Bayesian Modeling Strategies for Complex Data Structures, with Applications to Neuroscience and Medicine

Feihan Lu

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

ABSTRACT

Bayesian Modeling Strategies for Complex Data Structures, with Applications to

Neuroscience and Medicine

Feihan Lu

Bayesian statistical procedures use probabilistic models and probability distributions to summarize data, estimate unknown quantities of interest, and predict future observations. The procedures borrow strength from other observations in the dataset by using prior distributions and/or hierarchical model specifications. The unique posterior sampling techniques can handle different issues, e.g., missing data, imputation, and extraction of parameters (and their functional forms) that would otherwise be difficult to address using conventional methods. In this dissertation, we propose Bayesian modeling strategies to address various challenges arising in the fields of neuroscience and medicine. Specifically, we propose a sparse Bayesian hierarchical Vector Autoregressive (VAR) model to map human brain connectivity using multi-subject multi-session functional magnetic resonance image (fMRI) data. We use the same model on patient diary databases, focusing on patient-level prediction of medical conditions using posterior predictive samples. We also propose a Bayesian model with an augmented Markov Chain Monte Carlo (MCMC) algorithm on repeat Electrical Stimulation Mappings (ESM) to evaluate the variability of localization in brain sites responsible for language function. We close by using Bayesian disproportionality analyses on spontaneous reporting system (SRS) databases for post-market drug safety surveillance, illustrating the caution required in real-world analysis and decision making.

Contents

List of Figures

List of Tables

Acknowledgements

Foremost, I would like to express my upmost gratitude to my advisor Professor Michael Sobel and my co-advisor Professor David Madigan for their tremendous mentoring and invaluable support for my PhD research. I thank Professor Michael Sobel for taking his time and effort clarify my thoughts and guide me through every building block of my research work. I thank Professor David Madigan for continuously providing invaluable insights and inspiration for the data analysis and computation. The completion of this dissertation would not have been possible without guidance from them.

I would also like to thank my committee members Professor Tian Zheng, Professor Eva Petkova and Professor Martin Lindquist for kindly agreeing to serve on my committee. They offered tremendous suggestions to improve my research at various phases. I do hereby acknowledge all of them. I thank Professor Tian Zheng for providing helpful suggestions on conducting simulation studies. I thank Professor Eva Petkova for extending her help on literature reviews. I also thank Professor Martin Lindquist for providing data and interpretations in the real data analysis.

Finally, I wish to offer my deepest gratitude to my family for their consistent support and encouragement on every stage of my personal and academic life. This dissertation is dedicated to them.

To Tora and Jake

Introduction

Bayesian statistical procedures use probabilistic models and probability distributions to summarize data, estimate unknown quantities of interest, and predict future observations (Gelman et al. 2014, Green et al. 2015). Bayesian procedures can handle datasets with complex structures using various multilayer modeling strategies. By using prior distributions and/or hierarchical model specifications, the procedures allow partial pooling of data and parameters, leading in general to greater accuracy than methods that either admit no pooling (e.g., subject specific estimates) or demand complete pooling (e.g., estimates that average over all subjects). Further, Bayesian posterior sampling techniques can handle missing data imputation, and extraction of parameters (and their functional forms) more conveniently than conventional methods (e.g., maximum likelihood).

Over the past few decades, Bayesian methods have been widely applied in many fields, including economics (Greenberg 2012, Karlsson 2015), social science (Jackman 2009), ecology (McCarthy 2007) and healthcare (Spiegelhalter et al. 2004, Aven and Eidesen 2007). In healthcare practice, the sample size to estimate certain quantity of interest may be too small to draw accurate medicine conclusions. For example, the occurrence of a rare disease could be a very small number and thus treatment for such disease could be vague. This is similar for personalized treatment, where targeted medical decisions could be made based on limited historical information. One remarkable advantage of Bayesian methods in these settings is that it is able to borrow information from other less rare diseases or from other patients with more historical information, leading to deeper understanding of the underlying rare disease or more effective treatment for individual patients. Further, in neuroscience, the data could have complex structure that requires flexible modeling strategy to extract the desired information.

In this dissertation, we use Bayesian methodology to construct new models and methods for analyzing complex data from neuroscience and medicine.

In Chapter 1, we propose a sparse Bayesian hierarchical Vector Autoregressive (VAR) model to map human brain connectivity. In particular, we focus on resting state functional connectivity, using multi-subject multi-session functional magnetic resonance image (fMRI) time series data. The biggest challenge of this application is the complex structure of the dataset, which consists of multivariate time series observed for many subjects, with each subject observed at several occasions. Our model uses a flexible model specification to incorporate such complex data structure, allowing the decomposition of functional connectivity into its group, subject, and subject-session components and estimation of between subject and between session variabilities. This provides us with the opportunity to investigate whether connectivity is consistent across subjects and replications, something that to date has not been extensively explored. By adopting the doubly adaptive Elastic-net prior (Gefang 2014), our model induces sparsity in parameter estimation without assuming special structure for the error covariance matrix. Using parameter-expanded Gibbs sampler (Higdon 1998), our model greatly speeds up computation and can handle a larger number of brain regions than in previous works. A comprehensive simulation study and a real data example from the Human Connectome Project database (`http://www.humanconnectomeproject.org`) are used to illustrate the performance of our model.

In Chapter 2, we use the same sparse Bayesian hierarchical VAR model employed in Chapter 1 on the patient diary databases, focusing on patient-level prediction of medical conditions using posterior predictive samples. In many patient diary databases, longitudinal data on the activities of daily living, moods, personal habits, diseases or disorders of individual patients are collected, based on which patient-level prediction is often conducted. Existing studies commonly fit separate VAR models for each individual patient. While this method may be sufficient for patients who have an adequate number of observations, it may fail to provide

reliable prediction for others who may have a significantly smaller number of observations due to late participation or dropout in the study. Using the hierarchical model structure which allows partial pooling among subjects, we improved predictive performance for future observations compared to existing models, especially for subjects with a small number of observations. Our approach also enables estimation of between-subject heterogeneity by contrast with ad-hoc approaches such as the clustering analysis on the subject-specific VAR coefficient estimates. We apply the model to two diary datasets to predict 1). substance use craving, negative affect and tobacco use among college students and 2). functional somatic symptoms (FSS) and psychological discomforts among subjects who had persistent FSS's.

In Chapter 3, we propose a novel Bayesian model for repeat Electrical Stimulation Mappings (ESM) to quantify the change in localization of the brain regions responsible for language function over a long period of time. The major challenges in repeat ESM's are 1). the discrepancy in number of sites mapped across different occasions leading to missing observations in the dataset and 2). unknown matching of sites between the mappings. We adopt an augmented Markov Chain Monte Carlo (MCMC) posterior sampling procedure to quantify the change of locations in these sites over time, impute unobserved sites and provide a probabilistic solution for matching the sites across occasions. We use this chapter to illustrate the strength of the Bayesian methods in incorporating different data needs and recover parameters that are generally difficult to extract by conventional methods.

Due to the aforementioned advantages, Bayesian models have been applied to a wide range of healthcare data analysis. Nevertheless, application of these modeling strategies to real world data requires great cautions in terms of what data should be used and what interpretations can be made of the results. In Chapter 4, we provide a note to demonstrate potential pitfalls in applying Bayesian methods to real data problems, in particular, the application to large-scale spontaneous reporting system (SRS) databases for post-market drug safety surveillance. In such data mining problems, Bayesian disproportionality analyses emerged to identify signals

of safety concerns of a target drug by comparing it to the background reporting experience across all other drugs in the database. Head-to-head comparisons of the target drug to a single reference drug are also not uncommon. Using empirical approach (i.e., by comparing the two backgrounds for all drug-adverse event-combinations in the Food and Drug Administration SRS database), we find that the latter comparison using a single drug as background generally yields smaller disproportionality metrics, masking the true safety signaling. We therefore use this chapter to emphasize the awareness of appropriate usage of Bayesian methods and interpretation of the results in order to avoid unwarranted conclusions and decision making in practice.

Chapter 1

Modeling functional connectivity: a multi-subject multi-session Bayesian hierarchical sparse VAR model

1.1 Introduction

Functional magnetic resonance imaging (fMRI) (Ogawa et al. 1990) is a non-invasive imaging technology widely used by neuroscientists to identify brain regions responsible for specific functions (Friston 1994, Cohen and Bookheimer 1994, Downing et al. 2001, Beauchamp et al. 2010) and investigate the coordination, called *connectivity*, among distinct brain regions (He et al. 2003, Sporns et al. 2003, Buckner et al. 2008).

Task related connectivity studies, in which subjects are exposed to one or more experimental stimuli, have been used by neuroscientists to investigate, for example, interactions among brain regions in the processing of visual information (Greicius et al. 2004, Di et al. 2015, Goghari and Lawlor-Savage 2017), verbal information (Senden et al. 2017, Kepinska et al. 2017), and decision making (Kohno et al. 2015). Studies of resting state connectivity, in which subjects are scanned while awake, but not explicitly tasked, are also of great interest, in some measure due to the discovery that the amount of energy consumed by the brain on task is only minimally greater than that consumed at rest (Raichle and Mintun 2006). These studies (for example, Biswal et al. 1995, Gusnard and Raichle 2001, Fox et al. 2005, Buckner and Vicent 2007, Thomas Yeo et al. 2011) have identified various combinations of regions, called resting state networks (RSN), consistently activated while subjects are at rest. Resting state connectivity studies are also of clinical significance. Comparisons of fMRI scans of healthy control subjects with subjects suffering from various diseases or disorders have revealed group differences in connectivity (Sporns et al. 2003, Buckholtz and Meyer-Lindenberg 2012, Pievani et al. 2014, Iturria-Medina and Evans 2015), leading to the development of therapeutic interventions for diseases such as Alzheimer (Petrella et al. 2011), autism (Kennedy et al. 2006) and ADHD (Tian et al. 2006).

Various procedures have been proposed to study functional connectivity using multivariate fMRI time series data (for reviews, see Margulies et al. 2010, Lee et al. 2013, Cribben and Fiecas 2016), in which the blood oxygenation level dependent (BOLD) responses of each

subject are recorded at V = thousands of spatial locations (voxels) and T = several hundred or more time points separated by one or two seconds. In "static" studies, connectivity measures for single subjects are typically computed first. Group level connectivity is then assessed by averaging the single subject measures over sub-groups of subjects, e.g., patients or healthy control subjects.

