at each visits during the whole fiscal year. They basically indicate the occurrence of a specific disease group, and they are not mutually exclusive categories, meaning that a patient may have more than one ACC during the fiscal year and most actually do.

3 Methods

In this section we first present the proposed prediction model and its exclusive properties for modeling the PCMH demand variations. Then we develop a Bayesian framework with novel prior specifications for parameter estimation and model inference.

3.1 Model specification

The PCMH data is hierarchically organized into three nested levels as shown in Fig. 1, where patients are grouped within PCMH teams, and teams are in turn nested within VA facilities. Note that PCMH teams are tied to facilities, i.e., a specific team cannot work at different facilities (teams are nested within facilities). Risk factors can be associated with the response variables at each level while patients from the same team (facility) may have more similar outcomes than patients chosen at random from different teams (facilities). For example, we can study the effects of age (patient-level), PCMH assigned provider’s experience (team-level), and type of hospital (facility-level) on the outcomes with nested sources of variability. This setting, in addition to health services research, may happen in many other applications such as educational studies where students are nested within schools and successively within school district. It has been shown that ignoring a level of hierarchy in a data can greatly influence the estimated variances and sensitivity, can seriously inflate Type I error rates [18], and also can result in errors in interpreting the results of statistical significance tests [19]. As such, multilevel statistical models have been proposed to appropriately analyze the hierarchical (correlated) nesting of data, taking into account the variability associated with each level of the hierarchy [19].

To simplify, we begin by creating a univariate 2-level generalized linear model (GLM) that predicts the primary care RVU (PCRVU) in each PCMH team with one patient-level (age) and one team-level (assigned provider’s experience) predictors. The level-1 model would look like

\[ y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \epsilon_{ij}, \]  \hspace{1cm} (1)

where \( y_{ij} \) is the primary care workload for patient \( i \) in PCMH team \( j \) with an exponential family density, \( \beta_{0j} \) is the average primary care workload generated in team \( j \), \( X_{ij} \) is the patient-level predictor (age) for patient \( i \) in team \( j \), and \( \beta_{1j} \) is its coefficient or slope. This way, we assume that each team has a different (varying) intercept coefficient and a varying slope coefficient. These team-specific coefficients can be specified as either fixed effects or random effects. Treating them as fixed effects, however, leads to a large number of parameters with often very poor estimation results. A more conservative way is to think of them as random variables being modeled by some (level-2) hyperparameters. The last term, \( \epsilon_{ij} \) is the patient-level error term which is assumed to be normally distributed with covariance structure \( R \). Unlike most methods in the literature, which suppose that the residual variation is the same at the 2-level (teams) and/or the upper levels of hierarchy, we allow unequal variations of the residual to be passed on various levels of the hierarchy and on different response variables.

The next step is to explain the variation of the level-1 regression coefficients by introducing explanatory variables at the team level like

\[ \beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j} \]  \hspace{1cm} (2)

\[ \beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}. \]
In this equation, $\gamma_{00}$ is the grand mean of primary care workload across patients and across PCMH teams, $\gamma_{10}$ is the average effect of the patient-level predictor (age) across all teams, $Z_j$ is the team-level predictor (assigned provider’s experience) for team $j$, $\gamma_{01}$ and $\gamma_{11}$ are its (level-2) intercept and slope regression coefficient, and the $u$-terms are random errors at the team level, which are assumed to be normally distributed with covariance $G$. Similar to the $R$-side covariance matrix, we let these level-2 random errors have unequal variances and also leave them free to be correlated with each other. It is worth pointing out that, $Z_j$ in the second line of (2) acts as a moderator for the relationship between workload and patient age at level-1 analysis; that is, the relationship varies according to the value of the moderator variable. Following the same logic, we can extend this model to add further hierarchies at the facility-level, at the regional level, and so on.

Now a multivariate generalization of this hierarchical GLM is proposed in which both primary care and non-primary care workloads are predicted simultaneously. There are several advantages of using a multivariate approach instead of univariate method [19, 20]. One is that the multivariate analysis can better control the type I error rate compared to carrying out a series of univariate statistical tests. Second, this approach can shrink the prediction interval of the dependent variables to a large extent when compared to predicting one of them in isolation. Also using a multivariate scheme, the covariance structure of the responses can be decomposed over the separate levels of hierarchy, which can be of much value for multilevel factor analysis.

Suppose we have $P$ response variables and let $Y_{hijk}$ be the workload on outcome $h$ (primary care or non-primary care workload here) of patient $i$ in PCMH team $j$ and facility $k$. Here we put the responses on the lowest level of hierarchy, and represent the different outcome variables by defining $P$ dummy variable $d_{phijk}$ taking value of 1 whenever $p=h$ and zero otherwise. Then we formulate the lowest level as

$$Y_{hijk} = \pi_{pijk}d_{phijk} + e_{pijk}, \forall p \in \{1, \ldots, P\}$$

in which a separate index is utilized for denoting the dependent variable of interest. It is noted that with this approach one can fit different intercepts and slopes for different response variables and allow them to vary across any levels of the hierarchy. Following (2), at the team level, we can have

$$\beta_{p0jk} = \gamma_{p00k} + \beta_{p01k}Z_{jk} + u_{p0jk}$$

$$\beta_{p1jk} = \gamma_{p10k} + \beta_{p11k}Z_{jk} + u_{p1jk}$$

where we introduce our 2-level predictors (level-1 moderators) along with random intercepts and slopes and finally link them to the facility level equations by

$$\gamma_{p00k} = \lambda_{p000} + \lambda_{p001}W_k + u_{p00k}$$

$$\gamma_{p01k} = \lambda_{p010} + \lambda_{p011}W_k + u_{p01k}$$

$$\gamma_{p10k} = \lambda_{p100} + \lambda_{p101}W_k + u_{p10k}$$

$$\gamma_{p11k} = \lambda_{p110} + \lambda_{p111}W_k + u_{p11k}$$

Keeping on this way, one can straightforwardly extend the model to include more predictors at each level and study the effects of fixed and random parameters at any given point.
Another advantage of such modeling is that we can impose an equality constraint across all response variables to build a specific relation with certain effects. For example, we can force level-1 regression coefficients for \( p=1 \) (primary care workload) and \( p=2 \) (non-primary care workload) to be equal by adding the constraint \( \beta_{1jk} = \beta_{2jk} \). This makes the new model nested within the original model, and thus we can test whether simplifying the model is justified, using a chi-square test on deviances. Plus, if the predictor has random components attached to it, a similar approach would apply to the random part of the model.

