

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/269985440>

Bayesian Component Selection in Multi-response Hierarchical Structured Additive Models with an Application to Clinical Workload Prediction in Patient Centered Medical Homes

ARTICLE *in* IIE TRANSACTIONS · DECEMBER 2014

Impact Factor: 1.37 · DOI: 10.1080/0740817X.2014.982840

READS

35

3 AUTHORS, INCLUDING:



Kai Yang

Wayne State University

108 PUBLICATIONS 1,148 CITATIONS

SEE PROFILE

This article was downloaded by: [Kai Yang]

On: 01 January 2015, At: 11:10

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



[Click for updates](#)

IIE Transactions

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/uiie20>

Bayesian Component Selection in Multi-response Hierarchical Structured Additive Models with an Application to Clinical Workload Prediction in Patient Centered Medical Homes

Issac Shams PhD^a, Saeede Ajorlou PhD^a & Kai Yang PhD^b

^a Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, , ,

^b Department of Industrial and Systems Engineering, Healthcare Systems Engineering Group, Wayne State University, Detroit, MI 48202, , hse.eng.wayne.edu

Accepted author version posted online: 23 Dec 2014.

To cite this article: Issac Shams PhD, Saeede Ajorlou PhD & Kai Yang PhD (2014): Bayesian Component Selection in Multi-response Hierarchical Structured Additive Models with an Application to Clinical Workload Prediction in Patient Centered Medical Homes, IIE Transactions, DOI: [10.1080/0740817X.2014.982840](https://doi.org/10.1080/0740817X.2014.982840)

To link to this article: <http://dx.doi.org/10.1080/0740817X.2014.982840>

Disclaimer: This is a version of an unedited manuscript that has been accepted for publication. As a service to authors and researchers we are providing this version of the accepted manuscript (AM). Copyediting, typesetting, and review of the resulting proof will be undertaken on this manuscript before final publication of the Version of Record (VoR). During production and pre-press, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal relate to this version also.

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Bayesian Component Selection in Multi-response Hierarchical Structured Additive Models with an Application to Clinical Workload Prediction in Patient Centered Medical Homes

Issac Shams, PhD

Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, MI 48109
issacsh@umich.edu

Saeede Ajorlou, PhD

Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, MI 48109
saeede@umich.edu

Kai Yang, PhD

Department of Industrial and Systems Engineering
Healthcare Systems Engineering Group
Wayne State University
Detroit, MI 48202
ac4505@wayne.edu
<http://hse.eng.wayne.edu>

1. Introduction

Recent years have seen enormous revolutions in US health care systems due to advancements in database systems that rapidly collect and organize massive electronic health records. Standard operations management based medical practices have moved from relatively ad hoc and subjective to data driven decision making and evidence-based health care. Traditional “fee-for-service” payment methods have transitioned into models that tie reimbursement to outcomes such as pay-for-performance and bundled payment. The complexity and abundance of health care data have grown thanks to the development of new data gathering techniques as capturing devices, sensors, and mobile applications. Subsequently, there have been more and more incentives for hospitals to reduce costs and promote quality by using advanced data analytics tools that help find insights from large, noisy, heterogeneous, longitudinal, and hierarchical health care data.

Many kinds of health care data, including clinical data, billing/claims data, and patient specific data, involve hierarchical (nested) or clustered structures. For example, in a study of assessing differences in mortality rates across hospitals, data is randomly collected on samples of patients nested within each hospital. In this application, there are two levels of the hierarchy (level-1 for patients and level-2 for hospitals), and for each level, a set of specific covariates exists (such as age, gender, and severity of illness at the first level; and hospital size and hospital teaching status at the second level) that might have a relationship with the outcome. To handle these hierarchically structured data, multilevel models (also known as hierarchical linear models, variance components models, random-effect models, or split-plot designs) have been proposed and applied in different fields including psychometrics, biostatistics and econometrics (Goldstein, 2011). The basic idea is to link the covariates at higher levels to the predictor variables at lower

levels by imposing another set of regressions in which the lower-level (regression) coefficients are explained by higher-level predictors.

The assumption of parametric form of covariates in the hierarchical linear model makes it rather restricted. For example, in longitudinal growth studies where repeated measures of the response variable (e.g., height) are clustered within individuals, the relation between age and height is often found to be exponential. To relax the linearity constraints, covariates with nonparametric structure (such as local regression or smoothing spline) or semiparametric structure (such as partially linear models or varying-coefficient models) can be incorporated in the multilevel framework at each level of the hierarchy (Goldstein, 2011). One such extension is generalized additive mixed models, which enjoy the nonparametric properties of additive models and distributional flexibility of generalized linear mixed models. Another more recent class of this type is the hierarchical version of structured additive regression (STAR) models (Lang et al., 2013) that offers a broad and rich class of complex regression containing several important subclasses as special cases, such as generalized additive mixed models, state-space models for longitudinal studies, geoadditive models (Kamann & Wand, 2003), and varying-coefficient models (Hastie & Tibshirani, 1993).

As in many areas of statistical modeling and machine learning, the problem of variable selection (also known as feature selection, attribute selection, model selection, and variable subset selection) has become an important issue in multilevel models. Variable selection often aims to choose a subset of relevant covariates from a possibly large set of candidates that might include many redundant or irrelevant features. Due to its practical importance, this problem has attracted many researchers from diverse fields, leading to a vast amount of literature on selecting predictors of regression models. Classical methods in this area basically relied on 1) p -value

such as stepwise deletion or 2) information criteria like AIC, BIC, and more recently focused information criterion (Claeskens & Hjort, 2003), among others. However such approaches usually suffer from lack of stability and perform poorly in selecting random effect components (Breiman, 1996). In addition, they involve a combinatorial optimization comparing 2^{p+q} different models (p and q are numbers of fixed and covariance parameters, respectively) which is NP -hard and might be infeasible to solve even when $p + q \ll \text{sample size}$ is fixed (Pu & Niu, 2006). To address such drawbacks, regularization (or shrinkage) methods have been introduced that focus on selecting variables simultaneously with model estimation using some data oriented penalty functions. Popular examples may include the least absolute shrinkage and selection operator (Lasso; Tibshirani, 1996) or smoothly clipped absolute deviation (SCAD; Fan & Li, 2001) and modifications such as hierarchical or random Lasso. To get an overview of variable selection in linear models, see the review paper by Chen et al. (2013). Variable selection is also of great importance in high-dimensional data such as DNA microarray or functional MRI data (see Fan and Lv (2010) for a review). Likewise, various studies have been devoted to variable selection in nonparametric additive models and semiparametric linear models (see, for example, Huang et al. (2010) and Kundu and Dunson (2013)). Multivariate variable selection has also been investigated in a number of studies such as (Brown et al., 1998) and (Cai et al., 2005).

Compared to classical methods that are primarily based on Bayes factors, approaches for Bayesian variable selection are mostly built around spike-and-slab priors. The basic idea is to introduce a binary latent variable I_j associated with each regression coefficient so that the variable is forced to be zero when I_j is in the spike part, or kept unchanged if I_j is in the slab part. The posterior distribution of I_j is then interpreted as marginal posterior probabilities for inclusion or exclusion of the respective covariate. See stochastic search variable selection (SSVS) of

George and McCulloch (1993) and mixture of Zellner's g priors of Liang et al. (2008) as popular examples, and a recent review paper by O'Hara and Sillanpää (2009).

In multilevel models, however, the problem of selecting the random effects is more complicated since it involves boundary problems that can arise from either nonnegative constraints on fixed-effect parameters or positive semidefinite constraints on covariance matrices. To date, approaches for variable selection in this class mainly pertain to linear (or generalized linear) mixed models such as the generalized information criterion of Pu and Niu (2006), and Bayesian methods of Spiegelhalter et al. (2002), among others (see (Müller et al., 2013) for a review).

In contrast to variable selection, component (or function) selection deals with selecting an appropriate subset of covariates and, at the same time, determining whether linear or more flexible functional forms of covariates have to be chosen. Research in this area was started by Antoniadis and Fan (2001) who proposed a group SCAD penalty for regularization in wavelet approximation. Lin and Zhang (2006) developed the COSSO estimator in additive smoothing spline analysis of variance (SS-ANOVA) models with a fixed number of covariates. Recently, by extending the nonnegative garrote estimator of [Breiman \(1995\)](#), [Marra and Wood \(2011\)](#) developed a single step shrinkage approach for function selection in generalized additive models.

In this paper, consistent with the idea of modeling multivariate outcomes in multilevel data structures ([Goldstein, 2011, Chap. 6](#)), we first extend hierarchical STAR models introduced in (Lang et al., 2013) to include multivariate response variables from the exponential family distributions. This way, we will be able to simultaneously model the relationship of several responses on a set of structured additive predictors accounting for possible correlation among the dependent variables. Then, we propose spike-and-slab priors for automatic variable selection and

model choice within a Bayesian hierarchical framework similar to the work of Scheipl et al. (2012). We apply our model to real-world healthcare data obtained from the Department of Veteran Affairs (VA). The application analyzes Patient Centered Medical Home (PCMH) project data gathered from a large number of medical facilities during fiscal year 2011–12. Separate data tables from 1) patient health conditions and care utilization, and 2) patient demographic information are first combined to form patient-level data. The patient-level data is further aggregated to the provider and station levels to help predict patient's total care demands on primary and non-primary care on a yearly basis. By combining these multilevel data sources together, our proposal can assist health professionals in making operational decisions such as determining the number of primary care physicians based on expected clinic visits or expected clinical workloads for those visits.

The main contributions we make in this paper include formulating a multivariate version of the hierarchical STAR model, bridging the connection between multivariate hierarchical STAR models and generalized latent variable models, proposing a Bayesian function selection routine for the multivariate hierarchical STAR model based on spike and slab priors, and applying our proposal to real-world data from the VHA medical home project to demonstrate its performance and applicability and produce findings that convey key public and medical implications.

The rest of the paper is organized as follows. Section 2 reviews some literature in patient centered medical home and outlines the problem statement. Section 3 introduces some background on structured additive regression based on Bayesian P-spline and its hierarchical version. Section 4 describes the proposed multivariate extension to hierarchical STAR models followed by a Bayesian variable selection procedure. Section 5 provides an illustrative

application of our methods in VHA patient centered medical home data. Section 6 includes concluding remarks and directions for future research.

2. Overview of patient centered medical home

The patient centered medical home has been emerged as a new model for delivery system reform that has potentials to improve primary care quality with better outcomes and at lower costs. The model is a patient-oriented team-based approach consisting of different health professionals such as physician, registered nurse, nutritionist, and clerk that delivers accessible, coordinated, and comprehensive care in the context of patient's family and community (Stange et al., 2010). The medical home concept originated during the 1960's in pediatrics carrier but did not find its way to adult general practices until 2004. Theoretically, the PCMH model entails a broad set of fundamental principles such as having a physician directed medical practice taking responsibilities for all of the continuing care, an enhanced access to care through open scheduling systems with expanded hours and personalized communications, and an appropriate payment system that recognizes the added value provided to patients beyond the traditional fee-for-service encounters (Rittenhouse & Shortell, 2009).

As of 2007, there were some literature examining the prevalence and effectiveness of medical homes. For instance, [Fisher \(2008\)](#) outlined some recommendations for the success of medical homes such as sharing of information across health care providers, extending the performance measures to include patients' experience with care and assessments of outcomes, and establishing a PCMH based payment system that share savings among all providers involved. 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-one month period (Reid et al., 2010). Practically, as of

December 2009, there were about 26 pilot projects involving medical home being directed in 18 states. These includes over 14,000 physicians and approximately 5 million patients (Bitton et al., 2010). Of interest, the 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 one day and decreasing the percentage of inappropriate emergency department visits from 52% to 12% (Klein & Fund, 2011).

However, there are difficulties in fully achieving the benefits of the PCMH in practice. It is found that many efforts are required in the PCMH model to fully leverage the electronic health record technology and to develop new business rules and staffing structures than initially envisioned (Ajlrou et al., 2014; Rittenhouse & Shortell, 2009). From an operations management point of view, a key success factor in designing any healthcare delivery system is to achieve 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 in nature. Unlike health demands, the supply of healthcare services can be treated as deterministic and be calculated based on head counts and available service hours from all professional lines within a PCMH team on an annual basis. Yet, the estimation of clinical workload portfolio based on key patient factors is a challenging task, and to the best of our knowledge, our work is the first attempt tackling this problem within the OR/MS and IE community.

3. Background

3.1. STAR models based on Bayesian P-splines

Let (y_i, x_i, v_i) , $i = 1, \dots, n$, denote the i -th sampled vector of data, where y_i is the response variable, $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$ is a vector of continuous covariates, and $v_i = (v_{i1}, v_{i2}, \dots, v_{is})'$

is a vector of further (mostly categorical) predictors. Structured Additive Regression (STAR) models (Fahrmeir et al., 2004) assume that, given x_i and v_i , the distribution of y_i belongs to an exponential family $\pi(y_i|x_i, v_i, \phi) = c(y_i, \phi) \exp\left(\frac{y_i \theta_i - b(\theta_i)}{\phi}\right)$, where $b(\cdot)$, $c(\cdot)$, θ_i , and ϕ are determined by the type of distribution. The conditional expected value $\mu_i = E(y_i|x_i, v_i)$ is related to a semiparametric additive predictor η_i by $\mu_i = g(\eta_i)$ via a fixed (known) link function $g(\cdot)$ as in generalized linear models. The additive predictor η_i has the form

$$\eta_i = f_1(x_{i1}) + \dots + f_p(x_{ip}) + v_i' \gamma, \quad (1)$$

in which $f_1, \dots, f_p$ are unknown nonlinear (possibly smooth) functions of the continuous covariates, and $v_i' \gamma$ represents the usual linear part of the model. Following the Bayesian version of P(enalized)-splines (Lang & Brezger, 2004), the unknown function f_j is approximated by a polynomial spline of degree r defined over a set of (not necessarily equally spaced) knots $x_j^{min} = \zeta_j^0 < \zeta_j^1 < \dots < \zeta_j^{k_j-1} < \zeta_j^{k_j} = x_j^{max}$ within the domain of x_j . The spline can be expressed in terms of a linear combination of $M_j = k_j + r$ B-spline basis functions evaluated at the observation x_j , that is,

$$f_j(x_j) = \sum_{m=1}^{M_j} \beta_{jm} B_{jm}(x_j). \quad (2)$$

Here $B(\cdot)$'s are known basis functions and $\beta_j = (\beta_{j1}, \dots, \beta_{jM_j})'$ corresponds to a vector of unknown regression coefficients to be estimated. By defining a $(n \times M_j)$ design matrix $X_j(i, m) = B_{jm}(x_{ij})$, the predictor (1) can be rewritten in matrix form as

$$\eta = X_1 \beta_1 + \dots + X_p \beta_p + V \gamma, \quad (3)$$

where V is the usual design matrix for linear effects. Within this unified framework, components of (2) can represent various types of model terms, such as 1) linear terms ($f_j(x) = \beta_j x_j$); 2) nominal or ordinal predictors ($f(x_{ji}) = \beta_{x(k)}$ iff $x_{ji} = k$); 3) smooth functions of continuous

covariates (splines, kriging effects, tensor product splines, etc.); 4) Markov random field or its conditional specification, e.g. the conditional autoregressive model; 5) random effect models (cluster-specific intercept or slopes); and 6) interaction terms between different effects (varying-coefficient models, effect modifiers).