The contemporaneous association between the BOLD responses in different voxels (or the average of the BOLD responses over regions of interest (ROI) composed of aggregates of voxels) is typically used to measure single subject connectivity. Many studies have used the cross-correlation. This is the Pearson correlation coefficient (or sometimes its Fisher z-transformation) between voxel (or ROI) pairs. A predefined threshold value is then typically used to categorize pairs as correlated or uncorrelated (Biswal et al. 1995, Cordes et al. 2000, Buckner et al. 2008). Other studies have used the partial cross-correlation (typically obtained from the sample precision matrix), the correlation between voxel (or ROI) pairs controlling for the association due to all other voxels (or ROI's) (Zhen et al. 2007, Dawson et al. 2016). Cross-coherence and partial cross-coherence are, respectively, analogues of cross-correlation and partial cross-correlation in the frequency domain. These measures have also been used in a number of studies (Sun et al. 2004, Salvador et al. 2005, Liu et al. 2010, Peng et al. 2016).

Functional connectivity has also been studied using principal component analysis (PCA) and independent component analysis (ICA), often in tandem, and typically applied to voxel-level data (Kiviniemi et al. 2003, Beckmann et al. 2005, Damoiseaux et al. 2006). First, the covariance (or correlation) matrix Σ of the $NT \times V$ data matrix, where N is the number of subjects, is spectrally decomposed ($\Sigma = \sum_{i=1}^{V} \lambda_i \alpha_i \alpha_i'$, where $\lambda_i, i = 1, \ldots, V$ are eigenvalues and $\sum_{i=1}^{V} \alpha_i' \alpha_i = 1$), and C uncorrelated components (eigenvectors) corresponding to the C largest eigenvalues are chosen; typically C is much smaller than V.

ICA is then applied to factorize the data matrix as the product of a $NT \times C$ matrix L with a $C \times V$ matrix Z. Typically, a pre-defined threshold is applied to each row of Z to yield

7

independent clusters of "connected" voxels.

The preceding methods are easy to compute and in resting state studies, have consistently identified RSN's such as the motor, auditory, visual, sensory, cerebellum and default mode networks (Biswal et al. 1995, Damoiseaux et al. 2006, Buckner et al. 2008). Cross-correlation and cross-coherence are typically visualized using heat maps that display the significance (i.e., t-ratios or p-values) of the coefficients. But heat maps do not indicate the magnitude (strength) of the association between voxels. In studies using partial cross-correlation and partial cross-coherence, the measures are typically extracted from the sample precision matrix, which may be poorly estimated when V is much larger than T, as is commonly the case in fMRI. While PCA and ICA are useful exploratory procedures to discover connected ROI's when the ROI's are not defined a priori, the number of components selected is arbitrary and constraining the components to be uncorrelated may not be substantively reasonable.

Even more importantly, the procedures above have not been used to address asynchronous connectivity, and it would be awkward to use these to address both contemporaneous and asynchronous connectivity in a systematic way. Therefore, neuroscientists have turned to the use of vector autoregressive (VAR) models in conjunction with Granger causality (Granger 1969, Roebroeck et al. 2005, Webb et al. 2013, D'Souza et al. 2017). While these models are sometimes used to make inferences about effective connectivity, that is, about the effects of neural activity in one or more brain locations on the activity in other regions (Friston 1994), temporal precedence is not in and of itself a sufficient basis for inferring causation, and as Lindquist and Sobel (2016) have demonstrated, the common interpretation of non-zero coefficients in VAR models as effects is generally unwarranted. Although this does not diminish the utility of these models for studying functional connectivity, the models currently used by neuroscientists are limited in a number of significant ways.

Previous researchers have used VAR models to analyze single subject single session or multi-subject single session data (Deshpande et al. 2009, Schippers et al. 2011, Webb et al.

8

2013, Gorrostieta et al. 2013, Chiang et al. 2016). To analyze multi-subject single session data, a two step procedure is often used, in which VAR models are first applied to single subject data to obtain subject-specific estimates; these estimates are then used as the basis for group level inference. As an example, Schippers et al. (2011) are interested in whether two regions, say A and B, have equal "influence" on one another. They use single subject estimates to define the A-to-B "influence" and the B-to-A "influence" and they construct a t-test for the null hypothesis that the mean A-to-B "influence" is equal to the mean B-to-A "influence". Although the two step procedure is convenient for computation, because it does not allow information to be pooled over subjects, the subject level estimates may be highly variable. Moreover, no insight into between subject heterogeneity is obtained.

Several authors have proposed Bayesian hierarchical VAR models for multi-subject single session data (Gorrostieta et al. 2013, Chiang et al. 2016). In contrast to the two step procedure, which allows no pooling, these models allow partial pooling, resulting in more accurate estimates of subject and group level connectivity. Estimated variances of the VAR coefficients, which suggest substantial between subject heterogeneity, are obtained by sampling from the posterior distribution of the variance parameters.

However, these models are also limited in several important ways. First, they can only handle a small number of ROI's. This forces the investigator to either focus on a few ROI's within a network or a handful of ROI's spanning several networks. For example, Gorrostieta et al. (2013) studied five regions within the motor network and Chiang et al. (2016) investigated six regions within several resting state networks. Second, to impose sparsity, existing models assume the error covariance matrix is diagonal (Li et al. 2010, Gorrostieta et al. 2013, Chiang et al. 2016). These limitations can lead an investigator to make inferences that are incomplete and/or unwarranted.

Finally, and equally important, as evidenced by data collection efforts such as the Human Connectome Project (`www.humanconectome.org`), with increasing regularity, neuroscien-

tists are collecting data on many subjects, with each subject observed on several occasions. With such data, one can ask whether the variability in brain activity is due primarily to differences between subjects or differences within subjects across occasions. Previous studies have proposed a two stage procedure (similar to the aforementioned two step procedure) for general connectivity metrics such as cross-correlations and ICA (Wang et al. 2011, Zuo and Xing 2014, Ge et al. 2017). In this procedure, subject-session connectivity estimates are first obtained; individual mixed effect models are then applied on each component in the connectivity vector. While computationally simple, this approach uses the summary statistics obtained in the first stage, rather than the original data, to estimate the parameters in the second stage.

We propose a sparse Bayesian hierarchical VAR model that extends existing multi-subject Bayesian VAR models to the multi-subject multi-session case. Our model also addresses the limitations described above. Specifically, our model allows simultaneous inference about group level, subject level and subject-session level resting state functional connectivity, as well as between subject and between session variabilities. By adopting the doubly adaptive Elastic-net prior (Gefang 2014), our model induces sparsity in parameter estimation without assuming any special structure for the error covariance matrix. Finally, using parameter-expanded Gibbs sampler (Higdon 1998), we can study a larger number of connected ROI's than previous researchers could investigate. We apply our model to data from the Human Connectome Project to investigate resting state functional connectivity between 15 ROI's across 7 brain networks. Our unique modeling approach decomposes functional connectivity into group, subject, and session components, letting us investigate whether connectivity is consistent across subjects and replications, an important issue that has not yet been extensively explored. We find distinct interregional connections both within and between these brain networks and greater heterogeneity between subjects than within subjects over sessions.

We proceed as follows. Section 1.2 describes the proposed Bayesian model. Section

1.3 describes a simulation study to investigate the performance of the model under different scenarios and compares it to the two stage procedure described above. Section 1.4 illustrates the model with a real example from the Human Connectome Project. Section 1.5 concludes.

1.2 Method

We extend the work of Gefang (2014), who constructed a Bayesian VAR model using a doubly adaptive Elastic-net prior to induce sparsity, to the multi-subject multi-session case, allowing inference about group level, subject level and subject-session level connectivity, and between and within subject variabilities. We use a Gibbs sampler with parameter expansion (Higdon 1998, Gelman et al. 2014) to reduce correlations among parameters and speed up convergence, allowing us to simultaneously study more ROI's than in previous research using VAR models.

1.2.1 A Bayesian hierarchical sparse VAR model

For subject $i \in \{1, ...N\}$ during period $t \in \{1, ..., T\}$ of session $s \in \{1, ..., S\}$, let y_{istr} denote the average value (taken over the voxels in the ROI) of the BOLD responses in ROI $r \in \{1, ..., R\}$, and let $\boldsymbol{y}_{ist.} = (y_{ist1}, ..., y_{istR})^T$. For $t = p + 1, ..., T$, assume

$$\boldsymbol{y}_{ist.} = \sum_{l=1}^{p} A_{isl} \boldsymbol{y}_{is,t-l,.} + \boldsymbol{\epsilon}_{ist.}, \tag{1.1}$$

where A_{isl}, $l = 1, ..., p$ is a $R \times R$ coefficient matrix representing the lag-l association among the BOLD responses for subject i in session s, and $\boldsymbol{\epsilon}_{ist.}$, $t = p + 1, ..., T$, are independent and identically distributed (i.i.d.) multivariate normal with mean $\boldsymbol{0}$ and precision matrix Ω: $\boldsymbol{\epsilon}_{ist.}, t = p + 1, ..., T \sim i.i.d. \operatorname{MVN}(\boldsymbol{0}, \Omega^{-1})$. For subject i in session s, let

$$\boldsymbol{y}_{is..} = \begin{pmatrix} \boldsymbol{y}_{is,p+1.} \\ \vdots \\ \boldsymbol{y}_{isT.} \end{pmatrix},$$

11

and rewrite the model as:

$$y_{is..} = \left(H_{is}{}^T \otimes I_R\right) w_{is} + \epsilon_{is..}, \tag{1.2}$$

where $\otimes$ denotes the Kronecher product, I_R is an identity matrix of order R,

$$H_{is} = \begin{pmatrix} y_{isp.} & \cdots & y_{is,T-1,.} \\ \vdots & \ddots & \vdots \\ y_{is1.} & \cdots & y_{is,T-p,.} \end{pmatrix},$$

$$w_{is} = vec([A_{is1}, ..., A_{isp}])$$

is of length $R^2 p$, and

$$\epsilon_{is..} = \begin{pmatrix} \epsilon_{is,p+1,.} \\ \vdots \\ \epsilon_{isT.} \end{pmatrix}.$$

The coefficient vector w_{is} for the i-th subject and the s-th session is composed of three parts:

$$w_{is} = w + v_i + u_{is}, \tag{1.3}$$

where w is the group level coefficient vector, v_i is the subject deviation vector and u_{is} is the subject-session deviation vector.