Next, we specify the structure of random components in the model. As shown, we have two random parts in our method: first is the level-1 residual errors as appear in (3) by \( e \)-terms, and second relates to (higher level) varying intercepts and slopes introduced by \( u \)-terms in (4) and (5). We denote the covariance matrix of the former as \( R \) and the latter as \( G \) and then assume that both are normally distributed with

\[
E \left[ \begin{bmatrix} u \\ e \end{bmatrix} \right] = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{6}
\]

\[
Var \left[ \begin{bmatrix} u \\ e \end{bmatrix} \right] = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix},
\]

where bold face denote matrices. In the formula \( E(\cdot) \) and \( Var(\cdot) \) stand for mathematical expectation and variance, respectively. The current specification indicates that the residual and random parameters are independent having zero means. Generally \( G \) and \( R \) matrices are large and square with dimensions equal to the number of random coefficients and residuals. While several structures such as spatial or compound symmetry can be thought to formulate those, here we propose using an unstructured parameterization by the Kronecker (direct) product (see [21] for the use of Kronecker product in modeling covariance structures). The Kronecker product, denoted by \( \otimes \), is a product operation on two matrices that results in a block matrix. If \( A \) is an \( m \times n \) and \( B \) is a \( p \times q \) matrix, then \( A \otimes B \) is a \( mp \times nq \) block matrix of the form

\[
A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}.
\]

To explain the basic idea we take an example in our case study when modeling the \( G \) matrix, but the same logic is applied to modeling the \( R \) matrix. Suppose, we record two workload measures (one for primary care and one for non-primary care) for the patients in a sample of VA facilities. Generally we may expect dependence between the workloads in the sample. If we denote the random effects for the VA facilities in primary care workload by \( \nu_1 \), and the random effects for the same set of facilities in non-primary care workload by \( \nu_2 \), then the following facility variance component can be passed to the \( G \) matrix

\[
V_{\text{Facility}} = \begin{bmatrix} \sigma^2_{\nu_1} & \sigma_{\nu_1\nu_2} \\ \sigma_{\nu_2\nu_1} & \sigma^2_{\nu_2} \end{bmatrix} \otimes I\tag{7}
\]

Here the diagonal elements represent variances for the two workload measures, and the off-diagonals capture the covariances between them. The identity matrix in the Kronecker product is of dimension \( f \times f \), where \( f \) is the total number of VA facilities. Here, it implied that the facilities are independent. However, if there is evidence that facilities are not independent, such as when data are collected on related facilities within a geographic region, then the identity matrix should be replaced by a spatial structure where the covariance of two facilities depends on the distance between them. Likewise, we can model the PCMH team variations by \( V_{\text{Team}} \) with a same structure as (7). Then the complete form of the \( G \) matrix is given as

\[
G = \begin{bmatrix} V_{\text{Facility}} & 0 \\ 0 & V_{\text{Team}} \end{bmatrix}.
\tag{8}
\]

3.2 Model fitting and inference

Before describing model inferences, we give an equivalent description of our proposal. By substituting Eq. 2 into Eq. 1 and rearranging the terms, we have

\[
y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} X_{ij} Z_j + u_{1j} X_{ij} + u_{0j} + e_{ij},
\]

in which two distinct segments can be implied: the first is \([\gamma_{00}+\gamma_{10}X_{ij}+\gamma_{01}Z_j+\gamma_{11}X_{ij}Z_j]\), which we call the deterministic part, and the second is \([u_{1j}X_{ij}+u_{0j}+e_{ij}]\), which we call the stochastic part. That way, the moderator effect of (2) is expressed as \( X_{ij}Z_j \) and the multiplication \( u_{1j}X_{ij} \) directly reveals that the error is different for different values of \( X_{ij} \) (heteroscedasticity). Taking a matrix form, we may rewrite the right-hand-side of (9) as \( \eta = X\gamma + We \), which resembles a mixed effect model format, where \( X \) and \( W \) are the design matrices for deterministic and stochastic parts.

In order to fit the model we need to accomplish two preliminary tasks: (1) specifying the probability density functions of the two response variables, and (2) identifying the appropriate link function relating the linear predictor to the mean of the distribution function. For the first task, the
standard Quantile-Quantile plot along with the maximum likelihood (ML) method is used, but one can also employ non-parametric techniques such as kernel density estimation. We examine different base densities such as gamma, log-normal, log-logistic and Cauchy, then judge the best choice as having the best graphical pattern in QQ plot and the biggest likelihood value simultaneously. To determine the appropriate link function, a range of classical options including log link and inverse link are evaluated by two goodness-of-fit measures, namely DIC and modified Hosmer-Lemeshow test [22].

We take a fully DIC approach for model fitting and inference that utilizes numerical integration techniques by Markov chain Monte Carlo. This is basically done in two steps: first determining suitable priors for the parameters of interest, and second employing a simulation-based integration technique, such as Metropolis-Hastings or slice sampling, to iteratively sample the posterior until convergence. Afterwards, generated samples are used to estimate the approximate expectations of the parameters. Here, for the deterministic coefficients, we use Gaussian priors with zero means and diagonal matrix of large variances ($1e+10$). To sample from $\eta$, since its full conditional density cannot be identified, we apply the Metropolis-Hastings update of Damlen et al. [23]. In order to update the parameter vector $[\gamma, \epsilon]$, the single-block Gibbs sampler of García-Cortés, Sorensen [24] is applied. For updating the variance structures $R$ and $G$, single-block Gibbs samples of Inverse-Wishart distribution is used, and the expected (co)variances and the degree of freedom can be specified [25]. For example, in our study we have two response variables, two components for the $G$ matrix that includes PCMH team and VA facilities, and one component for the $R$ matrix that involves residuals at the patient level. We scale the responses to have a unit variance before the analysis.