For a fully Bayesian inference when selection of variables (and functions) is not considered, a diffuse prior $p(\gamma) \propto \text{const}$ is typically used for sampling from γ . The choice of priors for the unknown functions $f_1, \dots, f_p$, however, depends on the type of the covariate and the prior beliefs about smoothness. To avoid overfitting of a particular function f_j , the smoothness priors can be written into a general form as

$$p(\beta_j | \tau_j^2) \propto \left(\frac{1}{\tau_j^2}\right)^{\text{rk}(\mathbf{K}_j) / 2} \exp\left(-\frac{1}{2\tau_j^2} \beta_j' \mathbf{K}_j \beta_j\right), \quad (4)$$

in which $\mathbf{K}_j$ is a penalty matrix and τ_j^2 is the variance parameter. The goal of $\mathbf{K}_j$ is to shrink smoothness parameters towards zero, or penalize unexpected jumps between adjacent β_j 's. In most cases such as in the Gaussian random field, $\mathbf{K}_j$ is rank deficient (i.e., $\text{rk}(\mathbf{K}_j) < M_j$), leading to a partially improper prior for β_j . The variance parameter τ_j^2 controls the amount of smoothness and is sampled by an uninformative (conjugate) inverse Gamma hyperpriors $\tau_j^2 \sim IG(a_j, b_j)$ normally with small choices for a_j and b_j .

3.2. Hierarchical STAR models

When data are hierarchically structured in some levels, STAR models can be extended in a multilevel framework to account for possible correlations within units of a cluster (or a level) in the hierarchy. Such specification is usually expressed by imposing another regression model with structured additive predictors to the coefficient β_j in (3) as

$$\beta_j = \eta_j + \varepsilon_j = \mathbf{X}_{j1}\beta_{j1} + \dots + \mathbf{X}_{jp_j}\beta_{jp_j} + \mathbf{V}_j\gamma_j + \varepsilon_j. \quad (5)$$

Here it is assumed that $\boldsymbol{\varepsilon}_j \sim N(\mathbf{0}, \tau_j^2 I)$ is a vector of i.i.d. Gaussian random variables, but more complicated forms such as Dirichlet process mixture can be applied. Modeling higher levels of the hierarchy are also straightforward by again setting another STAR equation to the parameters in (5): for example, $\boldsymbol{\beta}_{jl} = \boldsymbol{\eta}_{jl} + \boldsymbol{\varepsilon}_{jl}, l = 1, \dots, p_j, \boldsymbol{\varepsilon}_{jl} \sim N(\mathbf{0}, \tau_{jl}^2 I)$ for level-3 regression. In this way, the whole model can be seen as a hierarchy of complicated STAR models with (possibly) nonlinear and smooth terms. In some applications of hierarchical models, observations are clustered according to their spatial (or geographical) positions. For example, in our VA medical home study, x_j may represent the district (or zip code) in which the patient lives. This way, x_j represents a group indicator taking values of $c \in \{1, \dots, C_j\}$. Then a regular way to model such cluster specific heterogeneity is to assume $f_j(c) = \beta_{jc} \sim N(0, \tau_j^2)$ with design matrix X_j being a 0/1 incidence matrix of dimension $n \times C_j$. Note that this approach is also taken when modeling random intercepts in multilevel structure. In other applications, we may study how the effect of a covariate is modified according to changes in the levels of a third variable. Such interactions can happen among the covariates at one given level or across multiple levels. As an instance in our case study, we are interested in how possessing a particular comorbid condition can moderate the relationship between patient age and healthcare demand. Here, it is presumed that x_j is a two-dimensional term as $x_j = (x_j^{(1)}, x_j^{(2)})'$. If $x_j^{(1)}$ is continuous and $x_j^{(2)}$ is categorical, their interaction is modeled by $f_j(x_j) = h(x_j^{(1)})x_j^{(2)}$, and the associated design matrix is given by $\text{diag}(x_1^{(2)}, \dots, x_n^{(2)})X_j^{(1)}$, in which $X_j^{(1)}$ is the usual design matrix for spline basis function evaluated at the observation $x_j^{(1)}$. If both covariates are continuous, a more flexible approach can be based on two dimensional P-spline, in which the unknown interaction surface can be approximated by the tensor product of the corresponding one-dimensional B-splines as

$f_j(x_j^{(1)}, x_j^{(2)}) = \sum_{m_1=1}^{M_{1j}} \sum_{m_2=1}^{M_{2j}} \beta_{j,m_1 m_2} B_{j,m_1}(x_j^{(1)}) B_{j,m_2}(x_j^{(2)})$. The related design matrix X_j is then $n \times (M_{1j} \cdot M_{2j})$ and it consists of products of basis functions. The appropriate priors for $\beta_j = (\beta_{j,11}, \dots, \beta_{j,M_{1j}M_{2j}})'$ are commonly found in spatial statistics.

Another common application in multilevel analysis is related to random slopes that appear when combining regression equations of higher levels with the lower levels to form a compound representation (Goldstein, 2011, Chap. 2). For example, in our case study of the VA medical home project, we would like to model the heterogeneity in the slope of relationship between healthcare demand and patient's age among all PCMH teams. Then, a random slope with regard to index variable $x_j^{(1)}$, which indicates the teams here, can be incorporated as $f_j(x_j) = h(x_j^{(1)})x_j^{(2)}$ with $h(x_j^{(1)}) = \beta_{jc} \sim N(0, \tau_j^2)$. Following this, the design matrix X_j is given by $diag(x_1^{(2)}, \dots, x_n^{(2)})X_j^{(1)}$ where $X_j^{(1)}$ is a 0/1 incidence matrix.

4. Proposed methods

4.1. Multi-response hierarchical STAR model

When we want to simultaneously study multiple response variables, a multivariate model should be developed to capture additional correlations among different measurements. One key advantage of such modeling lies in its ability to control the type I error rate better as compared to carrying out a series of univariate tests. In the context of multilevel analysis, different responses can be incorporated by placing them in a separate “response” level at the lowest level of the hierarchy. A series of d dummy variables, one for each response, is then defined and entered into regression equations of higher levels. For simplicity, we first focus on the three-level structure, $response^{(h)}$ within $patient_i$ within (medical home) $team_j$, with regular predictors, and then

show how this can be extended to the STAR context. A model with more than three levels is just a straightforward extension of what we propose here.

Suppose there are H response variables in the lowest level. We define $d_{hij}^{(h')} = 1$ if response h' -th is modeled and zero otherwise (Goldstein, 2011, Chap. 6). Let $x_{p,ij}$ and $z_{q,j}$ denote p -th and q -th covariate in the patient level and team level, respectively. Let $u_{0,j}^{(h')}$ and $u_{p,j}^{(h')}$ represent h' -th random intercept and h' -th random slope of the p -th predictor in the patient level, one-to-one. Then we model the outcome as

$$y_{hij} =$$

$$\sum_{h'} d_{hij}^{(h')} \beta_0^{(h')} + \sum_{h'} d_{hij}^{(h')} \sum_{p=1}^P \bar{\beta}_p^{(h')} x_{p,ij} + \sum_{h'} d_{hij}^{(h')} \sum_{q=1}^Q \bar{\beta}_q^{(h')} z_{q,j} + \sum_{h'} d_{hij}^{(h')} \sum_{p=1}^P \sum_{q=1}^Q \beta_{p,q}^{(h')} x_{p,ij} z_{q,j} + \sum_{h'} d_{hij}^{(h')} \sum_{p=1}^P u_{p,j}^{(h')} x_{p,ij} + \sum_{h'} d_{hij}^{(h')} u_{0,j}^{(h')} + \sum_{h'} d_{hij}^{(h')} \varepsilon_{ij}^{(h')} \quad (6)$$

$$\begin{bmatrix} u_{0,j}^{(1)} \\ \vdots \\ u_{p,j}^{(1)} \\ \vdots \\ \vdots \\ \vdots \\ u_{p,j}^{(H)} \end{bmatrix} \sim N(\mathbf{0}, \Omega_u), \quad \Omega_u = \begin{bmatrix} \tau_{u_0}^{2(1)} & \dots & \dots & \dots & \dots & \tau_{u_{0,p}}^{(1)(H)} \\ \vdots & \ddots & & & & \vdots \\ \tau_{u_{0,p}}^{(1)} & \dots & \tau_{u_p}^{2(1)} & & & \tau_{u_{p,p}}^{(1)(H)} \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ \vdots & & & & & \ddots \\ \tau_{u_{0,p}}^{(1)(H)} & \dots & \dots & \dots & \dots & \tau_{u_p}^{2(H)} \end{bmatrix} \quad (7)$$

$$\begin{bmatrix} \varepsilon_{ij}^{(1)} \\ \vdots \\ \varepsilon_{ij}^{(H)} \end{bmatrix} \sim N(\mathbf{0}, \Omega_\varepsilon), \quad \Omega_\varepsilon = \begin{bmatrix} \sigma_\varepsilon^{2(1)} & \dots & \sigma_\varepsilon^{(1)(H)} \\ \vdots & \ddots & \vdots \\ \sigma_\varepsilon^{(1)(H)} & \dots & \sigma_\varepsilon^{2(H)} \end{bmatrix} \quad (8)$$

The first term in (6) shows the grand mean for each of the response variables followed by patient level predictors and team level predictors, then cross-level interactions (effect modifiers) are included followed by random slopes and then random intercept terms, and finally patient level residuals are included. Note that there is no level-1 residual specified since level-1 exists only to define the multivariate structure. The random effects are defined in (7) with a general unstructured covariance Ω_u that contains the pairwise covariances between each set of these random effects for the intercept and slopes within each of the responses and between the response variables. The patient level residuals are defined in (8) with covariance structure Ω_ε that would include all variances and covariances between patient level residuals. Taking a matrix form, we can rewrite (6) as

$$y_{hij} = \sum_{h'} d_{hij}^{(h')} \mathbf{Z}_j^T \mathfrak{B}^{(h')} \mathbf{X}_{ij} + \sum_{h'} d_{hij}^{(h')} \mathbf{U}_j^{(h')T} \mathbf{X}_{ij} + \sum_{h'} d_{hij}^{(h')} \varepsilon_{ij}^{(h')}, \quad (9)$$

where we have

$$\mathbf{Z}_j = [1, z_{1,j}, \dots, z_{Q,j}]^T, \mathbf{X}_{ij} = [1, x_{1,ij}, \dots, x_{P,ij}]^T, \mathbf{U}_j^{(h')} = [u_{0,j}^{(h')}, u_{1,j}^{(h')}, \dots, u_{P,j}^{(h')}]^T \quad (10)$$

$$\mathfrak{B}^{(h')} = \begin{bmatrix} \underline{\beta}_0^{(h')} & \overline{\beta}_1^{(h')} & \dots & \overline{\beta}_P^{(h')} \\ \underline{\beta}_1^{(h')} & \beta_{1,1}^{(h')} & \dots & \beta_{P,1}^{(h')} \\ \vdots & \vdots & & \vdots \\ \underline{\beta}_Q^{(h')} & \beta_{1,Q}^{(h')} & \dots & \beta_{P,Q}^{(h')} \end{bmatrix}. \quad (11)$$

Note that $\overline{\beta}_1^{(h')}, \dots, \overline{\beta}_P^{(h')}$ in the first row of (11) show regression coefficients for patient level predictors while $\underline{\beta}_1^{(h')}, \dots, \underline{\beta}_Q^{(h')}$ placed in the first column of (11) indicate coefficients for team level variables.

To extend this within the STAR framework where the covariates are represented by a linear combination of B-splines basis functions, we simplify (6) for a particular outcome h' as

$$y_{hij}^{(h')} = \beta_0^{(h')} + \sum_{p=1}^P \left(\bar{\beta}_p^{(h')} + u_{p,j}^{(h')} \right) x_{p,ij} + \sum_{q=1}^Q \underline{\beta}_q^{(h')} z_{q,j} + \sum_{p=1}^P \sum_{q=1}^Q \beta_{p,q}^{(h')} x_{p,ij} z_{q,j} + u_{0,j}^{(h')} + \varepsilon_{ij}^{(h')} \text{ for } h' = 1, \dots, H. \quad (12)$$

We assume that, for response h' , patient level covariate $x_p, p = 1, \dots, P$ is represented by a set of

$$\bar{M}_p^{(h')} = \bar{k}_p + \bar{r} \text{ polynomial spline of degree } \bar{r} \text{ over } \bar{k}_p + 1 \text{ knots } \zeta_p^{0(h')} < \zeta_p^{1(h')} < \dots < \zeta_p^{k_j(h')}.$$

Similarly, team level predictor $z_q, q = 1, \dots, Q$ is represented by $\underline{M}_q^{(h')} = \underline{k}_q + \underline{r}$ polynomial splines of degree $\underline{r}$ over a domain defined by $\zeta_q^{0(h')} < \zeta_q^{1(h')} < \dots < \zeta_q^{k_j(h')}$. Hence,

a hierarchical STAR model with a multivariate response has the form

$$y_{hij}^{(h')} = \beta_0^{(h')} + \sum_{p=1}^P \sum_{m_1=1}^{\bar{M}_p^{(h')}} \left(\bar{\beta}_{m_1 p}^{(h')} + u_{m_1, p, j}^{(h')} \right) \bar{B}_{m_1 p}^{(h')} (x_{p,ij}) + \sum_{q=1}^Q \sum_{m_2=1}^{\underline{M}_q^{(h')}} \underline{\beta}_{m_2 q}^{(h')} \underline{B}_{m_2 q}^{(h')} (z_{q,j}) + \sum_{p=1}^P \sum_{q=1}^Q \sum_{m_1=1}^{\bar{M}_p^{(h')}} \sum_{m_2=1}^{\underline{M}_q^{(h')}} \beta_{m_1 m_2 p q}^{(h')} \bar{B}_{m_1 p}^{(h')} (x_{p,ij}) \underline{B}_{m_2 q}^{(h')} (z_{q,j}) + u_{0,j}^{(h')} + \varepsilon_{ij}^{(h')}, \quad h' = 1, \dots, H \quad (13)$$

$$\begin{bmatrix} u_{0,j}^{(1)} \\ \vdots \\ \vdots \\ u_{\bar{M}_P^{(H)}, P, j}^{(H)} \end{bmatrix} \sim N(\mathbf{0}, \Omega_u), \quad \Omega_u = \begin{bmatrix} \tau_{u_0}^{2(1)} & \dots & \dots & \dots & \tau_{u_{0, (\bar{M}_P^{(H)}, P)}}^{(1)(H)} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \tau_{u_{0, (\bar{M}_P^{(H)}, P)}}^{(1)(H)} & \dots & \dots & \dots & \tau_{u_{\bar{M}_P^{(H)}, P}}^{2(H)} \end{bmatrix} \quad (14)$$

$$\begin{bmatrix} \varepsilon_{ij}^{(1)} \\ \vdots \\ \varepsilon_{ij}^{(H)} \end{bmatrix} \sim N(\mathbf{0}, \Omega_\varepsilon), \quad \Omega_\varepsilon = \begin{bmatrix} \sigma_\varepsilon^{2(1)} & \dots & \sigma_\varepsilon^{(1)(H)} \\ \vdots & \ddots & \vdots \\ \sigma_\varepsilon^{(1)(H)} & \dots & \sigma_\varepsilon^{2(H)} \end{bmatrix} \quad (15)$$

In (13), $B(\cdot)$ and $\beta(\cdot)$ represent basis functions and B-spline coefficients, respectively. Random effect splines are defined in (14). For a particular outcome, the patient level random effects present each patient's deviance from the average intercept $u_{0,j}$ and from the average slope of each the splines $(u_{1,j}, \dots, u_{m_1,p,j})$. The patient level covariance matrix includes the pairwise covariances between each set of spline random effects for the intercept and slopes within each of the response variables as well as between the response variables. The patient level residuals are defined in (15) with covariance structure Ω_ϵ . Although covariances described in (14) and (15) are in general unstructured format, special forms such as Toeplitz or Kronecker type structure can be taken based on different applications.