For the group level coefficient vector w, we use the doubly adaptive Elastic-net prior proposed by Gefang (2014),

$$w|D \sim \text{MVN}\left(0, D^{-1}\right), \tag{1.4}$$

where the precision matrix D is diagonal, depending on hyperparameters $2\tau_k^2$ and $\lambda_{2,k} > 0, k = 1, ..., R^2 p$,

$$\text{diag}(D) = \left(\lambda_{2,1} + \frac{1}{2\tau_1^2}, ..., \lambda_{2,R^2 p} + \frac{1}{2\tau_{R^2 p}^2}\right). \tag{1.5}$$

The hyperparameters $2\tau_k^2, k = 1, ..., R^2 p$, are assumed to follow independent exponential distributions with rates $\frac{\lambda_{1,k}^2}{2\xi_k^2}, k = 1, ..., R^2 p$:

$$2\tau_k^2 | \xi_k^2, \lambda_{1,k} \sim E\left(\frac{\lambda_{1,k}^2}{2\xi_k^2}\right), \tag{1.6}$$

12

where $\xi_k^2, k = 1, ..., R^2p$, depends on the error precision matrix Ω:

$$\xi_k^2 = M_{k,k} - \left(M_{k,k+1}, ..., M_{k,R^2p} \right) \begin{pmatrix} M_{k+1,k+1} & \cdots & M_{k+1,R^2p} \\ \vdots & \ddots & \vdots \\ M_{R^2p,k+1} & \cdots & M_{R^2p,R^2p} \end{pmatrix}^{-1} \begin{pmatrix} M_{k,k+1} \\ \vdots \\ M_{k,R^2p} \end{pmatrix}, \quad (1.7)$$

$$M = I_{Rp} \otimes \Omega^{-1}. \quad (1.8)$$

The collection of hyperparameters $\lambda_{1,k}^2$ and $\lambda_{2,k}, k = 1, ..., R^2p$, are assumed to follow two independent gamma distributions with known parameters:

$$\lambda_{1,k}^2 \sim \Gamma(\mu_1, \nu_1), \quad (1.9)$$

$$\lambda_{2,k} \sim \Gamma(\mu_2, \nu_2), \quad (1.10)$$

where the density function is written as $f(x; \mu, \nu) = \frac{1}{\Gamma(\frac{\nu}{2})} \left(\frac{2\mu}{\nu} \right)^{-\frac{\nu}{\mu}} x^{\frac{\nu}{2}-1} e^{-\frac{\nu}{2\mu}x}, x > 0, \mu > 0,$ $\nu > 0$, where μ and ν are, respectively, the mean and degrees of freedom (df). Here, $\lambda_{1,k}$ and $\lambda_{2,k}$ represent, respectively, the L1 and L2 tuning parameters for the k-th component of w. It can be shown that the conditional prior for each of the components of w given the error precision matrix Ω is composed of a normal distribution and a Laplace distribution with individual shrinkage parameters, that is, this prior is equivalent to the Elastic-net regularization with component-wise shrinkage. Note that conditioning on the error precision matrix Ω is important as it ensures the posterior distribution for each of the components of w is unimodal (Park and Casella 2008).

For the subject and subject-session deviations, we apply conjugate i.i.d. priors with mean 0 and diagonal precision matrices Θ_v and Θ_u:

$$v_i|\Theta_v \sim \text{MVN} \left(0, \Theta_v^{-1} \right), i = 1, ..., N, \quad (1.11)$$

$$u_{is}|\Theta_u \sim \text{MVN} \left(0, \Theta_u^{-1} \right), i = 1, ..., N, s = 1, ..., S, \quad (1.12)$$

where the diagonal elements of Θ_v and Θ_u, respectively, are θ_{vk} and $\theta_{uk}, k = 1, ..., R^2p$, assumed to follow independent gamma distributions:

$$\theta_{vk} \sim \Gamma(k_1, s_1), k = 1, ..., R^2p, \quad (1.13)$$

13

$$\theta_{uk} \sim \Gamma(k_2, s_2), k = 1, ..., R^2 p, \tag{1.14}$$

where k_1, s_1 and k_2, s_2 are the known mean and df in (1.13) and (1.14), respectively. We assume the group level coefficients, subject level coefficients and the subject-session level coefficients are independent of each other given the hyperparameters.

Finally, for the error precision matrix Ω we assume a conjugate Wishart hyperprior:

$$\Omega \sim \text{Wishart}(K, \nu), \tag{1.15}$$

where K and ν are the known $R \times R$ scale matrix and degrees of freedom (i.e., the mean and variance of Ω increases as ν increases).

1.2.2   Posterior sampling and posterior inference

We use Markov Chain Monte Carlo (MCMC) to generate draws from the posterior distribution, specifically, a Gibbs sampler with parameter expansion (Gelman et al. 2014). For hierarchical regression models the sampler can be very slow if there is strong dependence between the coefficients and their variance parameters. In our model, $\boldsymbol{v}_i$ and Θ_v (and $\boldsymbol{u}_{is}$ and Θ_u) are dependent: when the current draw of Θ_v is large, the next draw of $\boldsymbol{v}_i$ will be small, making the next draw of Θ_v even larger, and so on. Thus the sampler can take a long time to cover the entire parameter space, resulting in slow convergence.

Using parameter expansion allows us to reduce the dependence among parameters and speed up convergence. Specifically, we add component-wise multiplicative factors $\boldsymbol{\alpha}$ to $\boldsymbol{v}_i$ and $\boldsymbol{\beta}$ to $\boldsymbol{u}_{is}$ (i.e., multiply α_k with v_{ik} and β_k with u_{isk}, $k = 1, ..., R^2 p$), where

$$\boldsymbol{\alpha} \sim \text{MVN}(\boldsymbol{0}, a I_{R^2 p}), \tag{1.16}$$

$$\boldsymbol{\beta} \sim \text{MVN}(\boldsymbol{0}, b I_{R^2 p}), \tag{1.17}$$

and a and b are known constants. When the current draw of Θ_v is large, the next draw of $\boldsymbol{v}_i$ will be small, but the next draw of $\boldsymbol{\alpha}$ will be large (see the conditional distribution of $\boldsymbol{\alpha}$ in

Appendix A), resulting in the next draw of v_i in a normal range (similarly for u_{is} and Θ_u). That is, the expanded parameterization allows the Gibbs sampler to move in more directions and avoid getting stuck.

The subject and subject-session deviation parameters are now

$$v_i^* = \alpha * v_i, \tag{1.18}$$

$$u_{is}^* = \beta * u_{is}, \tag{1.19}$$

and the diagonal elements of the precision matrices are

$$\theta_{vk}^* = \alpha_k^{-2}\theta_{vk}, k = 1, ..., R^2p, \tag{1.20}$$

$$\theta_{uk}^* = \beta_k^{-2}\theta_{uk}, k = 1, ..., R^2p. \tag{1.21}$$

In addition to parameter expansion, to speed up convergence, we use Method-of-Moments-like estimates as initial values for the posterior sampling. After obtaining the posterior samples, we use the posterior modes and posterior medians, respectively, to estimate the group level coefficients and the other parameters. Using posterior modes for the group level parameters induces sparsity and gives estimates equal to those obtained using penalized likelihood methods such as the Lasso (Genkin et al. 2007, Park and Casella 2008). Using posterior medians for other parameters increases robustness. We use the values with the highest empirical density to obtain the posterior modes. Finally, we use the AIC as a criterion to select the optimal order of the VAR model. See the Appendix A for the derivation of the initial values and the full conditionals used in the Gibbs sampler.

1.3   Simulation Studies

1.3.1   Simulation studies of the Bayesian model

We conduct a simulation study to investigate the behavior of our model. We vary 3 factors: 1) sample size (N = 5, 50, 100, 500), 2) the ratio of the standard deviations of the subject

deviations and the subject-session deviations (f=.5, 1, 2), 3) the magnitude and structure of the error covariance matrix, including high and low noise levels and block diagonal and diagonal structures. We consider 4 cases, all with $T = 200$ time points, $R = 5$ regions, $S = 4$ sessions, and $p = 2$ lags. We assume normal subject-session specific VAR models with coefficients $w_{is} = w + v_i + u_{is}, i = 1, \ldots, N, s = 1, \ldots, S$, and 70% of the group level coefficients w are set to zero, with the remainder drawn from the $N(0.112, 0.09^2)$ distribution.

1. Cases 1a-1d; $f = 1$, $N = 5$ (1a), 50 (1b), 100 (1c), 500 (1d). The subject deviations in v_i and the subject-session deviations in u_{is} are each drawn from the $N(0, 0.067^2), i = 1, \ldots, N, s = 1, \ldots, S$. The error covariance matrix Ω^{-1} is drawn from a Wishart distribution with 10 degrees of freedom and a symmetric Toeplitz scale matrix with diagonal elements equal to 0.2.

2. Cases 2a-2d; f=2, $N = 5$ (2a), 50 (2b), 100 (2c), 500 (2d). v_i and u_{is} are, respectively, drawn from the $N(0, 0.09^2)$ and the $N(0, 0.045^2)$ distributions, $i = 1, \ldots, N, s = 1, \ldots, S$. Ω^{-1} is as in case 1.

3. Case 3a-3d; f=0.5, $N = 5$ (3a), 50 (3b), 100 (3c), 500 (3d). v_i and u_{is} are, respectively, drawn from the $N(0, 0.045^2)$ and the $N(0, 0.09^2)$ distributions, $i = 1, \ldots, N, s = 1, \ldots, S$. Ω^{-1} is as in case 1.

4. Cases 4a-4d; f=1, $N = 5$. In cases 4a and 4b, the error covariance matrices are drawn from Wishart distributions with 10 degrees of freedom and Toeplitz scale matrices with diagonal elements equal to 0.3 (4a) and 0.1 (4b). In case 4c, the error covariance matrix is block diagonal and in case 4d diagonal, with elements equal to 2; for both these cases, v_i and $u_{is}, i = 1, \ldots, N, s = 1, \ldots, S$, are as in case 1.

For each case, we generated 100 repetitions using the R package 'sparsevar' (`https://cran.r-project.org/web/packages/sparsevar/index.html`), and we fit our VAR model with orders of p=1, 2 and 3 to the simulated data. We use p=2 to investigate the recovery of the parameters and p=1 and 3 to investigate whether the AIC selects the correct order. The

entire simulation procedure was implemented in R. The computations were performed using the Gibbs sampler strategy with expanded parameterization. For each repetition, we ran 30,000 iterations across 4 chains, using the first 15,000 iterations as warm-up; the next 15,000 iterations were thinned, keeping every 60-th iteration, yielding 250 posterior samples per chain (for cases with N=500, we ran 40,000 iterations and used the last 20,000 iterations, keeping every 80-th iteration, to ensure mixing of the chains). All chains mixed well. For details about the generation of the error covariance matrices, see the Appendix A.

1.3.1.1 Parameter recovery

For each component of the group level connectivity parameters ($\boldsymbol{w}$), the subject deviations ($\boldsymbol{v} = \{\boldsymbol{v}_i, i = 1, \ldots, N\}$), the subject-session deviations ($\boldsymbol{u} = \{\boldsymbol{u}_{is}, i = 1, \ldots, N, s = 1, \ldots, S\}$), the standard deviation of the subject deviations ($\mathbf{sdv} = \{\theta_{vk}^{*}{}^{-\frac{1}{2}}, k = 1, \ldots, R^2 p\}$) and the standard deviation of the subject-session deviations ($\mathbf{sdv} = \{\theta_{uk}^{*}{}^{-\frac{1}{2}}, k = 1, \ldots, R^2 p\}$), we estimate the mean square error (MSE) using the 100 repetitions. We then average the estimates over the components of $\boldsymbol{w}$, $\boldsymbol{v}$, $\boldsymbol{u}$, $\mathbf{sdv}$ and $\mathbf{sdu}$, respectively, obtaining estimated average MSE's (AMSE) for $\boldsymbol{w}$, $\boldsymbol{v}$, $\boldsymbol{u}$, $\mathbf{sdv}$ and $\mathbf{sdu}$ (the number of components for $\boldsymbol{w}$, $\mathbf{sdv}$ and $\mathbf{sdu}$ is 50 and for $\boldsymbol{v}$ and $\boldsymbol{u}$ are 50N and 200N, respectively). Figure 1.1 displays the MSE's under all cases, with each boxplot representing the distribution of the MSE's for all components in $\boldsymbol{w}$, $\boldsymbol{v}$, $\boldsymbol{u}$, $\mathbf{sdv}$ and $\mathbf{sdu}$ under a particular case. Columns 1-5 in Table 1.1 report the estimated AMSE's for all cases using our model.