For these three components, we choose (7) as the default structures with
\[
\begin{bmatrix}
\frac{1}{3} & 0 \\
0 & \frac{1}{3}
\end{bmatrix}
\]
for their priors, implying the total variance is equally split across the three hierarchical levels with a priori independence between primary care and non-primary care workloads. The degree of freedom is set to two indicating that only two elements (which are diagonal elements here) need to be estimated (with the starting value of 1/3) for each response variable.

Usually the goodness-of-fit of Bayesian models can be assessed using the deviance information criterion (DIC), which is a Bayesian alternative to AIC and Schwarz criterion. The DIC can be calculated at different levels of hierarchy, and a smaller amount indicates a better fit to the data while compensating for model complexity. Here, we adopt the method of Spiegelhalter et al. [26] and assess the DIC for the lowest level of the hierarchy (that is patient level) over all iterations after the burn in phase of MCMC simulations. Based on a rule of thumb, we favor the model with lower DIC when the DIC reduction of more than 10 units is observed.

### 4 Analyses

We conduct several analyses to estimate the effect of risk factors on annual primary care and non-primary care workloads. Although various modeling strategies could be selected to fit the model, here we concentrate on model parsimony and goodness-of-fit (as measured by the DIC) to choose from alternative fittings. To this end, six models (Table 1) from basic to comprehensive are run sequentially and the outputs are reported for each step in order to provide insights for a particular objective. Further, to avoid overfitting, we perform stepwise selection for the deterministic covariates with probabilities to enter and stay of 0.15 and 0.1, respectively. Different functional forms of the covariates, such as logarithmic and power relations, as well as within-level interactions are evaluated too, but only the statistically significant ones are included. As an example, we analyze 12 pairs of ACC code interactions that are notable for co-occurring in patients with multiple chronic illnesses and/or an acute disease combined with a chronic condition [27].

We examine alternative variance structures for the $G$ and $R$ matrices where the covariances are not set to zero like Eq. 7, and compare the results in terms of DIC values to see which structure is better. We also configure different prior specifications for the random effect parameters and assess the impacts on the DIC measure and MCMC diagnostic checks. Particularly, instead of 1/3 for the expected variances at the limit (see above), we evaluate two other values: (1) one, and (2) estimated value from the MLE approach. Finally, we perform comparisons between our proposal and the situations when one employs a series of univariate (multilevel) models for predicting the outcomes.

### 5 Results

Before presenting descriptives of the independent variables we perform some data preprocessing activities to prevent unexpected errors when we fit the models. These include: 1) imputing missing values in such features as VISN and CAN score by unconditional mode imputation, 2) removing outliers from such variables as age and assigned provider experience thus focusing on the first through ninety ninth percentiles, and 3) binning multimodal highly skewed features such as Distance and Length of stay into discrete factors. Following these preprocessing steps, the number of records was reduced to 81,190 patients. The results are summarized in Table 2. Note that SD stands for standard deviation and % denotes the percentages of the subgroup in the population. As appeared, the mean age of patients is 62.42 years (SD=15.26) and about half of the cohort were over 63. Not surprisingly, near 94% of our veteran population was male and approximately 61% of all were insured. Over half of the patients were married but...
lower than one third of all were reported as actively employed. The most frequently enrolled patients are the low income and Medicaid group followed by >50 % connected disability, and non-service connected patients with income above HUD (Housing and Urban Development). The majority of patients (93 %) did not spend a day as an inpatient admitted to the hospital, and most of them travelled only a short distance to receive care from the VA hospitals. The mean care assessment need score is roughly 47 with a great variation (SD = 28.88). Also, on average, most of patient’s assigned providers are well experienced working rather full time in their roles.

Next, we provide two schematic views of 1) the mean annual care demands and 2) disease prevalence, across multiple patient groups. In Fig. 2, the average relative value unit demands for the primary care (PCRVU) and non-primary care (NonPCRVU) are displayed across different priority groups with insurance status nested. Not unexpectedly, the non-primary care workload is always more than the primary care workload and its ratio changes from 1.8 in group 8-insured to 6.6 in group 4-uninsured. In all priority groups, uninsured VA patients compared to insured ones produce, on average, more workload both in primary care and non-primary care. In addition, the biggest (lowest) workload demands for both primary and specialty care services are associated with group 8-uninsured patients (group 6-insured patients).

Figure 3 displays a mosaic plot of illness types along with patient gender and marital status. We exclude ACC 28 (neonate’s diseases) and ‘unknown’ marital category because of absence or rarity in our sample study. Note that letters P, N, and M above the marital status bar denote ‘Previously married’, ‘Never married’, and ‘Married’ groups. The ACC labels are given in Table 6. As shown, the most commonly occurring conditions among all patient clusters is ACC 30 (Screening) followed by ACCs 5 (nutritional and metabolic) and 16 (heart). However, the least prevalent illnesses among the VA patients are ACC24 (pregnancy-related), ACC13 (developmental disability), and ACC15 (cardio-respiratory arrest). Plus, in almost all disease types, married males are more at risk than the two other male groups.

In order to specify the density functions of the response variables, we use QQ plots and the ML method introduced in Section 3.2. As a result, the lognormal distribution is found the most proper case for both responses. Figure 4 shows the QQ plots along with bootstrapped point-wise confidence envelopes at 0.95 accuracy rate. As shown, the primary care workload (left panel) displays a perfect linear pattern, and for the non-primary care workload (right panel), almost all points lie within the confidence band. We also get the minimum value of the minus log-likelihood based on the ML fitting when the lognormal distribution is taken. Also based on the ways described in Section 3.2 for determining the link function, we observe that the default identity link for the lognormal distribution does estimate the upper and lower tails of both responses more properly compared to the other links, and thus it is chosen for our study.