Following section 3.2, the interaction effect between patient level and team level covariates is modeled with varying coefficient $h(x_{p,ij})z_{q,j}$ if z is categorical, or through nonparametric two dimensional surface fitting of $f(x_p, z_q)$ by the tensor product of two univariate B-splines as in (13), if z is continuous. If variable selection is not looked at, the most commonly used priors for the latter case are established with the next four nearest neighbors on a regular lattice as

$$\beta_{m_1 m_2 p q}^{(h')} \mid \cdot \sim N \left(\frac{1}{4} \left(\beta_{(m_1-1)m_2 p q}^{(h')} + \beta_{(m_1+1)m_2 p q}^{(h')} + \beta_{m_1(m_2-1)p q}^{(h')} + \beta_{m_1(m_2+1)p q}^{(h')} \right), \frac{\tau_{pq}^{2(h')}}{4} \right), \quad (16)$$

for $m_1 = 2, \dots, \overline{M}_p^{(h')} - 1, m_2 = 2, \dots, \underline{M}_q^{(h')} - 1$. This can be seen as a direct generalization of a first-order random walk in one dimension. Other types of priors such as Kronecker product of penalty matrices of the main effects $\mathbf{K}_{pq,j} = \mathbf{K}_{p,j} \otimes \mathbf{K}_{q,j}$ can also be applied (see Lang & Brezger, 2004).

4.2. Relationship with a structural equation model

Here we show how the multilevel spline model with a multivariate response can equivalently be represented and estimated in the structural equation modeling framework. For simplicity we choose a model with only level-2 predictors, but this can be extended to more general cases with higher level predictors and possible interactions such as the one we developed in Section 4.1. In addition, we pick the linear spline model as a special case to help better understand the approach, but this can easily be generalized to other types of splines like the one we exploit in this paper.

Generally structural equation models (SEM) involve two specific parts with distinct objectives: a measurement equation and a structural equation (Kline, 2011). In the measurement equation, each of the responses $y_j^{(h')}$ loads on the latent variables $f_m^{(h')}$, $m = 0, 1, \dots, M^{(h')}$. The intercept term for response h' is $f_0^{(h')}$ and the loadings for any of the measurements $y_j^{(h')}$ on this latent variable are 1. The other $M^{(h')}$ factors serve as the slopes for each piece on domain x_p defined by the linear splines

$$s_{m,pj}^{(h')} = \begin{cases} 0 & \text{if } s_{pj} \leq s_{(m-1),p}^{(h')} \\ s_{pj} - s_{(m-1),p}^{(h')} & \text{if } s_{(m-1),p}^{(h')} < s_{pj} \leq s_{m,p}^{(h')} \\ s_{m,p}^{(h')} - s_{(m-1),p}^{(h')} & \text{if } s_{pj} > s_{m,p}^{(h')} \end{cases} \quad (17)$$

Applying the same $M^{(h')} + 1$ pieces, $m = 0, 1, \dots, M^{(h')}$, as above, the measurement equation can be written as

$$y_j^{(h')} = f_0^{(h')} + \sum_{m: s_{m,p}^{(h')} \leq s_{pj}} (s_{m,p}^{(h')} - s_{(m-1),p}^{(h')}) f_m^{(h')} + \sum_{m: s_{(m-1),p}^{(h')} < s_{pj} < s_{m,p}^{(h')}} (s_{pj} - s_{(m-1),p}^{(h')}) f_m^{(h')} + \sum_{m: s_{(m-1),p}^{(h')} \geq s_{pj}} 0 f_m^{(h')} + \varepsilon_j^{(h')}, \text{ for } h' = 1, \dots, H; \text{ for } p = 1, \dots, P. \quad (18)$$

Any rescaling of (18) proportional to the loadings can be employed as well. To see how this is equivalent to the multilevel spline model, an additional subscript showing patients, i , is included and $(\beta_{m,p}^{(h')} + u_{m,pj}^{(h')})$ is substituted for each of the $f_m^{(h')}$. This gives

$$\begin{aligned}
 y_{ij}^{(h')} &= (\beta_0^{(h')} + u_{m,pj}^{(h')}) + \sum_{m: s_{m,p}^{(h')} \leq s_{pj}} (s_{m,p}^{(h')} - s_{(m-1),p}^{(h')}) (\beta_{m,p}^{(h')} + u_{m,pj}^{(h')}) + \sum_{m: s_{(m-1),p}^{(h')} < s_{pj} < s_{m,p}^{(h')}} (s_{pj} - s_{(m-1),p}^{(h')}) (\beta_{m,p}^{(h')} + u_{m,pj}^{(h')}) + \sum_{m: s_{(m-1),p}^{(h')} \geq s_{pj}} 0 (\beta_{m,p}^{(h')} + u_{m,pj}^{(h')}) + \varepsilon_{ij}^{(h')} \\
 &= (\beta_0^{(h')} + u_{m,pj}^{(h')}) + \sum_{m=1}^M s_{m,pj}^{(h')} (\beta_{m,p}^{(h')} + u_{m,pj}^{(h')}) + \varepsilon_{ij}^{(h')}, \text{ for } h' = 1, \dots, H; \text{ for } p = 1, \dots, P,
 \end{aligned} \tag{19}$$

which can be derived from (13) with spline $B(\cdot)$ defined in (17) and excluding terms that contain level-3 covariates z_q 's.

The structural equation part of the SEM characterizes the mutual relationships between the factors. It can be shown that the coefficient for the univariate relationship between any two factors, $f_{m_1}^{(h_1)}$ regressed on $f_{m_2}^{(h_2)}$ is identical to that between two random effects $u_{m_1,p}^{(h_1)}$ and $u_{m_2,p}^{(h_2)}$ by substituting $(\beta_{m,p}^{(h')} + u_{m,pj}^{(h')})$ for each of the $f_m^{(h')}$ as

$$\beta_{m_1 m_2}^{(h_1)(h_2)} = \frac{\text{cov}(f_{m_1}^{(h_1)}, f_{m_2}^{(h_2)})}{\text{var}(f_{m_2}^{(h_2)})} = \frac{\text{cov}(\beta_{m_1,p}^{(h_1)} + u_{m_1,p}^{(h_1)}, \beta_{m_2,p}^{(h_2)} + u_{m_2,p}^{(h_2)})}{\text{var}(\beta_{m_2,p}^{(h_2)} + u_{m_2,p}^{(h_2)})} = \frac{\text{cov}(u_{m_1,p}^{(h_1)}, u_{m_2,p}^{(h_2)})}{\text{var}(u_{m_2,p}^{(h_2)})} = \frac{\tau_{m_1 m_2, p}^{(h_1)(h_2)}}{\tau_{m_2, p}^{2(h_2)}}, \tag{20}$$

in which the numerator and denominator can be found from (14). Similarly, other regression coefficients derived from the relationship between factors can be demonstrated to be equal to those between random effects.

A number of works have investigated the equivalence of linear multilevel models and SEMs in the literature (Rabe-Hesketh et al., 2004; Steele, 2008). Yet, it should be pointed out that

nonlinear multilevel models and generalized linear multilevel models do not always have identical parameterization within the SEM framework. Our goal here is to provide a basis for replacing a multivariate linear multilevel spline model with a standard SEM so that specific strengths of SEM analysis can be captured and they might help improve upon our multilevel analysis. Examples of such strengths may include the ability to explicitly model measurement errors through multiple indicator latent factor, and testing within-level and across-level mediation which are not straightforward in multilevel analysis. Also our attempts here can further be utilized in a way to parameterize and estimate generalized STAR models within a standard SEM framework.

4.3. Bayesian function selection

In real world data sets with complex hierarchical structures, choosing a suitable subset among many potential predictors and at the same time determining their appropriate shapes (smooth vs. linear) and interaction effects is a challenging and important task. For example, in our case of the VA medical home study, we want to select a small group from a set of 30 comorbidity indicator variables and to determine whether the effect of patient age and patient care assessment need score on the response variables are nonlinear or linear, whether an interaction between age and some of the comorbidities exists, and whether a district-specific heterogeneity arising from the location of medical facilities is necessary. To this end, we apply spike-and-slab prior structure for selecting single effect variables as well as grouped coefficients combined with smoothing parameters that represent particular model terms. The main idea of such an approach is to assume a mixture prior for each $\beta_j = (\beta_{j1}, \dots, \beta_{jM_j})'$ with one part being a narrow spike around the origin that imposes very strong shrinkage on the coefficients and the other part being a wide slab that forces very little shrinkage on the coefficients (Ishwaran & Rao, 2005). The posterior

mixture weights for the spike (or slab) component of a specific coefficient or coefficient batch can be interpreted as the posterior probability of its exclusion from (or inclusion in) the model.

According to Section 4.1, we note that any multi-response hierarchical STAR model of form (13) can be written in a unifying form $\mathbf{y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\eta} = \boldsymbol{\eta}_0 + \mathbf{X}_1\boldsymbol{\beta}_1 + \cdots + \mathbf{X}_p\boldsymbol{\beta}_p$ with $\boldsymbol{\eta}_0$ showing offset terms (e.g., grand means of multivariate responses) and effects that are not under selection procedure. Then the conventional spike-and-slab prior structure is given by the following hierarchical Bayesian model

$$\begin{aligned}\beta_j | \delta_j, \rho_j^2 &\stackrel{\text{prior}}{\sim} N(0, v_j^2) \text{ with } v_j^2 = \rho_j^2 \delta_j, \\ \delta_j | \omega &\stackrel{\text{prior}}{\sim} \omega I_1(\delta_j) + (1 - \omega) I_{v_0}(\delta_j), \\ \rho_j^2 &\stackrel{\text{prior}}{\sim} \Gamma^{-1}(a_\rho, b_\rho), \\ \text{and } \omega &\stackrel{\text{prior}}{\sim} \text{Beta}(a_\omega, b_\omega).\end{aligned}\tag{21}$$

This structure is called Normal-mixture of inverse Gammas (NMIG) prior that places a bimodal prior on the hyper-variance v_j^2 of the coefficients that leads to a spike-and-slab type prior on the STAR coefficient themselves. $I_z(\cdot)$ is an indicator function that takes 1 in z and zero otherwise and v_0 is a very small positive constant. This way, δ will be 1 with probability ω and close to zero with probability $(1 - \omega)$. Hence, the implied prior for (hyper-)variance v_j^2 is a bimodal mixture of inverse Gamma distributions, with one part focused on very small values—the spike with $\delta_j = v_0$ —and a second diffuse part with more mass on larger values—the slab with $\delta_j = 1$. The mixture weights ω , in addition, follow a Beta prior that captures any prior knowledge about the sparsity of coefficient β_j (Scheipl et al., 2012).

It is found that prior structure (21) does not work well for coefficient batches in the STAR models that are associated with spline basis functions or random effects. Briefly, the problem is that a small hyper-variance for a batch of coefficients entails small coefficient values and vice versa. This problematic dependence between a vector of coefficients and their associated hyper-variances makes the MCMC sampler unlikely to switch between basins of attraction around the two spike and slab modes. To reduce the dependence, a multiplicative parameter expansion for β_j is recommended that improves the mixing properties of δ_j and boosts the shrinkage characteristics of the resulting prior compared to (21). The idea is to expand β_j as $\beta_j = \alpha_j \Xi_j$ where scalar $\alpha_j \sim \text{NMIG}(v_0, \omega, a_\rho, b_\rho)$ is given as (21), and it is independent of Ξ_j . Elements of the M_j -dimensional vector Ξ_j are then assigned as

$$\Xi_{jm}|r_{jm} \sim N(r_{jm}, 1), \quad r_{jm} \sim \frac{1}{2}I_1(r_{jm}) + \frac{1}{2}I_{-1}(r_{jm}), \quad j = 1, \dots, p; m = 1, \dots, M_j \quad (22)$$

which corresponds to a mixture of two i.i.d Gaussian density with mean ± 1 and equal mixture weights. The current approach resolves the mixing problems of δ_j since the Markov blankets of both δ_j and ρ_j now include only α_j with dimension one instead of the vector β_j .

The MCMC posterior inference and component selection is performed by a block-wise Metropolis-within-Gibbs sampler which reduces to a standard Gibbs scheme when responses are Gaussian (see Appendix A). The full conditional densities (FDC) for parameters ω , ρ_j^2 , δ_j , and conditional means $\mathbf{r} = (r_l', l: 1, \dots, \mathcal{L})$ of normal variables $\Xi|r_l \sim N(r_l, 1), r_l = \pm 1$ are given in closed form regardless of the choice of exponential family for the responses (Appendix A). The full conditionals of α and Ξ are based on the conditional design matrices $\mathbf{X}_\alpha = \mathbf{X}\text{blockdiag}(\Xi_1, \dots, \Xi_p)$ and $\mathbf{X}_\Xi = \mathbf{X}\text{blockdiag}(\mathbf{1}_{e1}, \dots, \mathbf{1}_{ep})\alpha$, where $\mathbf{1}_e$ is a $e \times 1$ vector

of ones and $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_p)$ is the concatenation of the designs for the model terms as in (2).

Under the Gaussian assumption of the responses, these are given as follows

$\boldsymbol{\alpha} | \cdot \sim N(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha)$ where

$$\boldsymbol{\Sigma}_\alpha = \left(\frac{1}{\phi} \mathbf{X}_\alpha^T \mathbf{X}_\alpha + \text{diag}(\delta \rho^2)^{-1} \right)^{-1}, \boldsymbol{\mu}_j = \frac{1}{\phi} \boldsymbol{\Sigma}_\alpha \mathbf{X}_\alpha^T \mathbf{y} \quad (23)$$

and

$\boldsymbol{\Xi} | \cdot \sim N(\boldsymbol{\mu}_\Xi, \boldsymbol{\Sigma}_\Xi)$ where

$$\boldsymbol{\Sigma}_\Xi = \left(\frac{1}{\phi} \mathbf{X}_\Xi^T \mathbf{X}_\Xi + \mathbf{I} \right)^{-1}, \boldsymbol{\mu}_j = \boldsymbol{\Sigma}_\Xi \left(\frac{1}{\phi} \mathbf{X}_\Xi^T \mathbf{y} + \mathbf{r} \right). \quad (24)$$

If the response variables are not Gaussian, the penalized iteratively reweighted least squares (P-IWLS) is used within a Metropolis-Hastings iteration to sample from $\boldsymbol{\alpha}$ and $\boldsymbol{\Xi}$ (Scheipl et al., 2012). The posterior inclusion probability $P(\delta_j = 1 | \mathbf{y})$ can then be employed to decide upon insignificant, intermediate, and important model terms.