Inspection of Figure 1.1 and Table 1.1 indicates that as N increases, the MSE's of the components of $\boldsymbol{w}$, $\mathbf{sdv}$ and $\mathbf{sdu}$ become close to 0 in all cases, demonstrating that our model recovers these parameters well: this is due to the large sample consistency of the Bayesian estimates (Gelman et al. 2014). The number of parameters in $\boldsymbol{v}$ and $\boldsymbol{u}$ increases with N, so the property of large sample consistency does not hold for $\boldsymbol{v}$ and $\boldsymbol{u}$. Nevertheless, the estimated AMSE's for these parameters also decreases as N increases. Further, results for case 4 are similar to those above, suggesting the choice of error covariance matrix is not a key

17

Table 1.1: Estimated AMSE's for the group level connectivity parameters (***w***), the subject deviations (***v***), the subject-session deviations (***u***), the standard deviation of the subject deviations (**sdv**) and the standard deviation of the subject-session deviations (**sdu**) for all cases using the Bayesian model (columns 1-5) and the two stage procedure (columns 6-10). Columns 11-15 list the reduction in the estimated AMSE's for the Bayesian model relative to the two stage procedure.

Case	Bayesian model					Two stage procedure					Reduction in AMSE				
	w	***v***	***u***	**sdv**	**sdu**	***w***	***v***	***u***	**sdv**	**sdu**	***w***	***v***	***u***	**sdv**	**sdu**
1a	1.08E-3	2.99E-3	3.33E-3	1.55E-3	1.00E-3	1.71E-3	3.12E-3	8.67E-3	1.79E-3	2.78E-3	36%	4%	62%	14%	64%
1b	9.32E-5	1.88E-3	2.76E-3	1.56E-4	8.64E-5	1.38E-4	2.10E-3	8.81E-3	1.71E-4	2.71E-3	32%	10%	69%	9%	97%
1c	6.93E-5	1.77E-3	2.70E-3	7.08E-5	3.78E-5	9.12E-5	2.01E-3	8.82E-3	8.64E-5	2.73E-3	24%	12%	69%	18%	99%
1d	1.42E-5	1.72E-3	2.63E-3	1.36E-5	6.85E-6	1.89E-5	1.94E-3	8.63E-3	1.60E-5	2.65E-3	25%	12%	70%	15%	100%
2a	1.59E-3	3.59E-3	1.79E-3	2.16E-3	6.00E-4	2.43E-3	4.01E-3	8.02E-3	1.99E-3	3.93E-3	34%	10%	78%	-8%	85%
2b	1.22E-4	1.85E-3	1.53E-3	1.26E-4	1.10E-4	1.83E-4	2.19E-3	7.97E-3	1.49E-4	3.87E-3	33%	16%	81%	16%	97%
2c	9.30E-5	1.79E-3	1.49E-3	6.28E-5	4.74E-5	1.20E-4	2.13E-3	8.00E-3	7.53E-5	3.91E-3	22%	16%	81%	17%	99%
2d	1.91E-5	1.74E-3	1.44E-3	1.30E-5	7.09E-6	2.49E-5	2.05E-3	7.85E-3	1.45E-5	3.81E-3	23%	15%	82%	10%	100%
3a	7.27E-4	1.89E-3	4.54E-3	7.94E-4	1.23E-3	1.25E-3	2.09E-3	8.71E-3	1.37E-3	1.98E-3	42%	10%	48%	42%	38%
3b	7.65E-5	1.54E-3	3.78E-3	3.33E-4	8.59E-5	1.11E-4	1.55E-3	9.19E-3	3.37E-4	1.85E-3	31%	1%	59%	1%	95%
3c	5.55E-5	1.42E-3	3.72E-3	1.98E-4	4.55E-5	7.47E-5	1.44E-3	9.22E-3	1.77E-4	1.88E-3	26%	2%	60%	-12%	98%
3d	1.08E-5	1.03E-3	3.60E-3	2.30E-5	7.62E-6	1.52E-5	1.38E-3	9.07E-3	2.67E-5	1.83E-3	29%	6%	60%	14%	100%
4a	1.13E-3	3.04E-3	3.36E-3	1.58E-3	1.04E-3	1.73E-3	3.18E-3	9.20E-3	1.81E-3	3.07E-3	35%	4%	63%	13%	66%
4b	1.08E-3	2.98E-3	3.35E-3	1.53E-3	1.03E-3	1.67E-3	3.09E-3	8.31E-3	1.75E-3	2.60E-3	35%	4%	60%	13%	61%
4c	9.69E-4	2.78E-3	3.38E-3	1.47E-3	1.04E-3	1.52E-3	2.76E-3	5.50E-3	1.51E-3	1.28E-3	36%	-1%	38%	3%	19%
4d	9.69E-4	2.72E-3	3.37E-3	1.44E-3	9.89E-4	1.51E-3	2.73E-3	5.30E-3	1.48E-3	1.18E-3	36%	0%	36%	2%	16%
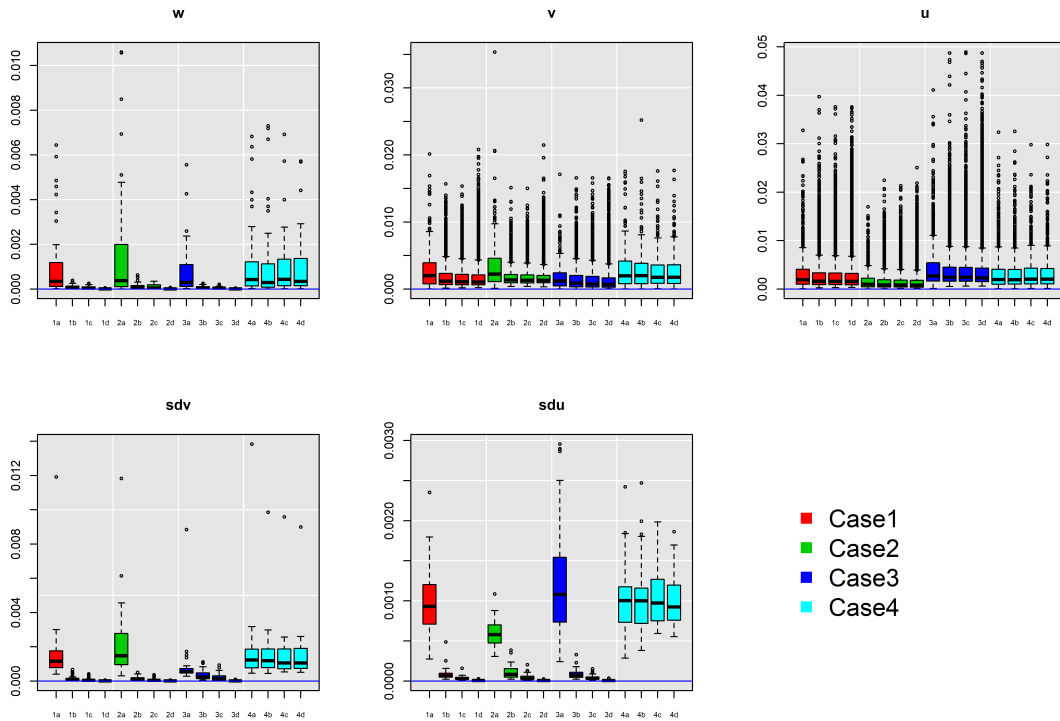
Figure 1.1: MSE's for the components of the group level connectivity parameters (*w*), the subject deviations (*v*), the subject-session deviations (*u*), the standard deviation of the subject deviations (**sdv**) and the standard deviation of the subject-session deviations (**sdu**) for all cases using the Bayesian model. Each boxplot represents the distribution of the MSE's for all components in *w*, *v*, *u*, **sdv** and **sdu** for a particular case.

determinant of model performance.

To assess the coverage of the 95% equal-tailed Bayesian credible intervals (based on the 2.5-th and 97.5-th posterior quantiles), we use the estimated coverage probability, based on the 100 repetitions. Figure 1.2 displays the coverage probabilities for each of the components in *w*, **sdv** and **sdu** under cases 1a-1d. (Although we would have preferred to use more repetitions, limited access to computational resources prevented us from doing so. Therefore, in addition to the estimate, displayed as a black dot in Figure 1.2, we also display a blue line at the value .95 and a blue band symmetric about .95 with a length of 2 standard deviations; the standard deviation (.021) is based on 100 draws from a binomial distribution with probability .95.) Black dots within the blue band are indicative of a coverage probability close to the