We set the significance level at 0.05, MCMC iterations at 50,000, burn-in period at 5000 iterations, and thinning interval at 25. We then fit the six models in Table 1 starting from Model 1 and try to answer the following three questions:

- How much of the variance in primary care and non-primary workload is associated with patients, PCMH teams, and VA facilities?
- Does the effect of any patient-level predictor change among PCMH teams or VA facilities? And does the effect of any team-level predictor vary among VA facilities?
- What is the impact of patient non-adherence (as measured by Changed provider count) on primary care workload, controlling for patient, PCMH team, and VA facility factors?

Table 1: Regression modeling strategy and specific results for the 3-level hierarchical model

| Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---------|---------|---------|---------|---------|---------|
| No predictors, just residual and random intercepts (Unconditional) | Model 1 + patient-level predictors | Model 2 + random slopes for patient-level predictors | Model 3 + team-level predictors | Model 4 + random slopes for team-level predictors | Model 5 + facility-level predictors |
| Results used to compute Interclass Correlation Coefficient (ICC) which assesses the degree of clustering among subsets of cases in the data. | Results show the relationships between patient-level predictors and outcomes. | Model 2 results + findings that show if the associations between patient-level predictors and the outcomes vary across team-level and facility-level units. | Model 3 results + results that reveal the relationships between team-level predictors and the outcomes. | Model 4 results + findings that show if the associations between team-level predictors and the outcomes vary across facility-level units. | Model 5 results + results that indicate the relationships between team-level predictors and the outcomes. |
estimated (co)variances for the $R$ matrix similar to the structure defined in Eq. 7. The first (third) row in each table shows the random intercept variances for the primary care (non-primary care) and the second row corresponds to the covariances between these two responses. The interclass correlation coefficient (or ICC) at the team level (level-2) for the primary care outcome is computed as the proportion of the variance in the team level (first cell in Table 4) to total variance for the primary care workload:

$$\frac{0.168}{0.609 + 0.168 + 0.218} = 0.218$$

That is, about 17% of all variation in primary care workload is due to different PCMH teams. Following the same way, we find that near 22% is there between VA facilities, leaving near 61% of the variance to be accounted for by the patients. Thus a practically meaningful proportion of all variation happens at higher levels, providing support for our use of a 3-level hierarchical model. These percentages are 5, 16, and 79% for the non-primary care workload respectively.

Other useful points can be made from Model 1 by interpreting the correlations between primary care and non-primary care at different levels. First, the results of a joint conditional independence test of Gueorguieva [28] show that the two response variables are positively associated at the...
patient level, which confirms the fact that a simultaneous modeling of both primary and non-primary care is more reasonable than using one of them in isolation. Second, we infer that the correlation is not significant when it comes to the PCMH team level (level-2), and it is poorly significant at the facility level (level-3).

We continue our modeling to include the predictors and random components at all levels, and then answer the two other questions based on the outputs from the best model. For brevity, we will not walk through all details in each model fitted, and instead summarize the results in Table 6. In this table, level-2 and level-3 predictors are underlined. In each row, the first number is related to the primary care and the second is pertained to the non-primary care outcome. We use (single quotation mark) and (double quotation mark) to display the significance level at 0.05, <0.001. It is worth noting that we suppress the overall intercept in fitting the models, since otherwise, the parameter estimates associated with the primary care are translated as contrasts with the non-primary care response. Also for the team-level, facility-level, and interactions, we only include those factors that are significant in at least one of the six models.

According to the DIC values shown at the bottom of Table 6, we figure out that each forward model exhibits a better fit to the data, so we take Model 6 as the best model and use it to answer the remaining questions. We repeat the joint independence test of Gueorguieva [28] for model 6 and reaffirm the positive correlation of responses at the patient level. In other words, after controlling for all sources of variation, if the primary care workload is increased from one patient to another, on average we will expect an increase in the related non-primary care. In Table 6, the estimates for deterministic effects are interpreted as prevalence ratios but variance components are reported in natural scale.
Further, we perform some diagnostic tests for verifying Model 6. First, to assess the Markov chain convergence and mixing properties, trace plots and smoothed posterior densities are provided for each parameter of interest. As an illustration, Fig. 5 shows the plots for age and gender across both outcomes and Fig. 6 displays them for R-side covariance components. As depicted in Fig. 5, the traces are trendless and the chains are mixing well travelling quickly to the target distribution with small autocorrelations. Nearly the same patterns are observed in Fig. 6, but chains are now mixing marginally at a bit slower traverse rate, which can easily be tackled by increasing the MCMC iterations. Nonetheless, the densities do smoothly estimate the mean posterior for the residual variances as reported in Table 6. For deterministic terms in Fig. 5 the posterior histogram is plotted in log scale. We additionally perform Gelman, Rubin [29] and effective sample size tests to all posterior estimates and we find no violations therein.

It is worth to highlight that some estimates are changed in terms of significance among models. For example, age, insurance, and CAN score are significant in Model 2 but no longer significant in later models once their related random slopes are accounted for in Model 3. Examining other random components in these models, we figure out that significant variability exists in their nested random intercepts and slopes, even after controlling for these patient-level predictors. Hence, we can say that the association between these variables and the outcomes varies considerably among the PCMH teams. We expect that the influence of patient oldness on care demands may be stronger or weaker from one PCMH team to another within a VA facility. The same thing happens in terms of effect magnitude for ‘changed provider count’ between Model 4 and Model 5; the relationship between this variable and both workloads changes meaningfully among different VA facilities. By these statements, we tackle our second research question.