5. Application to VA patient centered medical home data

5.1. Description of data set

The data we use in this study is gathered from a large number of VA medical facilities across the nation that undertook a patient centered medical home (PCMH) project as a way to reform their health care delivery system. The goal of our study here is to predict patient's annual care demands on primary care and non-primary care with the help of patient level and provider level attributes. Since patient data files are recorded separately, we first combine data tables belonging to patient health conditions (such as comorbidities) and patient care utilization (such as health care workload) with those tables associated with patient demographic and socioeconomic information in order to form a patient level data. This patient level is further aggregated to the provider level and station level data to create a three level hierarchical structure. At each level of

the hierarchy we have a set of risk factors that are selected based on relevant medical literature and confirmed by a group of VA health professionals.

We collected a random sample of 10,000 outpatients from 260 VA medical facilities through the nation during fiscal years 2011–12. All patient visits to primary care and women's health are assembled for a total capture period of one year. Visits from other primary care related clinics, such as internal medicine or geriatric primary care, are excluded from the analysis because health services requested by such visits are generally not rendered through medical homes; instead they are fulfilled by a specific physician, a licensed practical nurse, or a registered nurse.

Study variables

We identify and calculate two response variables for health care workloads generated by each unique patient during the fiscal years 2011–12. Particularly, we use Relative Value Unit (RVU) to measure the primary care and non-primary care workloads generated by a specific patient (Dummit, 2009). The RVU schema has been widely used for reimbursement and each value is assigned to a particular service (as defined by a coding system called Current Procedural Terminology or CPT) rendered by a provider. The values are adjusted by geographic regions so that, for example, a 99213 CPT code (refers to office/other outpatient services) performed in Manhattan is worth more than when performed in Dallas. Simply put, the primary care RVU represents the resources needed to provide all primary care services of a patient during a year, and non-primary care RVU refers to all of the non-primary care workload during the year, which could be from one or many visits to outpatient care units. 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 that are generated by telephone encounters.

The predictor variables are organized in three levels: level-1 is the patient level, on which patient's demographic and socioeconomic attributes are included; level-2 is the PCMH team level, on which covariates such as assigned provider's experience and frequency of times that the patient has changed his/her assigned provider are collected; and level-3 is the VA facility level, on which only one continuous covariate, zip-code based distance between patient's home and his/her assigned facility, is collected. The detailed descriptions of the variables and data types along with their summary statistics are shown in Table 1. In the table, enrollment priority is assigned based on the veteran's severity of service-connected disabilities and the VA income means test: groups 1, 2, 3 are generally veterans with service connected disabilities of > 50%, between 30% and 50%, and between 20% and 30%, respectively; 4, catastrophically disabled veterans; 5, low income or Medicaid; 6, Agent Orange or Gulf War veterans; 7, non-service connected with income being below HUD (The US Department of Housing and Urban Development); and 8, non-service connected with income being above HUD. Care assessment need score is a general illness severity measure ranging from 0 (lowest risk) to 99 (highest risk) that reflects the likelihood of hospitalization or death. Accxx indicators are aggregated condition categories determined based on the various ICD-9-CM (International Classification of Disease, ninth version, Clinical Modification) codes that are assigned to a patient at each visit during the fiscal years 2011–12. Note that acc codes are not mutually exclusive as most patients have more than one acc assigned during a year. Acc 28 is related to neonatal diseases and is absent in the studied population.

Descriptive statistics

Descriptive statistics are summarized in Table 1. In brief, vast majority of the outpatients are male and elderly living near their assigned VA medical facility. The most commonly occurring

condition is screening (about 92%) followed by nutritional diseases (about 70%) and heart diseases (about 66%). In order to find the distribution of the response variables, we build Quantile-Quantile plots of primary care and non-primary care RVUs against parametric densities with positive support such as Gaussian, lognormal, chi-squared, Gamma, and Weibull. In addition to QQ plots, we checked the approximate fit by the maximum likelihood method. Based on both criteria, the lognormal distribution is found the most proper choice for both responses. The QQ plots for primary care and non-primary care RVUs along with bootstrapped point-wise confidence envelopes at a 0.95 accuracy rate are displayed in Figure 1 and Figure 2, respectively.

Preprocessing

To preempt numerical problems in model fitting, the data is preprocessed as follows:

- a) Missing values in CAN score, provider's experience, patient marital status, and facility distance are imputed with the hot-deck method (Andridge & Little, 2010).
- b) Error values in age (e.g. greater than 130) and distance (e.g. greater than 1500 miles) are identified and removed.
- c) The scale of distance and length of stay are changed to natural logarithm since their distributions are strongly positively skewed (skewness greater than two), which can lead to volatile estimation results on a standard scale.

Following these steps, the number of records was reduced to 9,935.

5.2. Modeling

We use natural logarithm transformation for both response variables (primary care relative value unit or 'pcrvu' and non-primary care relative value unit or 'npcrvu') in order to convert them into Gaussian. We distinguish four levels of hierarchy: responses (level-1) are nested in $patient_i$

(level-2), patients are nested in PCMH $team_j$ (level-3), and PCMH teams are nested in VA medical $facility_k$ (level-4). The following four level hierarchical STAR model is suggested:

$$\text{Level} - 1: y_{ijk}^{(h')} = d_{ijk}^{(1)} \ln(\mathbf{pcrvu}) + d_{ijk}^{(2)} \ln(\mathbf{npcrvu})$$

$$\begin{aligned} \text{Level} - 2: \ln(\mathbf{pcrvu}) &= \mathbf{1}\eta_0^{(1)} + \mathbf{f}_1^{(1)}(age) + \mathbf{f}_2^{(1)}(los) + \mathbf{f}_3^{(1)}(can) + \mathbf{f}_4^{(1)}(age)acc1 + \dots + \\ &\quad \mathbf{f}_{32}^{(1)}(age)acc30 + \mathbf{f}_{33}^{(1)}(can, los) + \dots + \mathbf{V}^{(1)}\boldsymbol{\gamma}^{(1)} + \boldsymbol{\varepsilon}^{(1)} \\ &= \mathbf{1}\eta_0^{(1)} + \mathbf{X}_1^{(1)}\boldsymbol{\beta}_1^{(1)} + \dots + \mathbf{V}^{(1)}\boldsymbol{\gamma}^{(1)} + \boldsymbol{\varepsilon}^{(1)} \end{aligned}$$

$$\text{Level} - 2: \ln(\mathbf{npcrvu}) = \mathbf{1}\eta_0^{(2)} + \mathbf{X}_1^{(2)}\boldsymbol{\beta}_1^{(2)} + \dots + \mathbf{V}^{(2)}\boldsymbol{\gamma}^{(2)} + \boldsymbol{\varepsilon}^{(2)}$$

$$\begin{aligned} \text{Level} - 3: \eta_0^{(1)} &= \mathbf{1}\eta_{0,0}^{(1)} + \mathbf{f}_{0,1}^{(1)}(prov.exp) + \mathbf{f}_{0,2}^{(1)}(prov.fte) + \mathbf{f}_{0,3}^{(1)}(prov.chng) \\ &\quad + \mathbf{f}_{0,4}^{(1)}(prov.pos) + \mathbf{f}_{0,5}^{(1)}(prov.exp)prov.pos + \dots + \mathbf{V}_0^{(1)}\boldsymbol{\gamma}_0^{(1)} + \boldsymbol{\varepsilon}_0^{(1)} \\ &= \mathbf{1}\eta_{0,0}^{(1)} + \mathbf{X}_{0,1}^{(1)}\boldsymbol{\beta}_{0,1}^{(1)} + \dots + \mathbf{V}_0^{(1)}\boldsymbol{\gamma}_0^{(1)} + \boldsymbol{\varepsilon}_0^{(1)} \end{aligned}$$

$$\text{Level} - 3: \eta_0^{(2)} = \mathbf{1}\eta_{0,0}^{(2)} + \mathbf{X}_{0,1}^{(2)}\boldsymbol{\beta}_{0,1}^{(2)} + \dots + \mathbf{V}_0^{(2)}\boldsymbol{\gamma}_0^{(2)} + \boldsymbol{\varepsilon}_0^{(2)}$$

$$\text{Level} - 3: \boldsymbol{\beta}_1^{(1)} = \mathbf{f}_{3,1}^{(1)}(prov.exp) + \mathbf{V}_3^{(1)}\boldsymbol{\gamma}_3^{(1)} + \boldsymbol{\varepsilon}_3^{(1)} = \mathbf{X}_{3,1}^{(1)}\boldsymbol{\beta}_{3,1}^{(1)} + \mathbf{V}_3^{(1)}\boldsymbol{\gamma}_3^{(1)} + \boldsymbol{\varepsilon}_3^{(1)}$$

$$\text{Level} - 3: \boldsymbol{\beta}_1^{(2)} = \mathbf{f}_{3,1}^{(2)}(prov.exp) + \mathbf{V}_3^{(2)}\boldsymbol{\gamma}_3^{(2)} + \boldsymbol{\varepsilon}_3^{(2)} = \mathbf{X}_{3,1}^{(2)}\boldsymbol{\beta}_{3,1}^{(2)} + \mathbf{V}_3^{(2)}\boldsymbol{\gamma}_3^{(2)} + \boldsymbol{\varepsilon}_3^{(2)}$$

$$\text{Level} - 4: \eta_{0,0}^{(1)} = \mathbf{V}_{0,0}^{(1)}\boldsymbol{\gamma}_{0,0}^{(1)} + \boldsymbol{\varepsilon}_{0,0}^{(1)}$$

$$\text{Level} - 4: \eta_{0,0}^{(2)} = \mathbf{V}_{0,0}^{(2)}\boldsymbol{\gamma}_{0,0}^{(2)} + \boldsymbol{\varepsilon}_{0,0}^{(2)} \quad (25)$$

The top level equation contains the two responses. The level-2 equations are STAR models for logged primary and non-primary care workloads that are regressed on possibly nonlinear effects of patient age, care assessment need score, and length of stay using P-splines. We also include interaction effects between age, CAN score, priority, and all disease types, and between CAN score and length of stay with a two dimensional surface. The categorical covariates on the

patient level along with their possible interactions are encoded as dummy variables and subsumed in $\mathbf{V}^{(\cdot)}$ with parameters $\boldsymbol{\gamma}^{(\cdot)}$. Note that here we use the same set of effects for both response regressions, but this may change in other applications with a bivariate response. The first and the second level-3 equations model patient-specific offset by the team level covariates such as provider experience and its interaction with provider position plus random intercepts $\boldsymbol{\varepsilon}_0^{(\cdot)}$. In addition, the linear or index terms on this level such as provider position are included in $\mathbf{V}_0^{(\cdot)}$. The third and the fourth level-3 equations model slope-specific heterogeneity of age plus additional linear terms $\mathbf{V}_3^{(\cdot)}$, and random slopes $\boldsymbol{\varepsilon}_3^{(\cdot)}$. Finally team-specific intercepts are modeled through level-4 equations containing the logarithm of average facility distance $\mathbf{V}_{0,0}^{(\cdot)}$ and facility random intercepts $\boldsymbol{\varepsilon}_{0,0}^{(\cdot)}$.

5.3. Analyses

We perform sensitivity analysis for component selection with regards to different hyperparameter settings, i.e. $v_0 = 0.00025, 0.005, 0.01$ and $(a_\rho, b_\rho) = (5, 25), (5, 50), (10, 35)$. We also evaluate the prediction performance of models with and without higher level hierarchies based on deviance values obtained for a test subset containing 1,000 observations.

5.4. Results

The maximal model contains approximately 121 model terms with 640 coefficients in total. The hyperparameters are set to $(a_\omega, b_\omega) = (1, 1)$, $(a_\rho, b_\rho) = (5, 25)$, and $v_0 = 0.00025$. Since we convert our responses to Gaussian, a very flat hyperprior $\phi \sim \Gamma^{-1}(10^{-4}, 10^{-4})$ is chosen for the error variance. The estimates are constructed on MCMC samples from ten parallel chains with a burn-in run of 1,000 iterations each, followed by a sampling phase of 15,000 iterations, with

every tenth iteration used. For modeling smooth terms we use cubic P-spline basis functions with 20 equidistant inner knots over the range of the covariates plus second-order difference penalties penalizing deviations from linearity. For linear/polynomial terms we use orthogonal basis functions of the associated degree without an intercept. For modeling index effects we employ dummy variables with sum-to-zero contrasts. The correlation structures of the random effects ('team-ind' and 'fac-ind') are set to identity here, but more complex classes such as autoregressive or spatial correlations can also be applied.

The model terms with posterior inclusion probability $P(\delta_j = 1|\mathbf{y})$ greater than 0.10 are listed in Table 2 for the primary care relative value unit and in Table 3 for the non-primary care value unit. Compared with the non-primary care RVU, the model for the primary care RVU is rather sparse with only 10 terms with inclusion probability larger than 0.10. In both models, including the team and facility random intercepts accounting for hierarchical heterogeneity turns out to be imperative. Four other terms are also common in the two models, that is, linear part of CAN score, marital status, whether the patient has been diagnosed with a musculoskeletal or connective tissue condition, and whether the patient has had a screening or history of disease. In terms of disease variables, the non-primary care additive predictor is almost entirely dominated by cancer, eye, mental, skin, ear/nose/throat, and injury/poisoning, while nutrition/metabolic and heart diseases are more prominent in the primary care additive predictor. The posterior mean of the nonparametric additive predictor η associated with a number of selected effects along with 90% credible intervals are illustrated in Figures 3-5 for the primary care RVU, and in Figures 6-9 for the non-primary care RVU. As shown in Figure 3, the care assessment need score effect on the primary care RVU is increasing from about -0.2 to +0.2 with a zero effect around 50. However, on the non-primary care RVU, the CAN score has a greater effect changing from -1 to +1 (Figure

6). The effects of comorbidities are shown in Figures 4 and 8. As expected, having a comorbid condition is always associated with greater clinical workload in both primary and non-primary care settings. The interaction effect of CAN score and priority on the non-primary care RVU (Figure 9) shows that patients in priority groups such as 5 and 6 are likely to generate more workloads as their CAN scores increase. Yet, patients in other groups like 8 and 2 have a decreasing trend with regards to increasing in CAN score. The interaction effect of age and provider position on the non-primary care workload is oscillating with a direction change around the age of 58. The effect of length of stay on the non-primary care workload is much higher than all other covariates. Its interaction with priority group is also illustrated in Figure 10 showing a large positive effect in group 7 and a large negative one in group 6.

In the test set containing 1,000 independent patients, the selected covariate set is the same as in Table 2 for the primary care RVU, except that there is no interaction effect identified; for the non-primary care workload prediction, the model includes exactly the same terms as shown in Table 3. This finding assures the stability of our approach and reinforces its internal validity (or reproducibility) with related samples underlying a same population.