nominal level. The index of the panel for w is sorted by the true values of the components of w (from smallest to largest), so the first 36 indices correspond to 0 coefficients in w. There is no sorting for **sdv** and **sdu** since the true values are the same for each of the components.
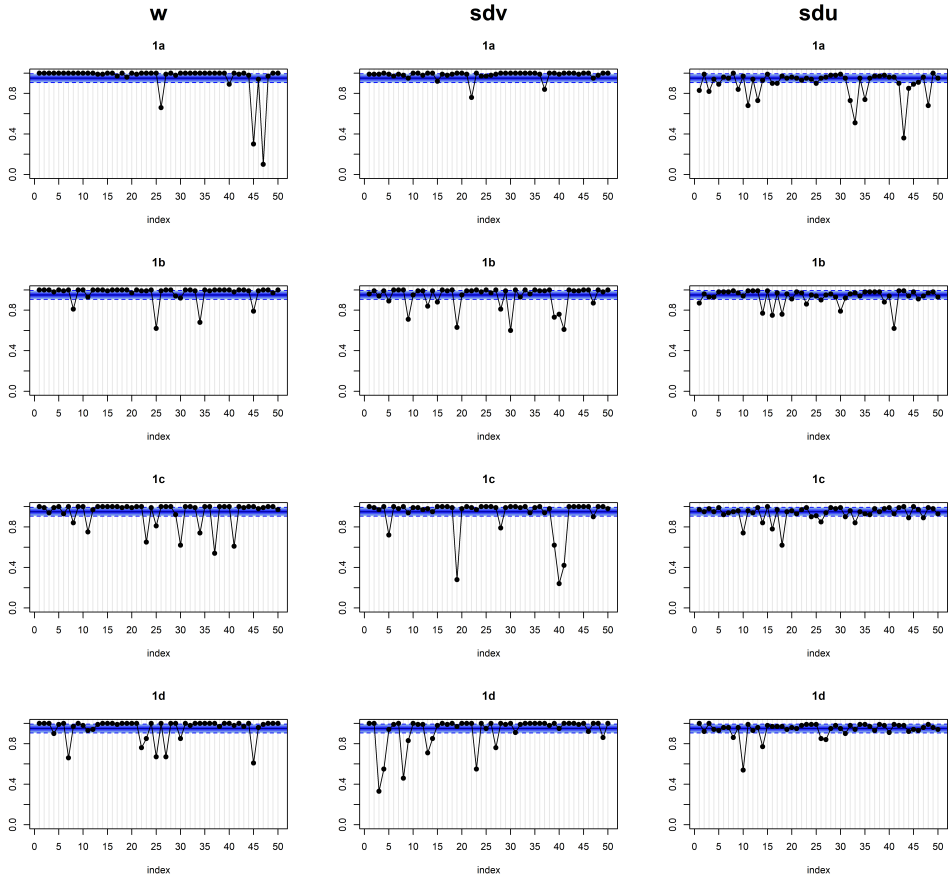
Figure 1.2: Coverage probabilities (black dots) of the 95% equal-tailed Bayesian credible intervals for the components of the group level connectivity parameters (w, column 1), the standard deviation of the subject deviations (**sdv**, column 2) and the standard deviation of the subject-session deviations (**sdu** , column 3) under cases 1a -1d (row 1-4). The index for w is sorted from smallest to largest by the true value of the components of w, so the first 36 indices correspond to 0 coefficients in w. There is no sorting for **sdv** and **sdu**.

Since 95% is the target level, we would expect about 2-3 of the black dots to fall outside of the blue bands. However, as sample size N increases, the number of overcovered parameters decreases but the number of undercovered parameters does not. We suspect this is due to two factors: 1) the tendency, reported in previous studies (Agresti and Min 2005, Zhang 2006,
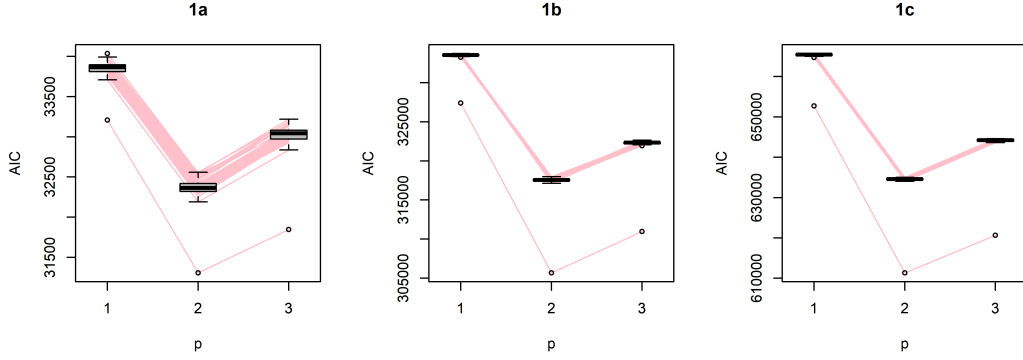
Figure 1.3: AIC's obtained by fitting VAR models with order p=1, 2 and 3 under cases 1a-1c. Each pink line in the plots represents one repetition.

Fraser 2011), for Bayesian credible intervals to undercover, and 2) the sample sizes here. To investigate this more thoroughly, we need to examine larger sample sizes, and to do so, we need to employ a more sophisticated computational strategy (such as the consensus Monte Carlo algorithm (Scott et al. 2013)).

1.3.1.2 Recovery of model order using AIC

We used AIC to select the order p for the VAR model. AIC recovered the corrected order (p=2) 100% of the time for every case. Figure 1.3 shows the boxplots of AIC's for p=1, 2 and 3 under cases 1a-1c.

1.3.2 Comparative study with a two stage procedure

Next, we compare the performance of our model with an analogous two stage procedure proposed for general functional connectivity metrics. Specifically, we first estimate individual VAR models for each subject and session to obtain subject-session specific MLE's: $\hat{w}_{is}, i = 1, \ldots, N, s = 1, \ldots, S$, and we then fit a mixed effect model for each component of $\hat{w}_{is}$:

$$\hat{w}_{isk} = w_k + v_{ik} + u_{isk}, k = 1, \ldots, R^2 p, i = 1, \ldots, N, s = 1, \ldots, S \qquad (1.22)$$

21

where w_k is the fixed effect, $v_{ik} \sim N(0, \sigma_{vk}^2)$, i.i.d., is the random effect for subject i, $u_{isk} \sim N(0, \sigma_{uk}^2)$, i.i.d, is the error term for subject i and session s, independent of v_{ik}, and σ_{vk}^2 and σ_{uk}^2 are, respectively, the between subject and between session variabilities.

1.3.2.1 Parameter recovery

Columns 6-10 in Tables 1.1 display the estimated AMSE's using the two stage method, and Columns 11-15 display the relative reduction in these for the Bayesian model, which outperforms the two stage procedure for almost all parameters across all cases. The reduction in the estimated AMSE is about 30% for the group level connectivity parameters (w), 6% for the subject deviations (v), 60% for the subject-session deviations (u), 8% for the standard deviation of the subject deviations (**sdv**) and 70% for the standard deviation of the subject-session deviations (**sdu**). For the coverage probability, neither method reaches the nominal level even for N=500 (the two stage method has an even lower coverage probability for w), but about 80% of the Bayesian credible intervals are shorter than those given by the two stage procedure. The two stage procedure does not provide range estimates for **sdv** and **sdu**.

Figures 1.4 and 1.5 display, respectively, boxplots of the estimates obtained in the 100 repetitions for each component of the group level connectivity parameters (w), the subject deviations (v) and the subject-session deviations (u) for cases 1a (N=5) and 1c (N=100) for the Bayesian and two stage methods. When N is small, the variance for the estimates of the coefficients in w is smaller under the Bayesian model, especially for coefficients whose true values are 0, due to the shrinkage imposed by the Elastic-net prior. This is also true for v and u, due to the pooling effect of the Bayesian model, which pools over subjects, sessions and components in the parameter vectors; the two stage procedure only pools over subjects and sessions. As N increases, both models recover parameters in w, as these parameters are consistently estimated under both procedures. The Bayesian model outperforms the two stage procedure for estimating v and u due to the aforementioned differences in pooling.

Figure 1.6 displays boxplots of the estimates obtained from the 100 repetitions for the

22
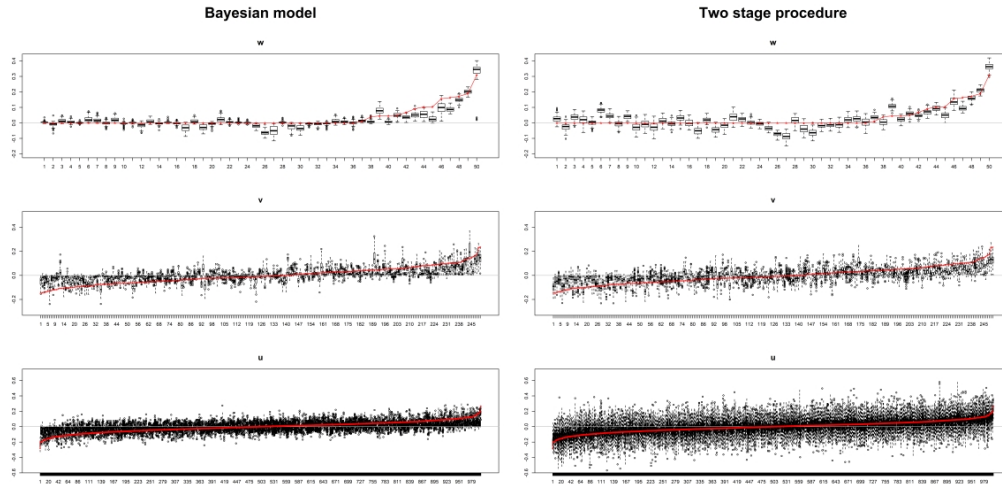
Figure 1.4: Boxplots of estimates obtained from the 100 repetitions for each component of the group level connectivity parameters (w), the subject deviations (v) and the subject-session deviations (u) using the Bayesian model (left column) and the two stage procedure (right column) under case 1a (N=5). The red dots represent the true values of the parameters. The boxplots are sorted by the true values for better visualization.
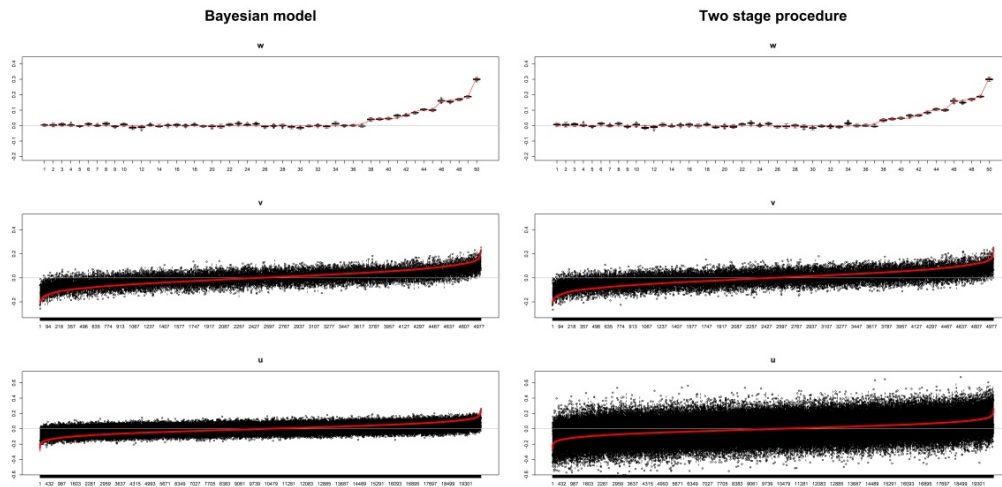
Figure 1.5: Boxplots of estimates obtained from the 100 repetitions for each component of the group level connectivity parameters (w), the subject deviations (v) and the subject-session deviations (u) using the Bayesian model (left column) and the two stage procedure (right column) under case 1c (N=100). The red dots represent the true values of the parameters. The boxplots are sorted by the true values for better visualization.
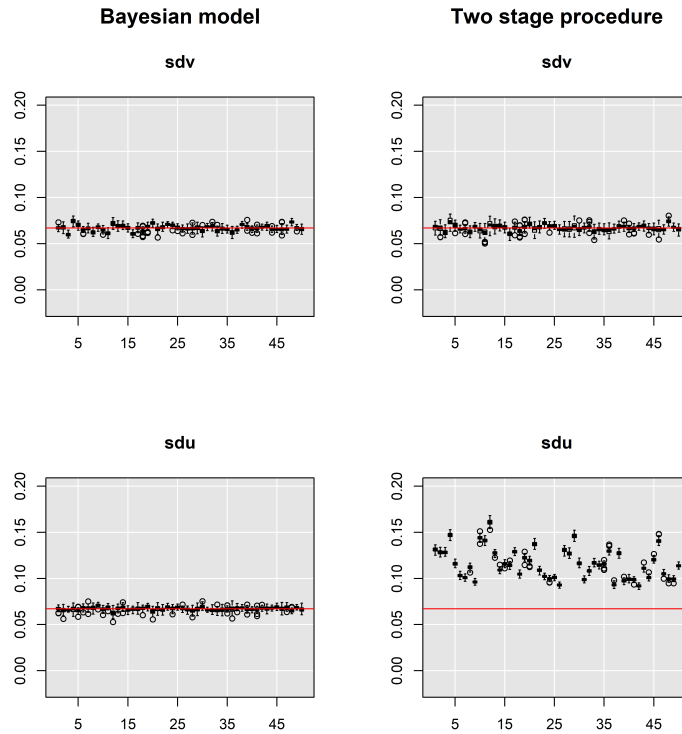
Figure 1.6: Boxplots of estimates from the 100 repetitions for each component of the standard deviation of the subject deviations (**sdv**, row 1) and the standard deviation of the subject-session deviations (**sdu**, row 2) using the Bayesian model (column 1) and the two stage procedure (column 2) under case 1d (*N*=500). The red lines display the true values.