To answer the last research question, we look at the deterministic effect of ‘changed provider count’ in Model 6. As shown, for each time that a patient switches assigned provider, we will expect an average of 6% more workload in his/her primary care, after accounting for variations of his/her non-primary care demands. Other selected key findings from Model 6 can be summarized as below:

### Table 3
Estimated random intercept (co)variances introduced by the VA facilities (level-3) with 95% lower and upper highest posterior density

| Components                  | Posterior Mean | L-95 % HPD | U-95 % HPD |
|-----------------------------|----------------|------------|------------|
| PCRVU, PCRVU                | 0.218          | 0.215      | 0.220      |
| PCRVU, Non-PCRVU            | −0.006         | −0.009     | −0.002     |
| Non-PCRVU, Non-PCRVU        | 0.157          | 0.153      | 0.162      |

### Table 4
Estimated random intercept (co)variances introduced by the PCMH teams (level-2) with 95% lower and upper highest posterior density

| Components                  | Posterior Mean | L-95 % HPD | U-95 % HPD |
|-----------------------------|----------------|------------|------------|
| PCRVU, PCRVU                | 0.168          | 0.158      | 0.176      |
| PCRVU, Non-PCRVU            | 0.035          | −0.014     | 0.059      |
| Non-PCRVU, Non-PCRVU        | 0.053          | 0.049      | 0.057      |

### Table 5
Estimated patient-level residual (co)variances (level-1) with 95% lower and upper highest posterior density

| Components                  | Posterior Mean | L-95 % HPD | U-95 % HPD |
|-----------------------------|----------------|------------|------------|
| PCRVU, PCRVU                | 0.609          | 0.604      | 0.614      |
| PCRVU, Non-PCRVU            | 0.316          | 0.298      | 0.330      |
| Non-PCRVU, Non-PCRVU        | 0.787          | 0.781      | 0.792      |
### Table 6 - Coefficient estimates from the 3-level hierarchical model for joint primary care and non-primary care workloads

| Deterministic Effect | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|----------------------|---------|---------|---------|---------|---------|---------|
| Changed provider count | 1.11 | 0.41 | 0.20 | 0.01 | 0.02 | 0.02 |
| Gender, Male | 1.02 | 1.04 | 1.03 | 1.03 | 1.01 | 1.02 | 1.04 |
| Age | 0.92 | 0.94 | 0.9 | 0.9 | 0.92 | 0.94 | 0.92 | 0.93 |
| Insurance, Yes | 0.95 | 0.92 | 0.94 | 0.95 | 0.91 | 0.93 | 0.9 | 0.93 | 0.92 |
| LOS, Zero | 0.07 | 0.74 | 0.06 | 0.73 | 1.07 | 0.72 | 1.07 | 0.72 |
| CAN Score | 1.12 | 1.07 | 1.08 | 1.02 | 1.09 | 1.03 | 1.08 | 1.03 | 1.1 | 1.02 |
| SQRT (CAN Score) | 1.15 | 1.19 | 1.11 | 1.12 | 1.12 | 1.13 | 1.11 | 1.12 | 1.12 | 1.13 |
| Priority (rcf=8) | 0.96 | 1.25 | 0.97 | 1.22 | 0.96 | 1.23 | 0.95 | 1.24 | 0.95 | 1.23 |
| 1 (disability>50 %) | 1.02 | 1.32 | 1.02 | 1.28 | 1.03 | 1.29 | 1.01 | 1.3 | 1.01 | 1.28 |
| 2 (disability 30–40 %) | 0.94 | 1.01 | 0.92 | 1.04 | 0.92 | 1.03 | 0.93 | 1.04 | 0.94 | 1.04 |
| 3 (disability 20–30 %) | 1.03 | 1.17 | 1.04 | 1.14 | 1.03 | 1.15 | 1.05 | 1.15 | 1.04 | 1.15 |
| 4 (catastrophically dis.) | 1.05 | 1.03 | 1.04 | 1.05 | 1.05 | 1.04 | 1.05 | 1.03 | 1.05 | 1.04 |
| 5 (Medicaid) | 1.06 | 1.34 | 1.03 | 1.32 | 1.03 | 1.33 | 1.04 | 1.35 | 1.05 | 1.35 |
| 6 (Agent Orange, …) | 0.09 | 1.1 | 0.08 | 1.07 | 0.09 | 1.07 | 0.08 | 1.09 | 0.08 | 1.1 |
| 7 (below HUD) | 0.08 | 1.02 | 0.1 | 0.04 | 1.04 | 1.22 | 1.05 | 1.22 | 1.05 | 1.22 |
| ACC001–Infected and Parasitic | 1.04 | 1.33 | 1.04 | 1.3 | 1.03 | 1.31 | 1.04 | 1.32 | 1.03 | 1.31 |
| ACC002–Malignant Neoplasm | 1.07 | 1.65 | 1.06 | 1.65 | 1.06 | 1.64 | 1.07 | 1.64 | 1.06 | 1.64 |
| ACC003–Benign/In Situ/Uncertain Neoplasm | 1.53 | 0.98 | 1.52 | 0.97 | 1.53 | 0.96 | 1.53 | 0.97 | 1.52 | 0.98 |
| ACC004–Diabetes | 1.18 | 1.02 | 1.19 | 1.03 | 1.2 | 1.02 | 1.2 | 1.04 | 1.19 | 1.03 |
| ACC005–Nutritional and Metabolic | 1.13 | 1.04 | 1.11 | 1.05 | 1.12 | 1.05 | 1.11 | 1.05 | 1.11 | 1.04 |
| ACC006–Liver | 1.09 | 1.13 | 1.07 | 1.14 | 1.07 | 1.14 | 1.08 | 1.15 | 1.08 | 1.14 |
| ACC007–Gastrointestinal | 1.18 | 1.27 | 1.17 | 1.27 | 1.16 | 1.28 | 1.16 | 1.27 | 1.17 | 1.26 |
| ACC008–Musculoskeletal and Connective Tissue | 1.09 | 1.05 | 1.08 | 1.06 | 1.07 | 1.06 | 1.08 | 1.04 | 1.08 | 1.05 |
| ACC009–Hematological | 1.12 | 0.98 | 1.11 | 1.11 | 0.99 | 1.12 | 1.11 | 1.12 |
| ACC010–Cognitive Disorders | 1.06 | 0.88 | 1.05 | 0.9 | 1.05 | 0.89 | 1.06 | 0.89 |
| ACC011–Substance Abuse | 1.03 | 1.73 | 1.04 | 1.7 | 1.03 | 1.71 | 1.04 | 1.71 | 1.04 | 1.73 |
| ACC012–Mental | 0.99 | 1.24 | 1.01 | 1.23 | 1.01 | 1.22 | 1.01 | 1.24 |
| ACC013–Developmental Disability | 1.07 | 1.15 | 1.06 | 1.14 | 1.07 | 1.16 | 1.07 | 1.15 | 1.07 | 1.14 |
| ACC014–Neurological | 1.07 | 1.02 | 1.03 | 1.04 | 1.05 | 1.03 | 1.05 | 1.03 | 1.06 | 1.03 |
| ACC015 – Cardio-Respiratory Arrest | 1.15 | 1.05 | 1.14 | 1.06 | 1.16 | 1.04 | 1.15 | 1.05 | 1.15 | 1.05 |
| ACC016 – Heart | 1.05 | 1.02 | 1.05 | 1.03 | 1.04 | 1.01 | 1.03 | 0.99 | 1.04 | 1.01 |
| ACC017 – Cerebrovascular | 1.08 | 1.26 | 1.1 | 1.26 | 1.09 | 1.27 | 1.11 | 1.27 | 1.09 | 1.26 |
| ACC018 – Vascular | 1.09 | 1.11 | 1.07 | 1.12 | 1.08 | 1.12 | 1.08 | 1.11 | 1.08 | 1.11 |
| ACC019 – Lung | 0.98 | 1.14 | 1.12 | 1.38 | 1.1 | 1.39 | 1.12 | 1.39 | 1.11 | 1.39 |
| ACC020 – Eyes | 0.94 | 0.98 | 0.92 | 0.94 | 0.92 | 0.94 | 0.93 | 1 | 0.93 | 1.01 |
| ACC021 – Ears, Nose, and Throat | 1.12 | 1.25 | 1.15 | 1.41 | 1.16 | 1.42 | 1.17 | 1.44 | 1.16 | 1.43 |
| ACC022 – Urinary System | 0.97 | 0.1 | 0.94 | 0.98 | 0.92 | 0.94 | 0.93 | 1 | 0.93 | 1.01 |
| ACC023 – Genital System | 1.12 | 1.42 | 1.13 | 1.43 | 1.12 | 1.43 | 1.13 | 1.41 | 1.12 | 1.41 |
| ACC025 – Skin and Subcutaneous | 1.12 | 1.28 | 1.11 | 1.29 | 1.12 | 1.3 | 1.12 | 1.29 | 1.12 | 1.29 |
| ACC026 – Injury, Poisoning, Complications | 1.17 | 1.45 | 1.15 | 1.41 | 1.16 | 1.42 | 1.17 | 1.44 | 1.16 | 1.43 |
| ACC027 – Symptoms, Signs, and Ill-Defined Conditions | 0.9 | 1.01 | 0.94 | 0.98 | 0.92 | 0.94 | 0.93 | 1 | 0.93 | 1.01 |
| ACC028 – Transplants, Openings, Amputations | 1.22 | 2.01 | 1.23 | 1.98 | 1.2 | 1.98 | 1.2 | 2 | 1.22 | 1.99 |
| Changed provider count | 1.11 | 1.09 | 1.04 | 1.03 | 1.03 | 1.02 |
Table 6 (continued)