We then perform predictive performance evaluation with different hyperparameter settings. To this end, the mean posterior deviance $\frac{1}{T} \sum_t^T -2l(\mathbf{y}|\boldsymbol{\eta}^{(t)}, \boldsymbol{\phi}^{(t)})$, the average of twice the negative log-likelihood of the observations over the saved MCMC iterations, is calculated and saved. Results confirm that the prediction accuracy is very robust across all the parameter combinations for both primary care and non-primary care workloads. However, variable selection is sensitive to varying hyperparameters, especially to the choice of v_0 . Generally, we observe that very small values of v_0 allow small effects to be included in the model, while larger

values of v_0 perform more conservatively. The model sparsity is found to be more sensitive with regard to v_0 than toward (a_ρ, b_ρ) .

Examining hierarchical versus nonhierarchical modeling, we notice that the mean posterior deviance is much smaller when we include random intercepts from level-3 and level-4 hierarchies. Specifically for the primary care RVU in the test set the reductions in deviance are 186 and 53 units with regards to the team and facility intercepts, respectively with the null deviance equal to 1932. For the non-primary care workload these cuts are found to be 197 and 64 units. Ignoring the hierarchical structure of data, which introduces nested correlations among observations, can result in a biased prediction of both outcomes. Finally, in terms of prediction quality, the reduced models consisting of the selected covariates produce about 68% and 73% predictive R-squared (see Appendix B) for the primary care workload and the non-primary care workload, respectively, showing a practically good fit.

6. Conclusion

In this paper, we propose a Bayesian function selection approach based on spike and slab priors for the hierarchical structured additive models with a multivariate response. The prior setting adopted in our work is a Bayesian hierarchical structure with a bimodal density on the hyper-variance of the coefficient blocks with one part being a narrow spike around the origin and the other part being a wide slab. We demonstrate how one can parameterize a special class of multi-response hierarchical structured additive model, that is, a multivariate linear multilevel spline model, within a standard structural equation modeling framework, and thus bridge the connection between multivariate multilevel STAR models and generalized latent variable models.

We then apply our methods to patient centered medical home data obtained from a large number of VA medical facilities during fiscal years 2011–12. Our work is the first attempt to

develop a portfolio based demand prediction model for patient centered medical home within the OR/MS or IE community. We aggregate three levels of hierarchical data including information from outpatients, the medical team responsible to render the care to the patients, and the VA facilities. We find that the sets of chosen predictors introduced by the model are different for the primary care and the non-primary workloads. Our findings also confirm that taking hierarchical heterogeneity into account is associated with better prediction accuracy, especially when the data has more than two levels.

Some methodological directions based on our approach can be investigated in future research. One challenging extension would be to develop Bayesian model choice and component selection for multi-response hierarchical auto-logistic or auto-Poisson regressions particularly used in ecology or hierarchical seemingly unrelated regressions in economics. Another direction can be Bayesian function selection in semiparametric quantile regressions with non-normal random effects modeled by Dirichlet process mixture.

Acknowledgments

The authors thank the editor and three anonymous referees for their constructive comments.

Funding

This research is supported by the National Science Foundation, Division of Civil, Mechanical, and Manufacturing Innovation (CMMI) under grant number 1233504.

Appendices

Appendix A

The MCMC algorithm in Section 4.3 is described as follows.

Initiate $\boldsymbol{\rho}^{(0)}, \boldsymbol{\delta}^{(0)}, \boldsymbol{\phi}^{(0)}, \boldsymbol{\omega}^{(0)}$, and $\boldsymbol{\beta}^{(0)}$ (via IWLS if the response is non-Gaussian)

Calculate $\boldsymbol{\alpha}^{(0)}, \boldsymbol{\Xi}^{(0)}, \mathbf{X}_\alpha^{(0)}$

for iterations $t = 1, \dots, T$ **do**

for blocks $\ell = 1, \dots, \ell_\alpha$ **do**

 update $\alpha_\ell^{(t)}$ by its FCD (formula (23) if Gaussian or IWLS if non-Gaussian)

 set $\mathbf{X}_\Xi^{(t)} = \mathbf{X} \text{blockdiag}(\mathbf{1}_{e1}, \dots, \mathbf{1}_{ep}) \boldsymbol{\alpha}^{(t)}$

 update $r_1^{(t)}, \dots, r_L^{(t)}$ via their FCD: $P(r_l^{(t)} = 1 | \cdot) = \left(1 + \exp(-2\Xi_l^{(t)})\right)^{-1}, l = 1, \dots, L$

for blocks $\ell = 1, \dots, \ell_\Xi$ **do**

 update $\Xi_\ell^{(t)}$ by its FCD (formula (24) if Gaussian or IWLS if non-Gaussian)

for model terms $j = 1, \dots, p$ **do**

 rescale $\Xi_j^{(t)}$ and $\alpha_j^{(t)}$ by $\Xi_j \rightarrow \frac{M_j}{\sum_i M_j |\Xi_{ji}|} \Xi_j$ and $\alpha_j \rightarrow \frac{\sum_i M_j |\Xi_{ji}|}{M_j} \alpha_j$

 set $\mathbf{X}_\alpha^{(t)} = \mathbf{X} \text{blockdiag}(\Xi_1^{(t)}, \dots, \Xi_p^{(t)})$

 update $\rho_1^{2(t)}, \dots, \rho_p^{2(t)}$ from their FCD: $\rho_j^{2(t)} | \cdot \sim \Gamma^{-1}\left(a_\rho + \frac{1}{2}, b_\rho + \frac{\alpha_j^{2(t)}}{2\delta_j^{(t)}}\right)$

 update $\delta_1^{(t)}, \dots, \delta_p^{(t)}$ from their FCD: $\frac{P(\delta_j^{(t)}=1|\cdot)}{P(\delta_j^{(t)}=v_0|\cdot)} = v_0^{1/2} \exp\left(\frac{(1-v_0)\alpha_j^{2(t)}}{2v_0\rho_j^{2(t)}}\right)$

 update $\omega^{(t)}$ from its FCD: $\omega^{(t)} | \cdot \sim \text{Beta}(a_\omega + \sum_j^p I_1(\delta_j^{(t)}), b_\omega + \sum_j^p I_{v_0}(\delta_j^{(t)}))$

if $\mathbf{y}$ is Gaussian then

$$\text{update } \phi^{(t)} \text{ from its FCD: } \phi^{(t)} | \cdot \sim \Gamma^{-1} \left(a_\phi + \frac{n}{2}, b_\phi + \frac{\sum_i^n (y_i - \eta_i^{(t)})^2}{2} \right)$$

Appendix B

The predictive R-squared can be defined similar to model based R-squared in order to assess the linear correlation between outcome y_o and its prediction y_p . It is bounded to the interval $[0,1]$. If we denote the arithmetic means of the observed and predicted outcomes as $\bar{y}_o$ and $\bar{y}_p$, respectively, the predictive R-squared is given as $R_p^2 = \frac{\sum_i^n (y_{p,i} - \bar{y}_p)(y_{o,i} - \bar{y}_o)}{\sqrt{\sum_i^n (y_{p,i} - \bar{y}_p)^2} \sqrt{\sum_i^n (y_{o,i} - \bar{y}_o)^2}}$. Note that, unlike the model based R-squared, the R_p^2 cannot be interpreted as percentage of variance explained because the decomposition of variance holding for estimated values $\hat{y}$ does not apply for predicted values y_p .

References

- Ajorlou, S., Shams, I. & Yang, K. (2014). An analytics approach to designing patient centered medical homes. *Health Care Management Science*, Ahead-of-print, <http://dx.doi.org/10.1007/s10729-014-9287-x>, accessed July 12, 2014.
- Andridge, R.R. & Little, R.J. (2010). A review of hot deck imputation for survey non - response. *International Statistical Review*, 78(1), 40-64.
- Antoniadis, A. & Fan, J. (2001). Regularization of wavelet approximations. *Journal of the American Statistical Association*, 96(455), 939-967.
- Bitton, A., Martin, C. & Landon, B.E. (2010). A nationwide survey of patient centered medical home demonstration projects. *Journal of General Internal Medicine*, 25(6), 584-592.
- Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, 37(4), 373-384.
- Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *Annals of statistics*, 24(6), 2350-2383.
- Brown, P.J., Vannucci, M. & Fearn, T. (1998). Multivariate Bayesian variable selection and prediction. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 60(3), 627-641.
- Cai, J., Fan, J., Li, R. & Zhou, H. (2005). Variable selection for multivariate failure time data. *Biometrika*, 92(2), 303-316.
- Chen, Y., Du, P. & Yuedong, W. (2013). Variable selection in linear models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(1), 1-9.

- Claeskens, G. & Hjort, N.L. (2003). The focused information criterion. *Journal of the American Statistical Association*, 98(464), 900-916.
- Dummit, L. (2009). Relative Value Units (RVUs). In: National Health Policy Forum Publications. Available at http://www.nhpf.org/library/the-basics/Basics_RVUs_02-12-09.pdf, Accessed on March 18, 2014.
- Fahrmeir, L., Kneib, T. & Lang, S. (2004). Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica*, 14(3), 731-762.
- Fan, J. & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348-1360.
- Fan, J. & Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1), 101-148.
- Fisher, E.S. (2008). Building a medical neighborhood for the medical home. *New England Journal of Medicine*, 359(12), 1202-1205.
- George, E.I. & McCulloch, R.E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423), 881-889.
- Goldstein, H. (2011). *Multilevel statistical models*: John Wiley & Sons.
- Hastie, T. & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 55(4), 757-796.
- Huang, J., Horowitz, J.L. & Wei, F. (2010). Variable selection in nonparametric additive models. *Annals of statistics*, 38(4), 2282-2313.
- Ishwaran, H. & Rao, J.S. (2005). Spike and slab variable selection: frequentist and Bayesian strategies. *Annals of statistics*, 33(2), 730-773.

- Kammann, E. & Wand, M.P. (2003). Geoadditive models. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 52(1), 1-18.
- Klein, S. & Fund, C. (2011). *The Veterans Health Administration: implementing patient-centered medical homes in the nation's largest integrated delivery system*: Commonwealth Fund.
- Kline, R.B. (2011). *Principles and practice of structural equation modeling*: Guilford press.
- Kundu, S. & Dunson, D.B. (2013). Bayes variable selection in semiparametric linear models. *Journal of the American Statistical Association*(just-accepted).
- Lang, S. & Brezger, A. (2004). Bayesian P-splines. *Journal of computational and graphical statistics*, 13(1), 183-212.
- Lang, S., Umlauf, N., Wechselberger, P., Harttgen, K. & Kneib, T. (2013). Multilevel structured additive regression. *Statistics and Computing*, 24(2), 223-238.
- Liang, F., Paulo, R., Molina, G., Clyde, M.A. & Berger, J.O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481), 410-423.
- Lin, Y. & Zhang, H.H. (2006). Component selection and smoothing in multivariate nonparametric regression. *Annals of statistics*, 34(5), 2272-2297.
- Marra, G. & Wood, S.N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis*, 55(7), 2372-2387.
- Müller, S., Scealy, J. & Welsh, A. (2013). Model selection in linear mixed models. *Statistical Science*, 28(2), 135-167.
- O'Hara, R.B. & Sillanpää, M.J. (2009). A review of Bayesian variable selection methods: what, how and which. *Bayesian analysis*, 4(1), 85-117.

- Pu, W. & Niu, X.-F. (2006). Selecting mixed-effects models based on a generalized information criterion. *Journal of Multivariate Analysis*, 97(3), 733-758.
- Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69(2), 167-190.
- Reid, R.J., Coleman, K., Johnson, E.A., Fishman, P.A., Hsu, C., Soman, M.P., Trescott, C.E., Erikson, M. & Larson, E.B. (2010). The group health medical home at year two: cost savings, higher patient satisfaction, and less burnout for providers. *Health Affairs*, 29(5), 835-843.
- Rittenhouse, D.R. & Shortell, S.M. (2009). The patient-centered medical home: will it stand the test of health reform? *Journal of the American Medical Association*, 301(19), 2038-2040.
- Scheipl, F., Fahrmeir, L. & Kneib, T. (2012). Spike-and-slab priors for function selection in structured additive regression models. *Journal of the American Statistical Association*, 107(500), 1518-1532.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 64(4), 583-639.
- Stange, K.C., Nutting, P.A., Miller, W.L., Jaén, C.R., Crabtree, B.F., Flocke, S.A. & Gill, J.M. (2010). Defining and measuring the patient-centered medical home. *Journal of General Internal Medicine*, 25(6), 601-612.
- Steele, F. (2008). Multilevel models for longitudinal data. *Journal of the Royal Statistical Society : Series A (Statistics in Society)*, 171(1), 5-19.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 58(1), 267-288.

Biographies

Issac Shams is a postdoctoral research fellow in the Department of Industrial and Operations Engineering at the University of Michigan. He received his B.Sc. and M.Sc. in Industrial Engineering from Iran University of Science and Technology in 2008 and 2011 respectively, and his Ph.D. in Industrial and Systems Engineering from Wayne State University in 2014. His research interests include healthcare-driven statistical modeling, statistical network analysis, and statistical learning for knowledge discovery and process improvement. He is a member of ASQ, IIE, INFORMS, ASA, and IMS.

Saeede Ajorlou is a postdoctoral visiting scholar in the Department of Industrial and Operations Engineering at the University of Michigan. She received her B.Sc. in computer engineering from Mazandaran University of Science and Technology in 2007, her M.Sc. in Industrial Engineering from Iran University of Science and Technology in 2008, and her Ph.D. in Industrial and Systems Engineering from Wayne State University in 2014. Her research focus is on the application and developments of operations research methods in modeling and control of stochastic systems in healthcare operations and production and operations management. She is a member of IIE and INFORMS.

Kai Yang is a Professor in the Department of Industrial and System Engineering, and the director of Healthcare Systems Engineering Group at Wayne State University. His areas of research include statistical methods in quality and reliability, healthcare systems engineering and engineering design methodologies. Dr Yang's research has been funded by such organizations as NSF, VA, GM, Ford, and Siemens. Dr Yang is currently a leading faculty member in US Veteran Administration (VA) Center for Applied System Engineering, which is a nationwide VA initiative to use industrial engineering to improve healthcare industry since 2009, in which Dr Yang is leading many projects involving healthcare access improvement, healthcare data analytics, readmission reduction, real time location system in healthcare, and patient centered medical homes. Dr. Yang obtained both his MS and PhD degrees from the University of Michigan.