components of **sdv** and **sdu** under case 1d (*N*=500). With larger *N*, the Bayesian estimates are close to the true values for these parameters, whereas the two stage procedure exhibits a positive bias for **sdu**, due to the fact that while the Bayesian model uses the original data to estimate these parameters, the two stage procedure uses the summary MLE's obtained in the first stage to estimate these parameters in the second stage.

1.3.2.2 Recovery of sparsity in group level connectivity

We also compare the two methods in terms of the recovery of zero coefficients in the group level connectivity parameters (w), assessed by the area under the precision-recall curve (AUC-

PRC) (He and Garcia 2009, Powers 2011, Saito and Rehmsmeier 2015), where *precision* = TP/(TP+FP), *recall*= TP/(TP+FN), and TP, FP and FN represent, respectively, true positives, false positives and false negatives. We define the zero coefficients as positive, non-zero coefficients as negative instances, and thus TP = the number of zero coefficients that are recovered as 0, FP = the number of non-zero coefficients that are recovered as 0, and FN = the number of zero coefficients that are recovered as non-zero. Therefore, in our case, precision measures how many of the coefficients recovered as 0 are truly zero, and recall measures how many of the zero coefficients are recovered as 0. A high value of the AUC- PRC indicates good recovery of sparsity in the coefficients.

To obtain the AUC-PRC, we conduct an α-level hypothesis test (H$_0$: $w_k = 0$ vs H$_1$: $w_k \neq 0$) for each component of the group level coefficient $w_k, k = 1, \ldots, R^2p$, by looking at whether the $(1 - \alpha) \times 100\%$ confidence interval contains 0 or not. For each value of α, this gives a pair of precision-recall values, and varying the value of α gives the AUC-PRC. Table 1.2 displays the averaged AUC-PRC over the 100 repetitions for the two methods. We can see that the Bayesian model outperformed the two stage procedure in most cases, demonstrating the advantage of the Elastic-net prior in recovering sparsity in the group level coefficients.

1.4   Application: the Human Connectome Project

1.4.1   Data and the preprocessing

We apply our model to the preprocessed and artifact-removed resting state fMRI data from 50 participants as provided by the HCP820-PTN data release of the Human Connectome Project (`http://www.humanconnectomeproject.org`). The data were acquired on a customized 3T Siemens connectome-Skyra 3T scanner. Participants completed two scanning sessions on two separate days. A T1w MPRAGE structural run was acquired on each day (acquisition time =7.6 min, TR/TE/TI = 2400/2.14/1000 ms, resolution = 0.7×0.7×0.7 mm3, SENSE factor = 2,

Table 1.2: Average area under the precision-recall curve (AUC-PRC) over the 100 repetitions.

Case	Bayesian model	Two stage procedure
1a	0.905	0.898
1b	0.991	0.988
1c	0.963	0.960
1d	0.999	0.997
2a	0.872	0.884
2b	0.986	0.984
2c	0.967	0.960
2d	0.997	0.996
3a	0.943	0.926
3b	0.993	0.990
3c	0.961	0.958
3d	0.999	0.998
4a	0.908	0.902
4b	0.900	0.900
4c	0.901	0.896
4d	0.897	0.893

flip angle = 8 degree). A simultaneous multi-slice pulse sequence with an acceleration factor of eight was used to acquire two resting state fMRI runs on each day, with each run consisted of 1200 volumes sampled every 0.72 seconds, at 2-mm isotropic spatial resolution (acquisition time:14 min 24 sec, TE = 33.1 ms, flip angle = 52 degree, 72 axial slices). Participants were instructed to keep their eyes open and fixated on a cross hair on the screen, while remaining as still as possible. Each run was minimally preprocessed, and artifacts were removed using the Oxford Center for Functional MRI of the Brain's (FMRIB) Independent Component Analysis-based X-noiseifier (ICA +FIX) procedure. At this point in the processing pipeline, the data acquired on each day were represented as a time series of gray ordinates, a combination of cortical surface vertices and subcortical standard-space voxels. Each run was temporally demeaned and variance normalized. All four runs for all subjects were fed into MELODIC's Incremental Group-Principal Component Analysis (MIGP) algorithm, which estimated the top 4500 weighted spatial eigenvectors. GICA was applied to the output of MIGP using FSL's MELODIC tool using six different dimensions (i.e., number of independent components: 15, 25, 50, 100, 200 and 300). Dual-regression was used to map group level spatial maps of

the components onto each subject's time series data. The full set of group level maps were used as spatial regressors against each subject's full time series (4800 volumes) to obtain a single representative time series per IC. The functional assignment of each component was determined using the Yeo Resting State Networks (Thomas Yeo et al. 2011), leading to 15 ROI's across 7 brain networks including visual, somatomotor, dorsal attention, ventral attention, frontoparietal, default mode and cerebellum (see Figure 1.7). The 2 runs per session for each subject are concatenated by temporal order, yielding T=2400 points per subject per session in the analysis.

1.4.2 Hyper-parameter values, posterior sampling and order selection

We used weakly informative hyperpriors that are conjugate but almost flat on the parameter space. In particular, the tuning parameters $\lambda_{1,k}^2$ and $\lambda_{2,k}$, $k = 1, \ldots, R^2 p$, follow the $\Gamma(1, 0.001)$ and $\Gamma(1, 0.01)$ distributions, respectively. The diagonal elements of Θ_v and Θ_u are follow the $\Gamma(1, 0.01)$ distribution, and the precision matrix Ω follows the Wishart$(1, (R - 1)I_R)$ distribution. The multiplicative variables α and β follow diffuse normal distributions MVN$(0, 100I_{R^2 p})$.

Previous studies (Martínez-Montes et al. 2004, Valdés-Sosa et al. 2005) find that the order of a VAR model for functional connectivity is usually quite small (1 or 2). We therefore fitted a VAR(1) model and a VAR(2) model, using the AIC as a model selection criterion. For each model, we ran 50,000 iterations across 3 chains, using the first 25,000 iterations as warm-up. The second 25,000 iterations were thinned, keeping every 75-th iteration, to break the autocorrelation in the posterior samples, yielding 250 samples per chain for posterior inference. All 3 chains mixed well and converged to the same stationary distribution for all parameters using the criteria $\hat{R} = \sqrt{\frac{n-1}{n} + \frac{B}{nW}} < 1.1$, where B and W are between and within chain variances of the posterior samples (Gelman et al. 2014), with no autocorrelation in the posterior samples. For these data, the VAR(2) model, with AIC =27245873, is marginally
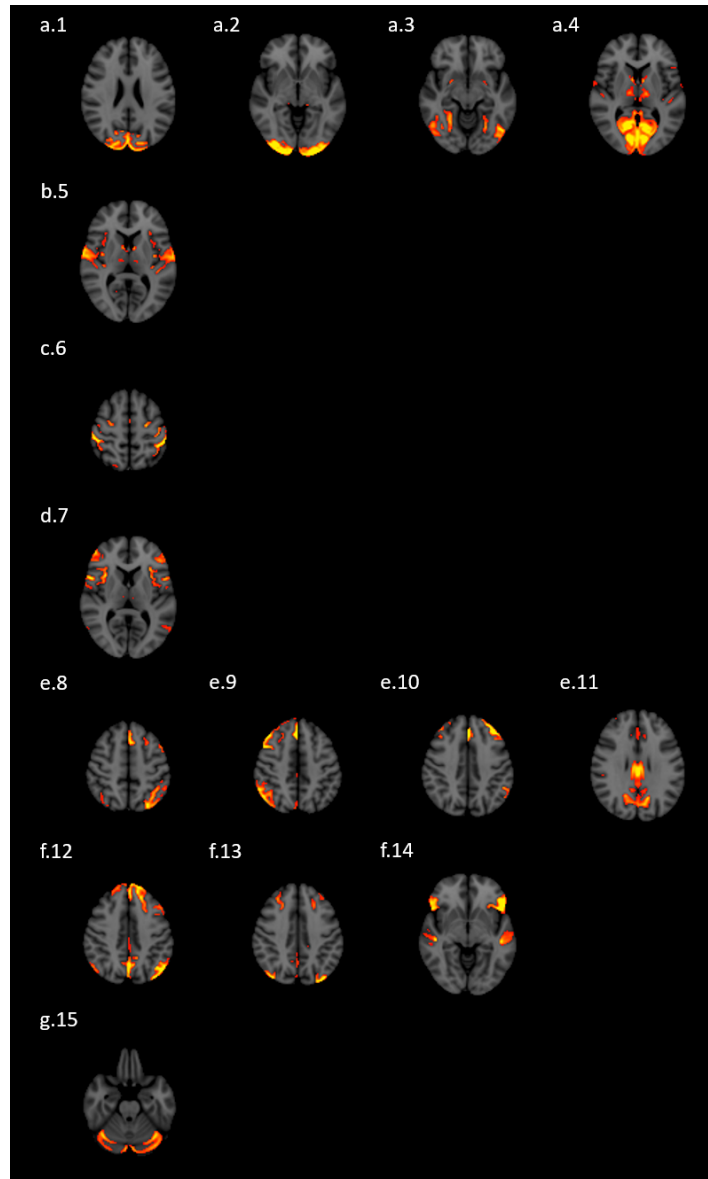
Figure 1.7: An illustration of the 15 independent components from the HCP example separated into 7 brain networks. In each panel, hot colors (yellow-to-red) indicate the spatial locations of voxels in a single representative slice of the brain that compose the components. Rows a-g show, respectively, components from the visual (a.1-a.4), somatomotor (b.5), dorsal attention (c.6), ventral attention (d.7), frontoparietal (e.8-e.11), default mode (f.12-f.14) and cerebellum (g.15) networks. Images provided by the Human Connectome Project.
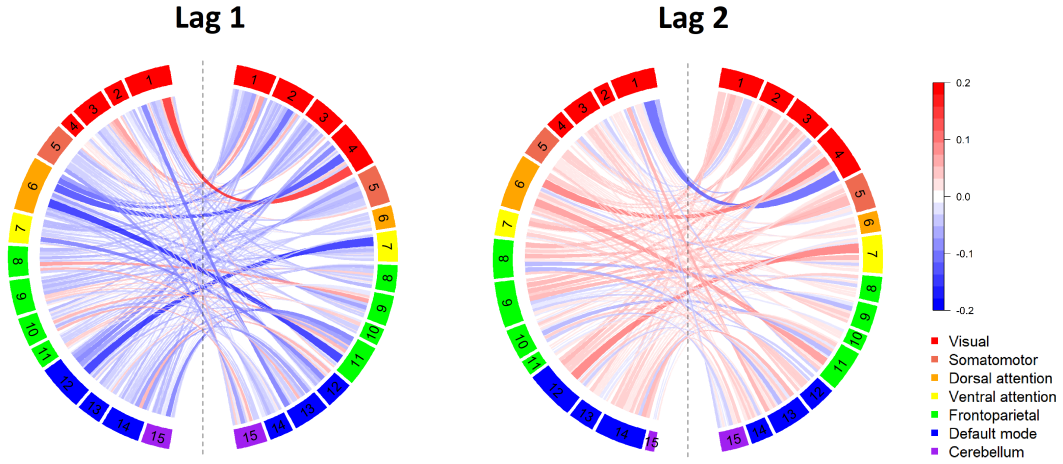
Figure 1.8: Posterior modes of the lag one (left) and lag two (right) between region group level connectivity parameters (w); numbers 1-15 represent the 15 ROI's and the colors of the sections represent the 7 brain networks. The lines from the regions on the left half circle to those on the right half circle represent the lag one or lag two coefficients, with negative coefficients colored blue and positive coefficients red. The color scale ranges from -0.2 to 0.2.