| Distance (ref = Far) | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---------------------|---------|---------|---------|---------|---------|---------|
| Middle              |         |         |         |         |         |         |
| Diabetes × Liver    | 1.02, 1.13' | 1.03, 1.15' | 1.03, 1.16' | 1.01, 1.16' | 1.01, 1.14' |
| Near                |         |         |         |         |         |         |
| Diabetes × Cardio-Respiratory Arrest | 1.12', 1.11'' | 1.11', 1.13' | 1.13', 1.12'' | 1.15', 1.14' | 1.13', 1.14'' |
| Diabetes × Heart    | 1.03, 1.1'' | 1.04, 1.12'' | 1.03, 1.11'' | 1.01, 1.11'' | 1.01, 1.11'' |
| Diabetes × Cerebrovascular | 1.07', 1.17' | 1.06', 1.14' | 1.06', 1.15' | 1.06', 1.16' | 1.06', 1.15' |
| Diabetes × Urinary System | 1.04, 1.12' | 1.06, 1.11' | 1.05, 1.1' | 1.07, 1.13' | 1.06, 1.13' |

Variance Component

| Residual            | 0.609', 0.79' | 0.446', 0.55' | 0.357', 0.46' | 0.352', 0.44' | 0.259', 0.43' | 0.255', 0.43' |
| Intercept (team)    | 0.168', 0.05' | 0.093', 0.04' | 0.076', 0.04' | 0.064', 0.04' | 0.059', 0.03' | 0.054', 0.03' |
| Intercept (facility)| 0.218', 0.16' | 0.125', 0.1' | 0.106', 0.08' | 0.091', 0.08' | 0.085', 0.07' | 0.083', 0.05' |
| Slope (age: team)   | 0.088', 0.09' | 0.081', 0.09' | 0.075', 0.1' | 0.073', 0.08' |         |         |
| Slope (age^2: team) | 0.042', 0.06' | 0.047', 0.07' | 0.053, 0.06' | 0.056', 0.06' |         |         |
| Slope (CAN Score: team) | 0.078', 0.09' | 0.072', 0.1' | 0.069', 0.09' | 0.066', 0.09' |         |         |
| Slope (CAN Score(0.5): team) | 0.037', 0.05' | 0.042', 0.04' | 0.038', 0.04 | 0.041, 0.05' |         |         |
| Slope (insurance: facility) | 0.051', 0.06' | 0.047', 0.07' | 0.049', 0.06' | 0.045', 0.07' |         |         |
| Slope (changed provider count: facility) | 0.053', 0.06' | 0.046', 0.05' |         |         |         |         |
| Model Fit           |         |         |         |         |         |         |
| DIC                 | 461019.6 | 227245.2 | 225469.7 | 225411.4 | 225351.3 | 225337.8 |