Table 1: Description of the predictors and response variables included in the study ($n = 10,000$)

Variable	Description	Summary statistics
<i>Patient level predictors</i>		
gen	Gender	Male (93.55%), Female (6.45%)
age	Age (years)	Mean: 61.74, StdDev: 15.19, Min: 20, Max: 98
mar	Marital status	Married (55.41%), Not married (15.43%), Previously married (28.62%), Unknown (0.54%)
ins	Insurance status	Insured (58.46%), Not insured (41.54%)
emp	Employment status	Active military service (0.15%), Employed full time (21.03%), Employed part time (5.07%), Not employed (37.18%), Retired (33.19%), Self-employed (2.41%), Unknown (0.97%)
prio	Enrollment priority	Group 1 (24.86%), Group 2 (8.23%), Group 3 (12.35%), Group 4 (2.78%), Group 5 (27.58%), Group 6 (4.44%), Group 7 (2.62%), Group 8 (17.14%)
los	Length of stay (days)	Mean: 0.88, StdDev: 6.18, Min: 0, Max: 210
can	Care assessment need score	Mean: 55, StdDev: 28.01, Min: 0, Max: 99
team-ind	Index of PCMH team	1301 categories; 800000054 (0.06%), 1000003172 (0.05%), ...
acc1-ind	Has been diagnosed with infectious or parasitic condition?	Yes (12.45%), No (87.55%)
acc2-ind	Has been diagnosed with malignant neoplasm?	Yes (10.28%), No (89.72%)
acc3-ind	Has been diagnosed with benign/in situ/uncertain neoplasm?	Yes (10.79%), No (89.21%)
acc4-ind	Has been diagnosed with diabetes?	Yes (28.7%), No (71.3%)
acc5-ind	Has been diagnosed with nutritional or metabolic disease?	Yes (70.01%), No (29.99%)
acc6-ind	Has been diagnosed with liver disease?	Yes (5.07%), No (94.93%)

Table 1—continued

Variable	Description	Summary statistics
acc7-ind	Has been diagnosed with gastrointestinal condition?	Yes (33.84%), No (66.16%)
acc8-ind	Has been diagnosed with musculoskeletal or connective tissue condition?	Yes (59.93%), No (40.07%)
acc9-ind	Has been diagnosed with hematological condition?	Yes (10.22%), No (89.78%)
acc10-ind	Has been diagnosed with cognitive disorders?	Yes (5.25%), No (94.75%)
acc11-ind	Has been diagnosed with substance abuse?	Yes (23.25%), No (76.75%)
acc12-ind	Has been diagnosed with mental condition?	Yes (37.38%), No (62.62%)
acc13-ind	Has been diagnosed with developmental disability?	Yes (0.89%), No (99.11%)
acc14-ind	Has been diagnosed with neurological condition?	Yes (16.73%), No (83.27%)
acc15-ind	Has been diagnosed with cardio-respiratory arrest?	Yes (1.4%), No (98.6%)
acc16-ind	Has been diagnosed with heart disease?	Yes (66.36%), No (33.64%)
acc17-ind	Has been diagnosed with cerebrovascular condition?	Yes (6.25%), No (93.75%)
acc18-ind	Has been diagnosed with vascular condition?	Yes (11.69%), No (88.31%)
acc19-ind	Has been diagnosed with lung disease?	Yes (18.12%), No (81.88%)
acc20-ind	Has been diagnosed with eyes condition?	Yes (38.66%), No (61.34%)
acc21-ind	Has been diagnosed with ears, nose, and throat condition?	Yes (37.82%), No (62.18%)
acc22-ind	Has been diagnosed with urinary system disease?	Yes (15.89%), No (84.11%)
acc23-ind	Has been diagnosed with genital system disease?	Yes (21.9%), No (78.1%)

Table 1—continued

Variable	Description	Summary statistics
acc24-ind	Has been diagnosed with pregnancy-related condition?	Yes (0.19%), No (99.81%)
acc25-ind	Has been diagnosed with skin or subcutaneous condition?	Yes (23.69%), No (76.31%)
acc26-ind	Has been diagnosed with injury, poisoning, or complications?	Yes (15.09%), No (84.91%)
acc27-ind	Has been diagnosed with symptoms, signs, or ill-defined conditions?	Yes (59.95%), No (40.05%)
acc29-ind	Has been diagnosed with transplants, openings, or amputations condition?	Yes (1.44%), No (98.56%)
acc30-ind	Has been diagnosed with screening/history?	Yes (92.08%), No (7.92%)
<i>Team level predictors</i>		
fac-ind	Index of PCMH facility	260 categories; Dallas VA Medical Center (0.84%), San Diego Community-based Outpatient Clinic (0.64%), ...
prov.pos	Assigned provider position	Primary care physician (68.74%), Nurse practitioner (15.90%), Attending physician (8.87%), Assistant physician (6.49%)
prov.exp	Assigned provider experience (years)	Mean: 8.55, StdDev: 7.79, Min: 0, Max: 41
prov.chng	# times the patient has changed his/her assigned provider	Mean: 0.75, StdDev: 0.90, Min: 0, Max: 9
prov.fte	Provider full time equivalent	Mean: 0.85, StdDev: 0.24, Min: 0, Max: 1
<i>Facility level predictors</i>		
fac.dist	Distance between patient's home and his/her assigned facility (miles)	Mean: 79.910, StdDev: 744.3, Min: 0.018, Max: 12430

Table 1—continued

Variable	Description	Summary statistics
<i>Patient level response variables</i>		
pcrvu	Primary care relative value unit	Mean: 3.96, StdDev: 2.82, Min: 0.17, Max: 36.74
npcrvu	Non-primary care relative value unit	Mean: 14.93, StdDev: 22.83, Min: 0.06, Max: 371.6

Table 2: Posterior means of marginal inclusion probabilities for primary care relative value unit

Term	$P(\delta_j = 1 \mathbf{y})$
Team, random intercept	1.000
Facility, random intercept	0.762
Care assessment need score, linear	0.614
Has been diagnosed with screening/history, factor	0.583
Has been diagnosed with symptoms, signs, or ill-defined conditions, factor	0.335
Has been diagnosed with nutritional or metabolic disease, factor	0.309
Has been diagnosed with musculoskeletal or connective tissue condition, factor	0.282
Has been diagnosed with heart disease, factor	0.247
Marital status	0.105
Linear (age) : smooth (prov.exp)	0.100

Table 3: Posterior means of marginal inclusion probabilities for non-primary care relative value unit

Term	$P(\delta_j = 1 y)$
Team, random intercept	1.000
Care assessment need score, linear	1.000
Age, linear	1.000
Age, smooth	1.000
Has been diagnosed with benign/in situ/uncertain neoplasm, factor	1.000
Has been diagnosed with eyes condition, factor	1.000
Has been diagnosed with mental condition, factor	0.999
Has been diagnosed with skin or subcutaneous condition, factor	0.998
Facility, random intercept	0.998
Care assessment need score, smooth	0.997
Has been diagnosed with symptoms, signs, or ill-defined conditions, factor	0.993
Has been diagnosed with screening/history, factor	0.978
Has been diagnosed with malignant neoplasm, factor	0.976
Enrollment priority, factor	0.953
Has been diagnosed with ears, nose, and throat condition, factor	0.946
Has been diagnosed with injury, poisoning, or complications, factor	0.851
Has been diagnosed with musculoskeletal or connective tissue condition, factor	0.805
Linear (care assessment need score) : factor (enrollment priority)	0.537
Marital status	0.394

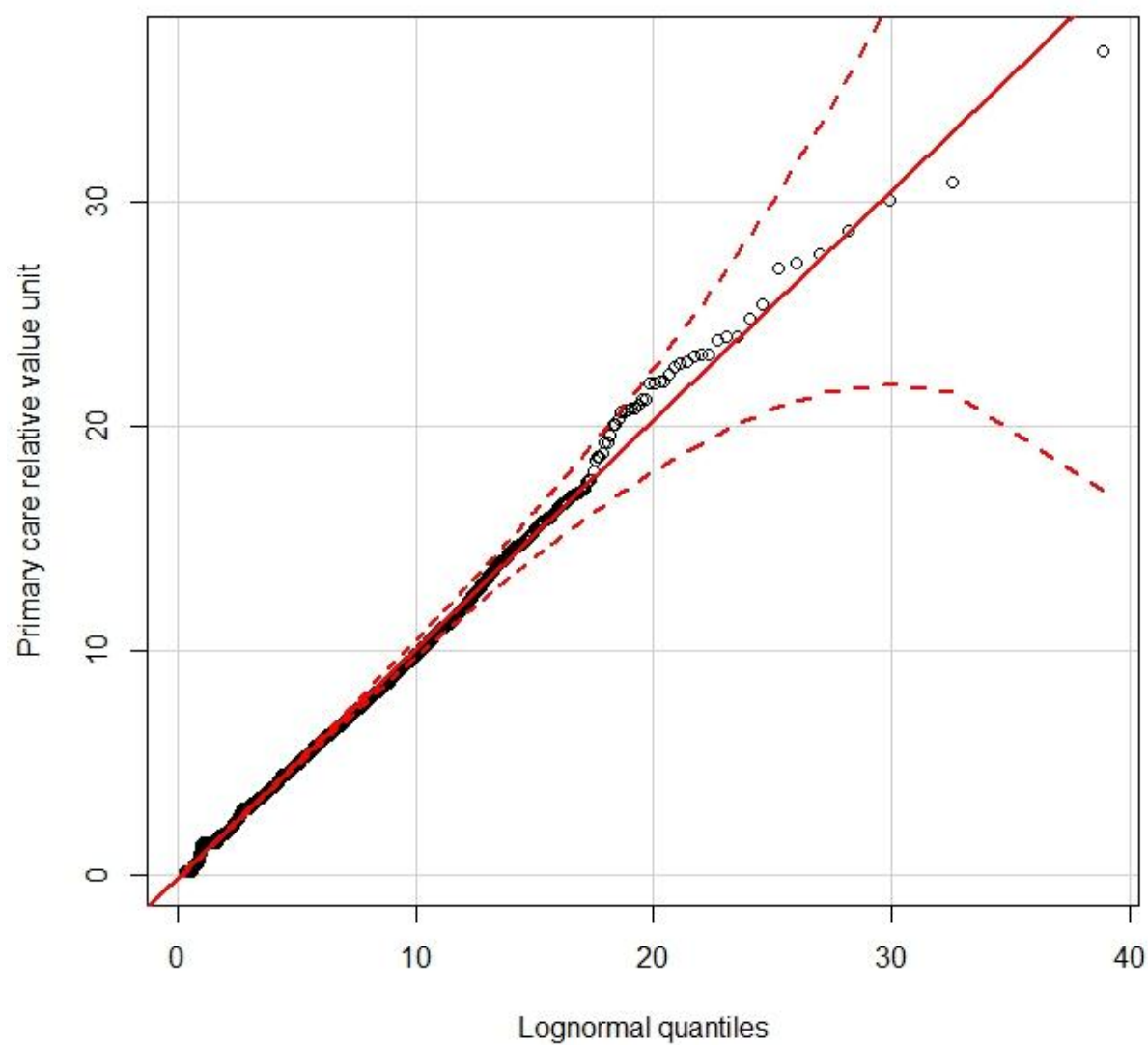


Figure 1: QQ plot of primary care relative value unit with 95% confidence bands

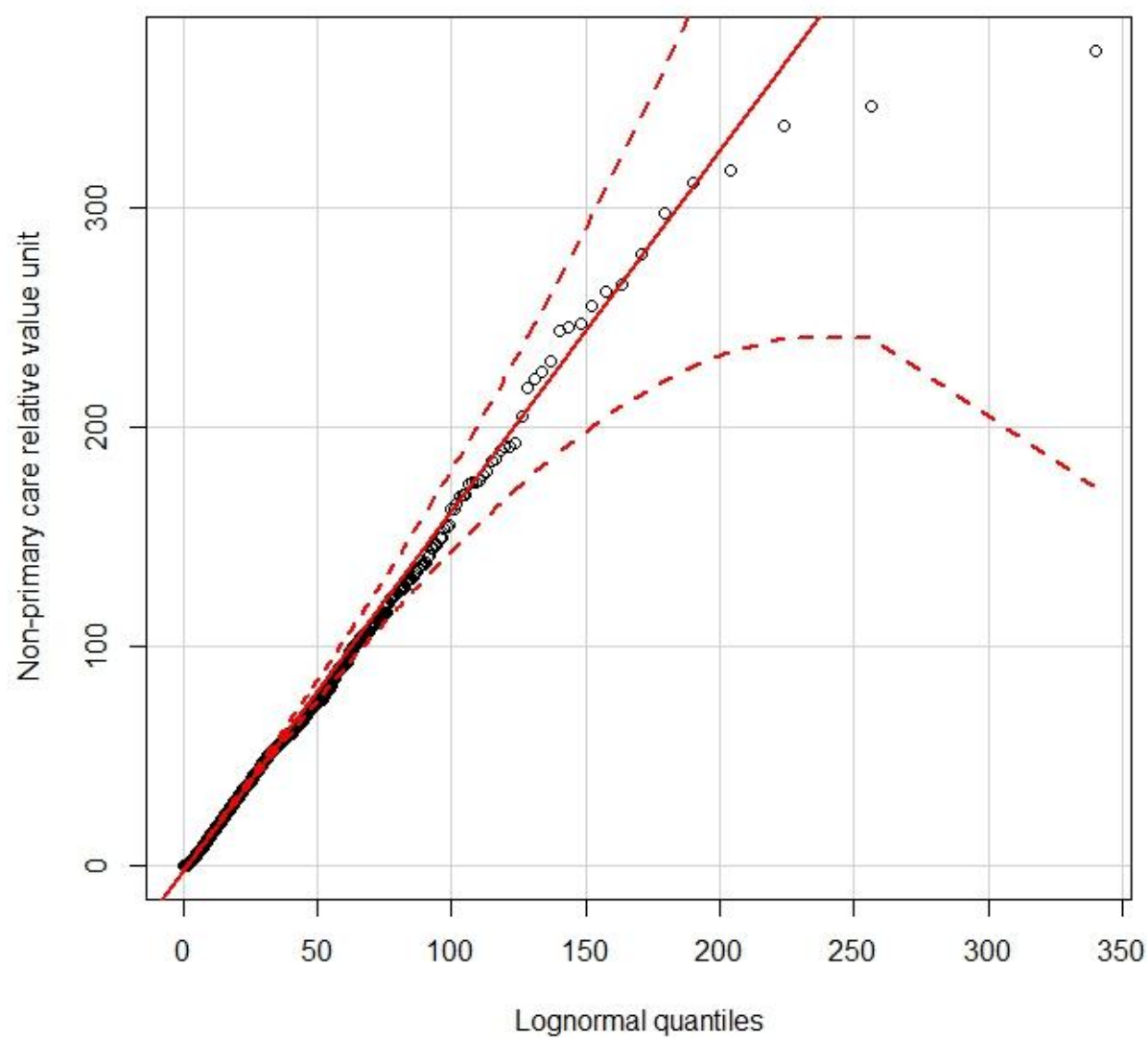


Figure 2: QQ plot of non primary care relative value unit with 95% confidence bands