favored over the VAR(1) model, with AIC = 27297369.

1.4.3 Findings

1.4.3.1 Group level connectivity

Figure 1.8 displays the posterior modes of the lag one and lag two between region group level connectivity parameters (w) between the 15 ROI's across 7 brain networks: the numbers 1-15 represent the ROI's and the colors of the sections represent the 7 brain networks. A curved line from region r' on the left half circle to region r on the right half circle represents coefficient (rr') in the lag one (lag two) group level coefficient matrix, that is, the lag one (lag two) association between the BOLD response in region r at time t and region r' at time $t - 1$, $t = 2, \ldots, T$. The blue and red lines indicate, respectively, negative and positive coefficients.

There are strong negative lag one coefficients between most of the 15 ROI's across the 7 networks, with the strongest coefficients (below -0.12) between the 6-th ROI in the dorsal attention network (orange section) and the 4-th ROI in the visual network (red section), the

same 6-th ROI in the dorsal attention network and the 11-th ROI in the frontoparietal network (green section), and the 12-th ROI in the default mode network (blue section) and the 7-th ROI in the ventral attention network (yellow section). Moreover, there are strong positive coefficients (above 0.08) within the visual network (red section) between the 1-st and 4-th ROI's and the 3-rd and 1-st ROI's. Analogous lag two and lag one coefficients generally have opposite signs, a common occurrence in VAR(2) models (e.g., see Davis et al. 2016). Nevertheless, the lag two coefficients are, in most cases, smaller in magnitude than the lag one coefficients, suggesting the associations among the 15 ROI's are generally negative.

1.4.3.2 Subject deviations and subject-session deviations

Figures 1.9 and 1.10 display, respectively, the posterior medians of the lag one and lag two between region subject deviations (v) for the first 16 subjects and the subject-session deviations (u) for the first 8 subjects. Comparison of the figures indicates that the subject deviations tend to be stronger than the subject-session deviations across different ROI's for both lag one and lag two coefficients. Moreover, the subject deviations differ among subjects in both lags. For example, subjects 9 and 13 have several large positive deviations, while subjects 3 and 10 have large negative values, suggesting substantial between subject variability. In contrast, most subject-session deviations are close to 0 (except for 1 of the first 8 subjects), suggesting less variability within subjects across sessions.

1.4.3.3 Between Subject and Between Session Variabilities

Figure 1.11 displays the posterior medians of the variances of the between region subject deviations (column 1), the subject-session deviations (column 2), and the intra-class corre-lation (ICC), the proportion of the total variance represented by the variance of the subject deviations (column 3) for lag one (row a) and lag 2 (row b) parameters. In most instances the estimated ICC > .5, suggesting more variability in between subject connectivity than within subject connectivity. Our findings are broadly consistent with those obtained by several pre-
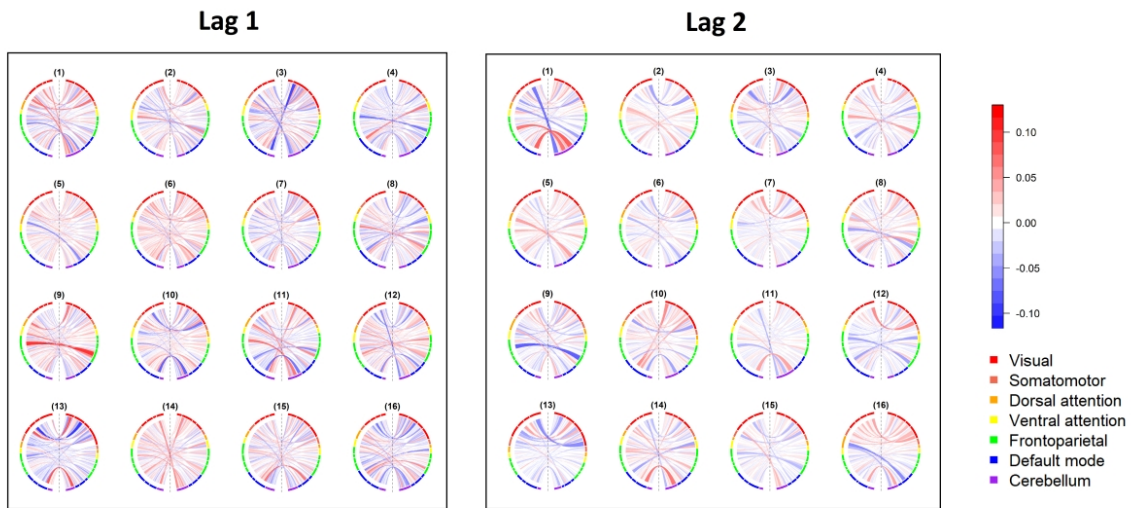
Figure 1.9: Posterior medians of the lag one (left box) and lag two (right box) between region subject deviations (v) for the first 16 subjects. The number at the top of the chord diagram is the subject number. The curved lines from the regions on the left half circle to those on the right half circle represent the lag one or lag two deviations, with negative deviations colored blue and positive deviations red; the color scale ranges from -0.13 to 0.13.
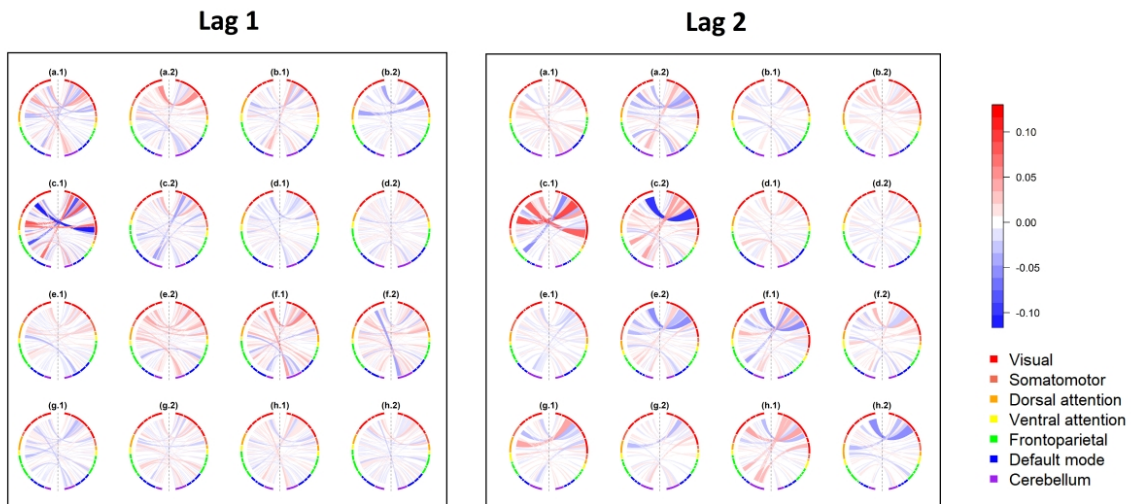
Figure 1.10: Posterior medians of the lag one (left box) and lag two (right box) between region subject-session deviations (u) for the first 8 subjects (letters a-h) with 2 sessions (numbers 1-2). The curved lines from the regions on the left half circle to those on the right half circle represent the lag one or lag two deviations, with negative deviations colored blue and positive deviations red; the color scale ranges from -0.13 to 0.13.
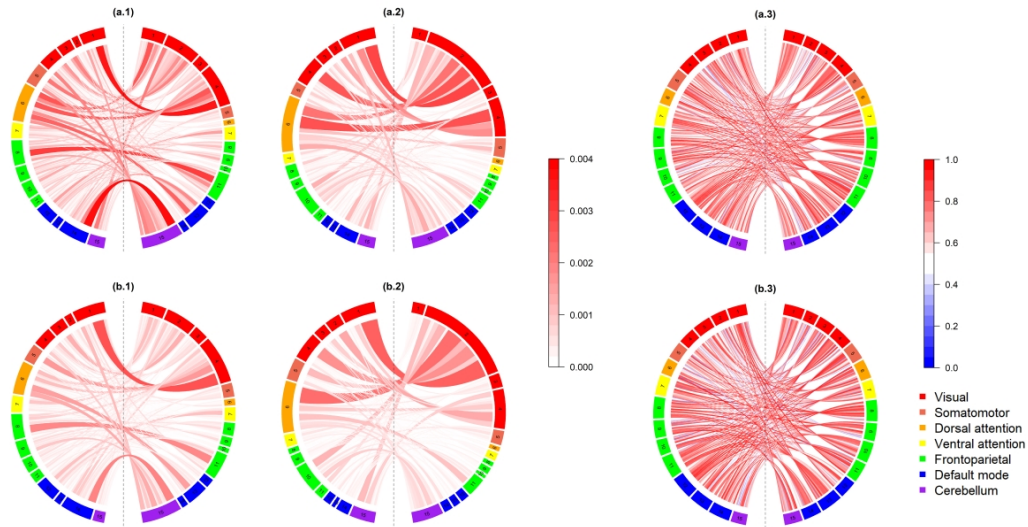
Figure 1.11: Posterior medians for the variances of the between region subject deviations (column 1) and the subject-session deviations (column 2), and the proportion of the total variance represented by the variance of the subject deviations (column 3) for lag one (row a) and lag two (row b) parameters. The curved lines from the regions on the left half circle to those on the right half circle represent the lag one or lag two deviations. For columns 1 and 2, the color scale ranges from 0 to .004 and for column 3 from 0 to 1.

vious researchers (Shehzad et al. 2009, Zhang et al. 2011, Braun et al. 2012) who used the ICC to study the test-retest reliability of some other measures of connectivity. However, these researchers used two stage methods to estimate the ICC. Further, interpreting the ICC as a test-retest reliability seems inadequate in the neuroimaging context, and can lead a researcher to inappropriately view a small ICC purely as evidence of a methodological or measurement problem, as versus a substantive finding of interest.