- Adjusting for the contributions of all other variables, female VA patients tend to produce about 57 % more primary care (98 % more non-primary care) workload compared to males. This is not unexpected due to gender imbalance issue existed in VA patients.
- Inpatient cohort generally creates 28 % more workload in non-primary care compared to outpatients, after accounting for variations of their primary care relative value units.
- Catastrophically ill veterans (P4) have 1.15 times the non-primary care demands of the P8 comparison group. The increase rates are about 35 and 23 % for veterans exposed to Agent Orange (or other herbicides) and >50 % for disabled veterans. Having been exposed to such chemicals also notably affects the increased care for cardio-respiratory arrest.
- Change rates in primary cares vary from 7 % decrease for ACC29 (Transplant) to 52 % increase for ACC4 (Diabetes). For non-primary cares, this varies from 11 % reduction for ACC11 (Substance Abuse) to 99 % rise for ACC30 (Screening).
- Both the team-level (patient non-adherence) and facility-level (distance) predictors are significantly associated with the outcomes; patients travelling more miles to VA hospitals are likely to generate a larger amount of care than closely located patients.
- In co-occurring diseases studies, diabetes greatly interacts with some acute and chronic conditions. For instance, in
patients with cardio-respiratory arrest, having diabetes is associated with a 13% (14%) increase in primary care (non-primary care) workload. Another comorbid condition that poses a similar pattern is heart disease, especially for cerebrovascular patients.

- Risk adjustment for disease types and their interactions improves the model fit to a great extent (about 160 K reduction in DIC) and makes most of their related effects statistically significant.

To test the effect of baseline structure for the $G$ and $R$ (co)variance matrix, we prepare two scenarios: (1) set structure $\begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \otimes I$ for the workload covariances at all three hierarchical levels, and (2) pick structure (7) for the workload covariances. This results in 8 different combinations as illustrated in Table 7. We run Model 6 for each case twice to take control of the Monte Carlo error and keep all other factors constant among different fittings. As shown, the best fit is corresponding to the
first row in which the variance structure (7) is applied at all levels of the hierarchy. In order to evaluate the choice of priors for the random effects, we follow the analyses described in Section 4 and find that almost no change occurs in deterministic estimates, DIC measure, and directions of (co)variance components in Model 6. However, the absolute range of alternations in variance estimates is around 2.3 % of those in Model 6. We detect that better chain convergence and mixing property is observed when using priors with smaller limit (co)variances. Further, the posterior correlation estimates remain reasonably unchanged while examining different types of priors, which provide some reassurance that our priors do not dominate the model to an unacceptable extent.

Finally to compare between the univariate and bivariate approach for prediction, we keep Model 6 settings constant and consider two scenarios: (1) A bivariate 3-level model with joint primary care and non-primary care workloads, and (2) Two univariate 3-level models one for the primary care and one for non-primary care workload predictions. Fitting both models, we aggregate the credible intervals for the mean outcomes and then compare them with the actual values. Interestingly, the probability of joint correct prediction (for both responses) is about 67 % for the first scenario and about 58 % for the second. Then we pick those correct intervals, compute their ranges, and calculate basic statistics for the resulting ranges in Table 8. As displayed, the credible intervals are substantially narrowed when applying the multivariate approach. Thus we can conclude that a joint modeling of primary care and non-primary care workloads would provide more robust and realistic predictions for medical home practices.

### Table 7 DIC values for models fitted with different covariance structures

| Facility | Team | Patient | Deviance information criterion |
|----------|------|---------|--------------------------------|
| 2        | 2    | 2       | 225337.8–227448.1              |
| 2        | 2    | 1       | 225491.7–225494.1              |
| 2        | 1    | 2       | 225401.1–225396.9              |
| 2        | 1    | 1       | 225582.5–225580.3              |
| 1        | 2    | 2       | 225378.5–225375.7              |
| 1        | 2    | 1       | 225444.9–225441.2              |
| 1        | 1    | 2       | 225457.8–225460.5              |
| 1        | 1    | 1       | 225550.7–225554.0              |

### Table 8 Summary statistics for the range of joint correct intervals

|             | Multivariate | Univariate |
|-------------|--------------|------------|
| Primary care| 0.431        | 0.514      |
| Non-primary care | 1.023 | 1.083      |
| Mean        | 0.381        | 0.439      |
| Median      | 0.977        | 1.058      |

### 6 Conclusion

A key factor in the success of medical homes in delivering quality and coordinated care lies in their teams’ ability to handle uncertainties that can be caused by different sources such as patient/physician appointment scheduling, care logistics, and more importantly patients’ health demands. This paper addresses the problem of clinical demand prediction in the presence of nested sources of variation at different operational levels. We collect outpatient visit data from a large sample of Veterans Affairs hospitals and investigate the relationship between risk factors at three operational levels and total care demands on a yearly basis. We propose a multivariate multilevel generalized linear model in a Bayesian framework to predict the care demand portfolio in medical home practices. The proposal can fit heteroscedastic variances and unstructured covariance matrices for nested random effects and residuals as well as their interactions with categorical and continuous covariates simultaneously.

We find that utilizing a multilevel analysis with nested random components can greatly contribute to model fit in hierarchical healthcare systems. Further, we show that risk-adjustment for patient disease conditions and their comorbidities extensively enhance the prediction power of our model. Our results confirm that using a multivariate as opposed to a univariate approach can significantly shrink the correct credible intervals for workload predictions thus allowing for a more precise estimation of either outcome. The approach used in this paper has a general application and could also be employed for analysis of multiple health outcomes in a variety of health analytics contexts.

Turning to specific results from recent VA data, we see that overall, the primary care is positively associated with the non-primary care (correlation coefficient of 0.027 as appeared in Fig. 6 middle panel) after accounting for all studied sources of variability. We find the association between patient-level predictors such as age and the care workloads varies considerably among PCMH teams within a hospital. Further, the effect of patient non-adherence on care demands is subject to change from one hospital to another. Moreover, it is found that patient oldness can contribute to the increased care demands required for heart, nutritional, and gastrointestinal diseases.