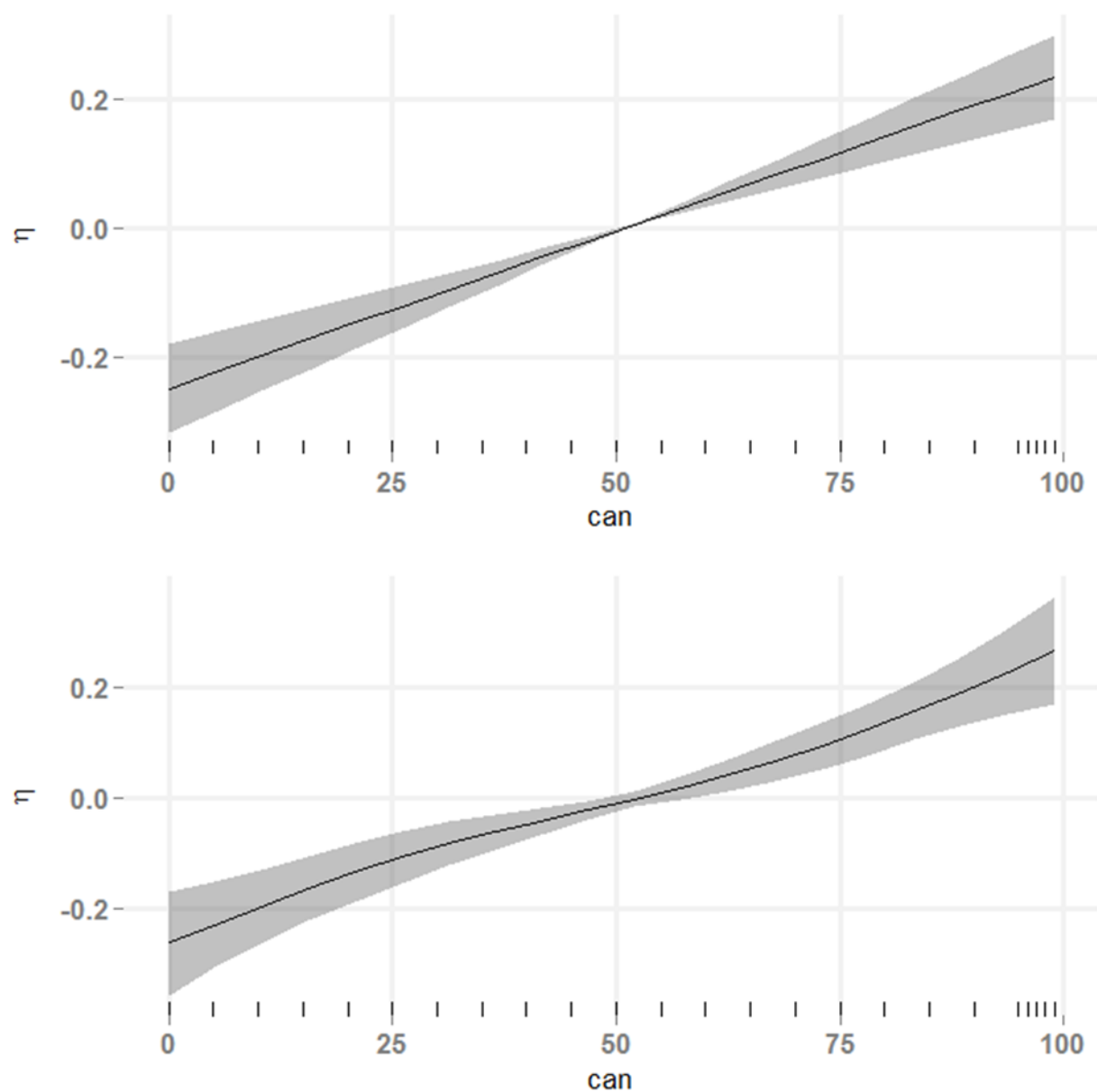


Figure 3: Linear (top) and nonlinear (bottom) effects of care assessment need score on the primary care predictor with 90% credible intervals

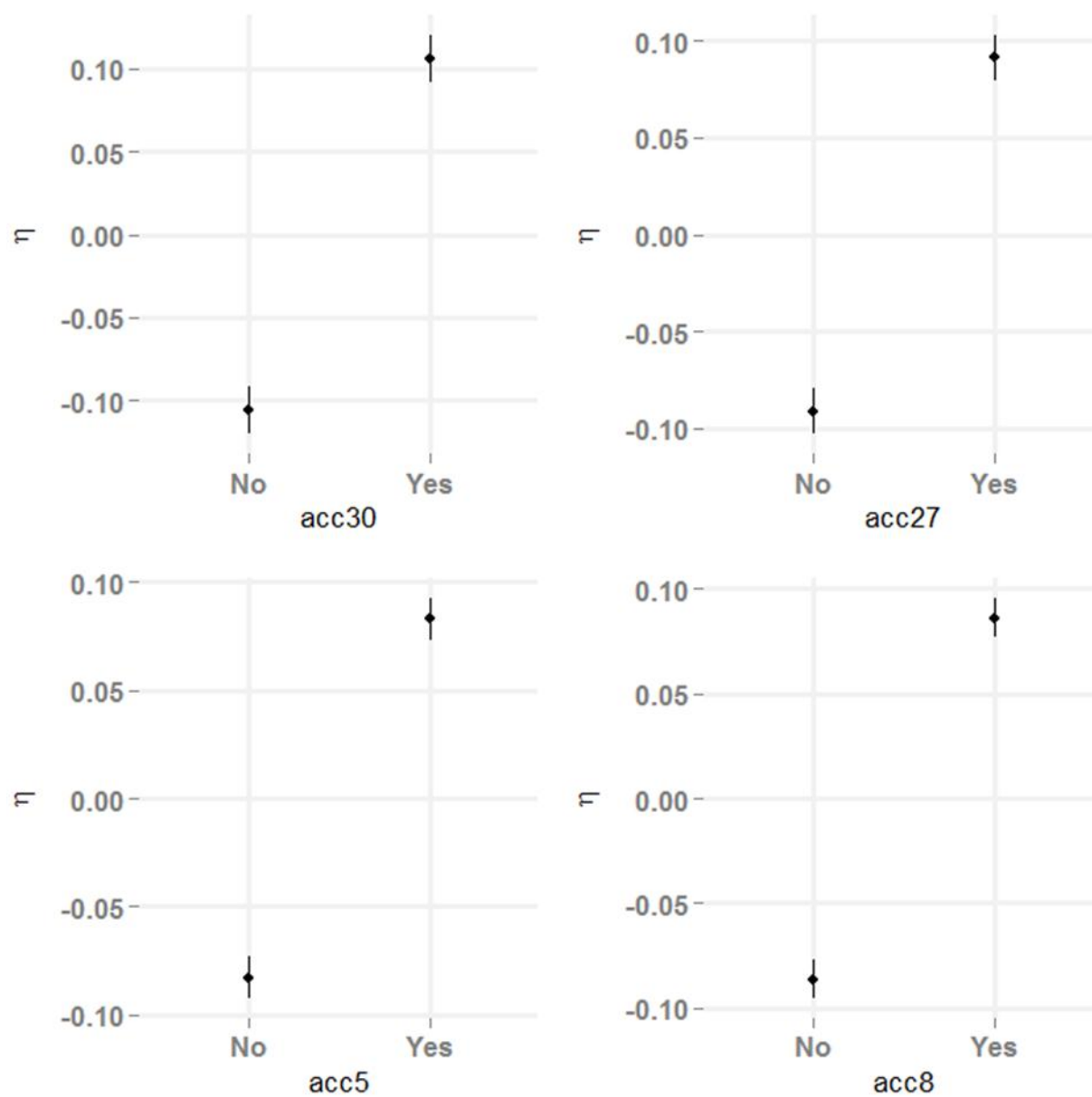


Figure 4: Effects of different comorbid conditions on the primary care predictor with 90% credible intervals

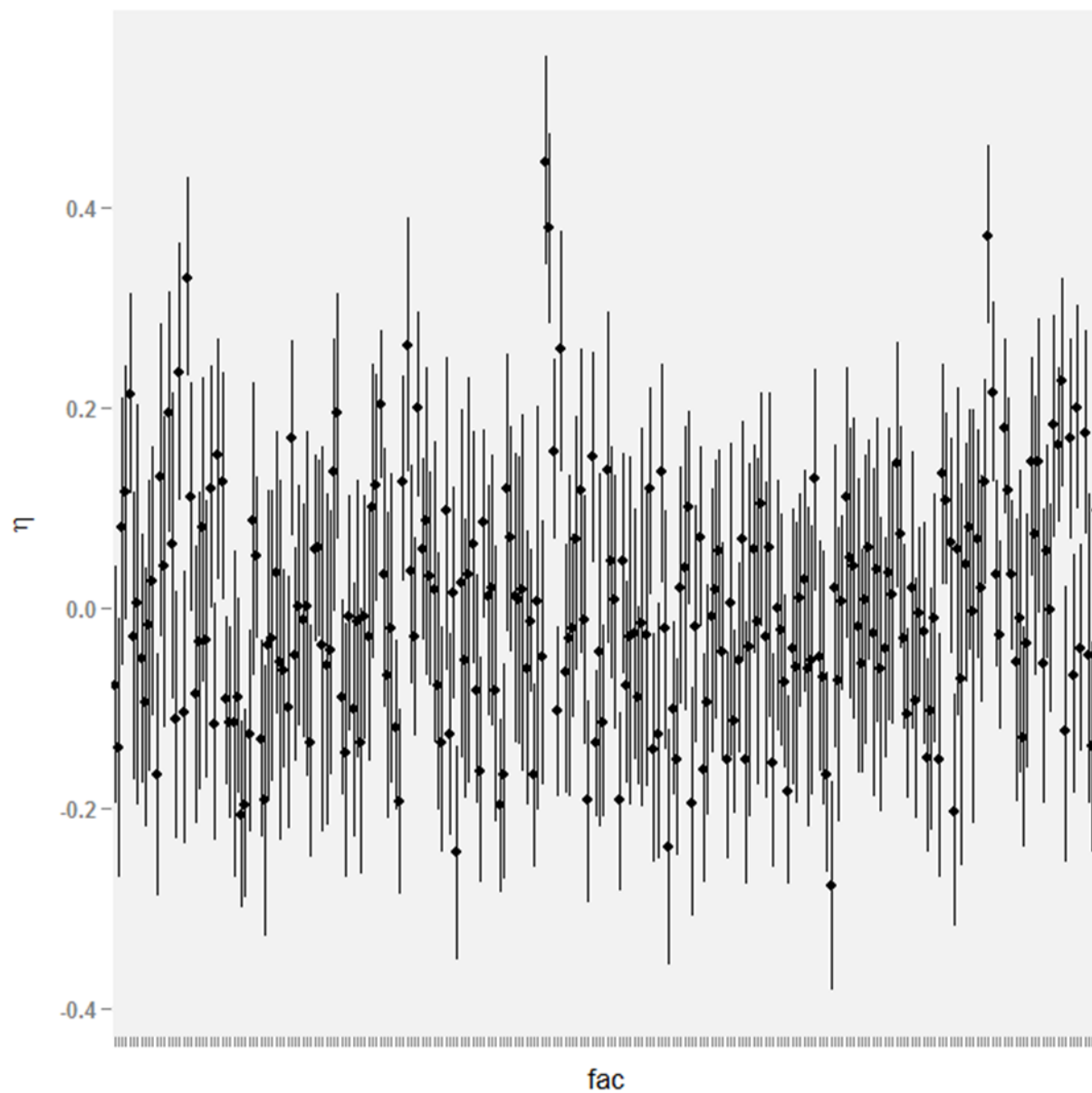


Figure 5: Effects of different facility on the primary care predictor with 90% credible intervals

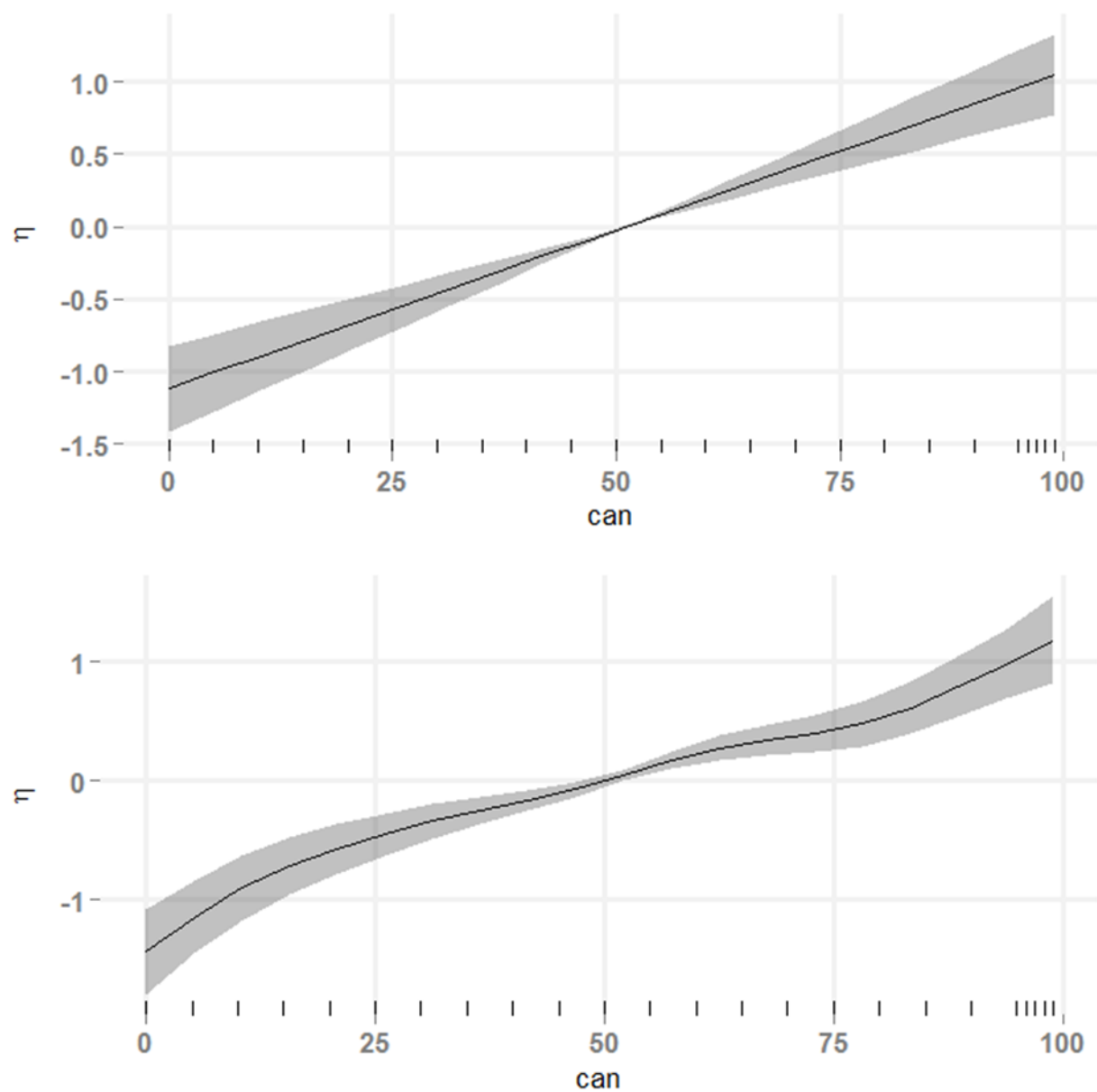


Figure 6: Linear (top) and nonlinear (bottom) effects of care assessment need score on the non primary care predictor with 90% credible intervals

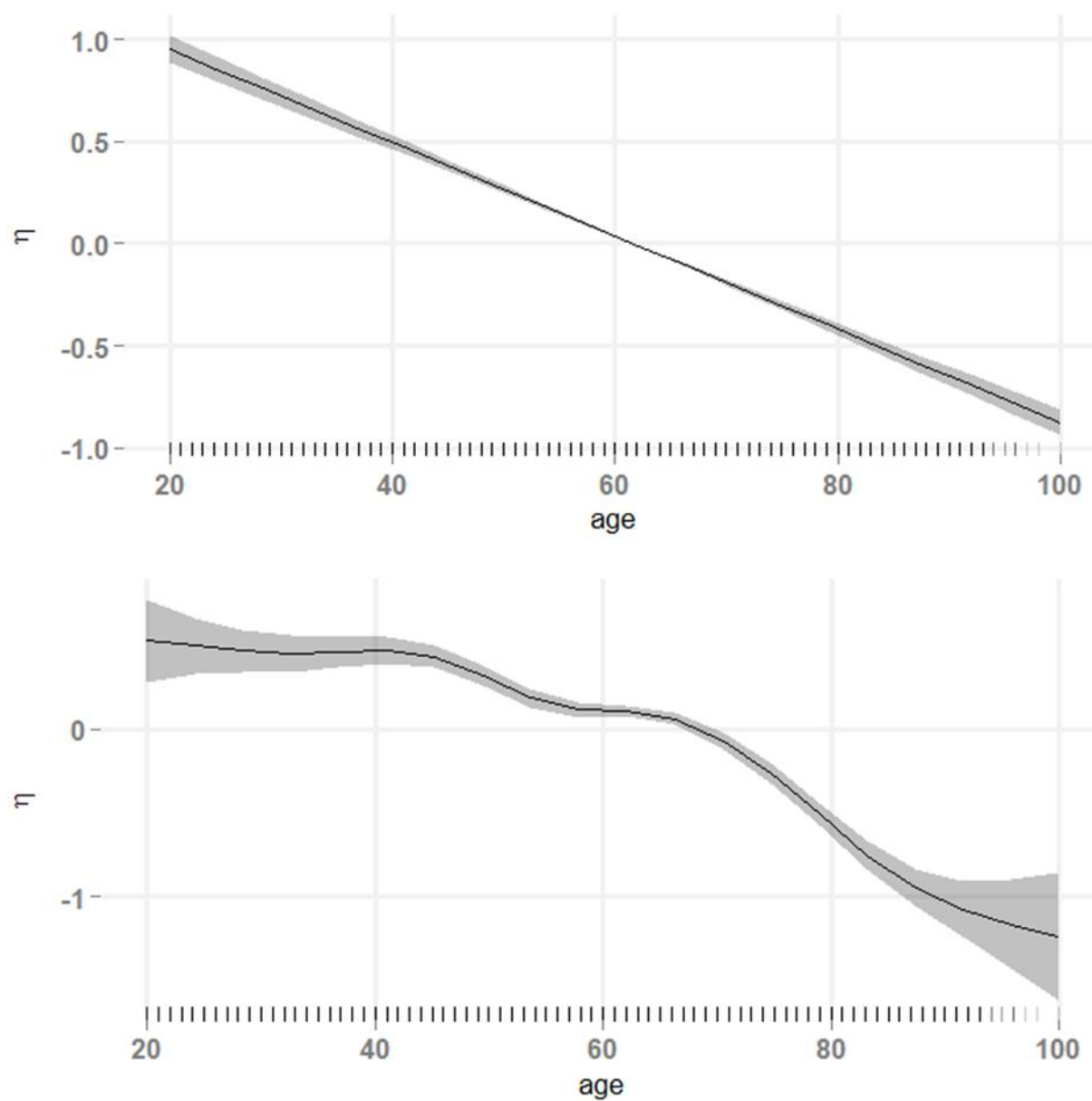


Figure 7: Linear (top) and nonlinear (bottom) effects of age on the non primary care predictor with 90% credible intervals

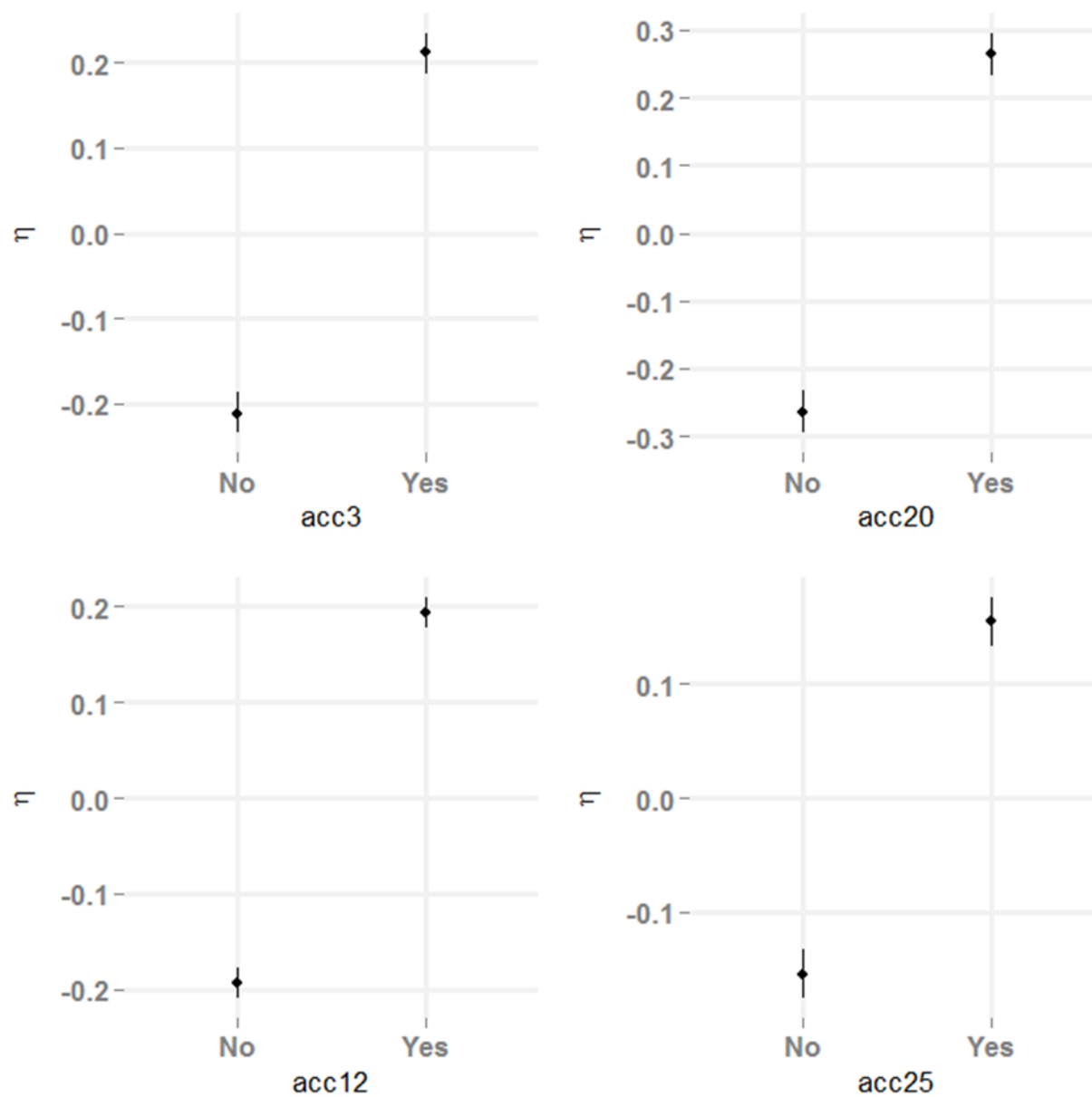


Figure 8: Effects of different comorbid conditions on the non primary care predictor with 90% credible intervals

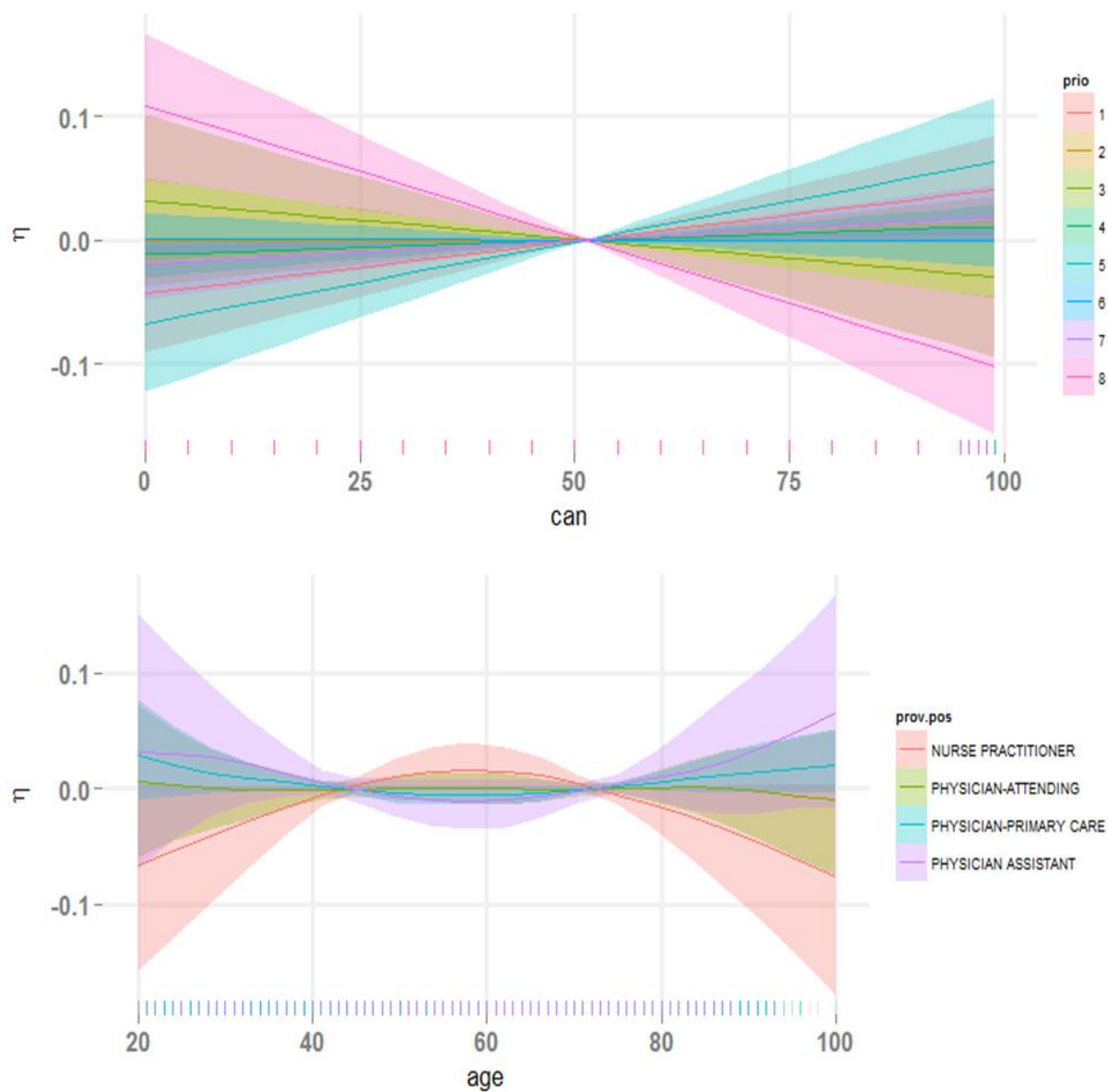


Figure 9: Interaction effects of care assessment need score and enrollment priority (top), age and assigned provider position (bottom) on the non primary care predictor with 90% credible intervals

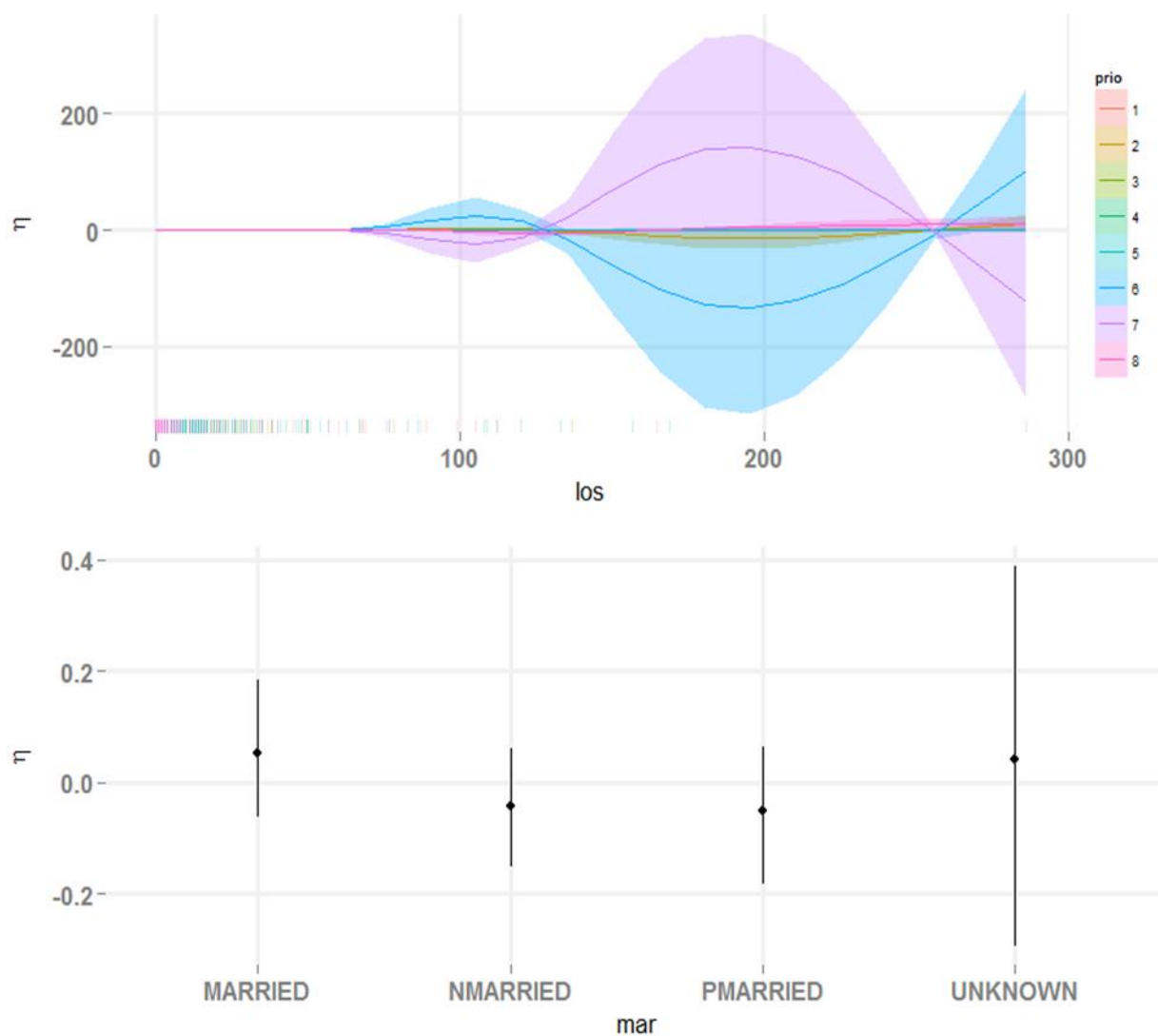


Figure 10: Interaction effects of length of stay and enrollment priority (top), and effect of marital status (bottom) on the non primary care predictor with 90% credible intervals. ('nmarrried' and 'pmarrried' stand for not married and previously married)