Finally there are some limitations to this research that need to be mentioned. First, the data in our study are collected solely from a veteran population (with fewer female and more senior patients) who receives support from government budgets. Thus the results from our study may not fully generalize to other health care systems. Second the data used is administrative and not real time, so some issues such as model tuning and calibration should be taken into account when dealing with online prediction efforts.

Our work can further be extended in some fronts. One challenging direction would be to modify the proposed
approach to handle longitudinal observations from past history of care demands for a specific patient profile. This may be done by expanding the multivariate distribution of outcomes to include a temporal dimension which requires great care in model specification and implementations thanks to various inter-correlations. Alternatively, one can combine some autoregressive terms to the variance structure introduced in this work. Another issue worth exploring is related to the way that one can adjust for patient risk or comorbidities. Although several algorithms such as Clinical Risk Group (CRG), veriskhealth DxCG [30], and CMS’s HCC software have been used in the literature, no scientific study is available to systematically evaluate the impacts of each algorithm on prediction modeling of care demands.

References

1. Green LV, Savin S, Murray M (2007) Providing timely access to care: what is the right patient panel size? Jtcomm j on quality and patient saf 33(4):211–218
2. Murray M, Davies M, Boushon B (2007) Panel size: answers to physicians’ frequently asked questions. Fam Pract Manag 14(10):29–32
3. Ostbye T, Yarnall KS, Krause KM, Pollak KL, Gradison M, Michener JL (2005) Is there time for management of patients with chronic diseases in primary care? Ann Fam Med 3(3):209–214
4. Naessens JM, Stroebel RJ, Finnie DM, Shah ND, Wagle AE, Litchy WJ, Killinger PJ, O’Byrne TJ, Wood DL, Nesse RE (2011) Effect of multiple chronic conditions among working-age adults. The Am j of manag care 17(2):118–122
5. Potts B, Adams R, Spadin M (2011) Sustaining primary care practice: a model to calculate disease burden and adjust panel size. The Permanente j 15(1):53–56
6. Balasubramanian H, Banerjee R, Denton B, Naessens J, Stahl J (2010) Improving clinical access and continuity through physician panel redesign. J Gen Intern Med 25(10):1109–1115
7. Rittenhouse DR, Shortell SM, Fisher ES (2009) Primary care and accountable care–two essential elements of delivery-system reform. N Engl J Med 361(24):2301–2303
8. Patient-Centered Primary Care Collaborative (PCPCC). (2009). Available at http://www.pcpcc.org/about/medical-home. Accessed August 14, 2013
9. Backer LA (2009) Building the case for the patient-centered medical home. Fam Pract Manag 16(1):14–18
10. Fisher ES (2008) Building a medical neighborhood for the medical home. N Engl J Med 359(12):1202–1205
11. Beal AC, Fund C (2007) Closing the divide: how medical homes promotes equity in health care. Commonwealth Fund
12. Rittenhouse DR, Shortell SM (2009) The patient-centered medical home: will it stand the test of health reform? J Am Med Assoc 301(19):2038–2040
13. Reid RJ, Coleman K, Johnson EA, Fishman PA, Hsu C, Soman MP, Tresco CE, Erikson M, Larson EB (2010) The group health medical home at year two: cost savings, higher patient satisfaction, and less burnout for providers. Health Aff 29(5):835–843
14. Bates DW, Bitton A (2010) The future of health information technology in the patient-centered medical home. Health Aff 29(4):614–621
15. Bitton A, Martin C, Landon BE (2010) A nationwide survey of patient centered medical home demonstration projects. J Gen Intern Med 25(6):584–592
16. Klein S (2011) The Veterans Health Administration: implementing patient-centered medical homes in the nation’s largest integrated delivery system. Commonwealth Fund
17. Liu CF, Sales AE, Sharp ND, Fishman P, Sloan KL, Todd-Stenberg J, Nichol WP, Rosen AK, Loveland S (2003) Case-mix adjusting performance measures in a veteran population: pharmacy- and diagnosis-based approaches. Health Serv Res 38(5):1319–1337
18. Moerbeek M (2004) The consequence of ignoring a level of nesting in multilevel analysis. Multivar Behav Res 39(1):129–149
19. Goldstein H (2011) Multilevel statistical models. Wiley series in probability and statistics, 4 edn. John Wiley & Sons
20. Shams I, Ajorlou S, Yang K (2014) A multivariate hierarchical Bayesian framework for healthcare predictions with application to medical home study in the Department of Veteran Affairs. arXiv preprint arXiv:14030674
21. Srivastava MS, von Rosen T, von Rosen D (2008) Models with a Kronecker product covariance structure: estimation and testing. Math methods of statistics 17(4):357–370
22. Hosmer Jr DW, Lemeshow S, Sturdivant RX (2013) Applied logistic regression. John Wiley & Sons
23. Damlen P, Wakefield J, Walker S (1999) Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. J R Stat Soc Ser B (Stat Methodol) 61(2):331–344
24. Garcia-Cortés LA, Sorensen D (2001) Alternative implementations of Monte Carlo EM algorithms for likelihood inferences. Genet Sel Evol 33(4):443–452
25. Hadfield JD (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J Stat Softw 32(2):1–22
26. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. J R Stat Soc Ser B (Stat Methodol) 64(4):853–639
27. Petersen LA, Byrne MM, Daw CN, Hasche J, Reis B, Pietz K (2010) Relationship between clinical conditions and use of veterans affairs health care among medicare-enrolled veterans. Health Serv Res 45(3):762–791
28. Gueorguieva R (2001) A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. Stat Model 1(3):177–193
29. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Statistical science:457–472
30. VERISK Heath Inc (2011) Verisk Health DxCG medical classification system – Version 7 structural summary Available at http://www.veriskhealth.com/verisk-advantage/DxCG-Medical-Classification-V7-Structural-Summary.pdf. Accessed March 18, 2014