1.5   Discussion

We construct a sparse Bayesian hierarchical VAR model for studying resting state functional connectivity for multi-subject multi-session fMRI data. Our model extends existing multi-subject single subject Bayesian VAR models to allow simultaneous inference regarding group level, subject level and subject-session level connectivity. In particular, our model allows us

to systematically address an important question: Is the major portion of variability in resting state neural activity due to differences between subjects or differences within subjects? In addition, our computational strategy allows us to consider more subjects and ROI's than previous researchers could handle, and we relax the assumption in previous work, used to induce sparsity, of a diagonal error covariance matrix.

As demonstrated in our simulations, the model also improves on two stage procedures for connectivity metrics in which single subject models are estimated in a first step, then pieced together in a second step. The simulations also indicate that model estimates are quite accurate across a wide variety of conditions, including in small samples. However, there is a tendency for the equal-tailed credible intervals to undercover, as illustrated in Figure 2.

We study resting state data for 50 subjects from the Human Connectome Project, applying our model with 15 ROI's spanning 7 brain networks. Based on the AIC, we selected a VAR(2) model. We find negative associations between ROI's in the dorsal attention network and the visual and frontoparietal networks, and ROI's in the default mode network and the ventral attention network. Key to our approach is the decomposition of functional connectivity into group, subject, and session components allowing us to study differences within and between subjects in functional connectivity. We find greater variability between subjects than within subjects across sessions separated by a day, suggesting the major source of differences in human brain connectivity is due to differences among subjects.

In future work, we intend to increase the computational capacity of the model. Using parallel computing techniques such as the consensus Monte Carlo algorithm (Scott et al. 2013) in conjunction with our parameter expanded Gibbs sampler will allow us to work with more ROI's and subjects. Increasing the number of ROI's will allow us to 1) subdivide existing regions and assess whether the connections among regions we have discovered require additional refinement, and 2) extend the number of regions under investigation to a whole brain ROI model, allowing us to potentially discover new connections between regions and

networks. The ability to handle larger samples will allow us to estimate parameters more reliably. It will also allow us to better differentiate among types of individuals, e.g., healthy controls and autistic individuals, and more generally allow for the exploration of sub-group differences, e.g., men and women.

Appendix A

Appendix to Chapter 1

A.1 Initial values

We use Method-of-Moments-like estimates as initial values to speed up posterior sampling. Specifically, we first obtain the maximum likelihood estimates (MLE's) of the subject-session specific VAR coefficients $\hat{\boldsymbol{w}}_{is}, i = 1, ..., N, s = 1, ..., S$:

$$\hat{\boldsymbol{w}}_{is} = \left[\left(H_{is} H_{is}{}^T \right)^{-1} H_{is} \otimes I_R \right] \boldsymbol{y}_{is..}. \tag{A.1}$$

We then use these MLE's to form initial values for the group level coefficient parameters, subject deviations and subject-session deviations:

$$\boldsymbol{w}^0 = \frac{1}{NJ} \sum_{i=1}^{N} \sum_{s=1}^{S} \hat{\boldsymbol{w}}_{is},$$

$$\boldsymbol{v}_i{}^0 = \frac{1}{S} \sum_{s=1}^{S} (\hat{\boldsymbol{w}}_{is} - \hat{\boldsymbol{w}}), i = 1, \ldots N,$$

$$\boldsymbol{u}_{is}{}^0 = \hat{\boldsymbol{w}}_{is} - \hat{\boldsymbol{w}} - \hat{\boldsymbol{v}}_i, i = 1, \ldots N, s = 1, \ldots, S.$$

Using these initial values reduces the computational time by 60%.

A.2 Full conditional distributions

To draw from the posterior distribution using the Gibbs sampler, one draws each parameter from its conditional distribution, given all other parameters at their current values. Let $g(\cdot)$ denote the conditionals of one parameter given the others. We now give the full conditionals for all parameters:

1. $g(\boldsymbol{w}) = \text{MVN}\left(M_w, S_w^2 \right)$, where

$$S_w^2 = \left[\sum_{i=1}^{N} \sum_{s=1}^{S} \left(H_{is} H_{is}{}^T \right) \otimes \Omega + D \right]^{-1},$$

$$M_w = S_w^2 \left[\sum_{i=1}^{N} \sum_{s=1}^{S} \left(H_{is} H_{is}{}^T \right) \otimes \Omega \left(\hat{\boldsymbol{w}}_{is} - \boldsymbol{\alpha} * \boldsymbol{v}_i - \boldsymbol{\beta} * \boldsymbol{u}_{is} \right) \right],$$

where $\hat{\boldsymbol{w}}_{is}$ is the MLE given by (A.1);

2. $g(\mathbf{v}_i) = \text{MVN}\left(M_{v_i}, S^2_{v_i}\right), i = 1, ..., N$, where

$$S^2_{v_i} = \left[\sum_{s=1}^{S} \text{Diag}(\boldsymbol{\alpha}) \left(\left(H_{is}H_{is}{}^T\right) \otimes \Omega\right) \text{Diag}(\boldsymbol{\alpha}) + \Theta_v)\right]^{-1},$$

$$M_{v_i} = S^2_{v_i} \left[\sum_{s=1}^{S} \text{Diag}(\boldsymbol{\alpha}) \left(\left(H_{is}H_{is}{}^T\right) \otimes \Omega\right) (\hat{\mathbf{w}}_{is} - \mathbf{w} - \boldsymbol{\beta} * \mathbf{u}_{is})\right],$$

where $\text{Diag}(\cdot)$ represents assigning the $R^2 p$ vector on a diagonal matrix;

3. $g(\mathbf{u}_{is}) = \text{MVN}\left(M_{u_{is}}, S^2_{u_{is}}\right), i = 1, ..., N, s = 1, ..., S$, where

$$S^2_{u_{is}} = \left[\text{Diag}(\boldsymbol{\beta}) \left(\left(H_{is}H_{is}{}^T\right) \otimes \Omega\right) \text{Diag}(\boldsymbol{\beta}) + \Theta_u\right]^{-1},$$

$$M_{u_{is}} = S^2_{u_{is}} \left[\text{Diag}(\boldsymbol{\beta}) \left(\left(H_{is}H_{is}{}^T\right) \otimes \Omega\right) (\hat{\mathbf{w}}_{is} - \mathbf{w} - \boldsymbol{\alpha} * \mathbf{v}_i)\right];$$

4. $g(\boldsymbol{\alpha}) = \text{MVN}\left(M_\alpha, S^2_\alpha\right)$, where

$$S^2_\alpha = \left[\sum_{i=1}^{N}\sum_{s=1}^{S} \text{Diag}(\mathbf{v}_i) \left(\left(H_{is}H_{is}{}^T\right) \otimes \Omega\right) \text{Diag}(\mathbf{v}_i) + aI\right]^{-1},$$

$$M_\alpha = S^2_\alpha \left[\sum_{i=1}^{N}\sum_{s=1}^{S} \text{Diag}(\mathbf{v}_i) \left(\left(H_{is}H_{is}{}^T\right) \otimes \Omega\right) (\hat{\mathbf{w}}_{is} - \mathbf{w} - \boldsymbol{\beta} * \mathbf{u}_{is})\right];$$

5. $g(\boldsymbol{\beta}) = \text{MVN}\left(M_\beta, S^2_\beta\right)$, where

$$S^2_\beta = \left[\sum_{i=1}^{N}\sum_{s=1}^{S} \text{Diag}(\mathbf{u}_{is}) \left(\left(H_{is}H_{is}{}^T\right) \otimes \Omega\right) \text{Diag}(\mathbf{u}_{is}) + bI\right]^{-1},$$

$$M_\beta = S^2_\beta \left[\sum_{i=1}^{N}\sum_{s=1}^{S} \text{Diag}(\mathbf{u}_{is}) \left(\left(H_{is}H_{is}{}^T\right) \otimes \Omega\right) (\hat{\mathbf{w}}_{is} - \mathbf{w} - \boldsymbol{\alpha} * \mathbf{v}_i)\right];$$

6. $g(\theta_{vk}) = \Gamma(k_v, s_{vk}), k = 1, ..., R^2 p$, where

$$k_v = \frac{N}{2} + k_1,$$

$$s_{vk} = \left(\frac{1}{2}\sum_{i=1}^{N} v^2_{nk} + \frac{1}{s_1}\right)^{-1},$$

and v_{nk} is the k-th component of $\mathbf{v}_i$;

7. $g\left(\theta_{uk}\right) = \Gamma(k_u, s_{uk}), k = 1, \ldots, R^2 p$, where

$$k_u = \frac{NS}{2} + k_2,$$

$$s_{uk} = \left(\frac{1}{2} \sum_{i=1}^{N} \sum_{s=1}^{S} u_{isk}^2 + \frac{1}{s_2}\right)^{-1},$$

and u_{isk} is the k-th component of $\boldsymbol{u}_{is}$;

8. $g(\lambda_{1,k}^2) = \Gamma\left(\mu_{\lambda_{1,k}}, \nu_{\lambda_{1,k}}\right), k = 1, \ldots, R^2 p$, where

$$\nu_{\lambda_{1,k}} = \nu_1 + 2,$$

$$\mu_{\lambda_{1,k}} = \frac{(\nu_1 + 2)(\mu_1 \xi_k^2)}{\nu_1 \xi_k^2 + \tau_k^2 \mu_1};$$

9. $g(\lambda_{2,k}) = \Gamma\left(\mu_{\lambda_{2,k}}, \nu_{\lambda_{2,k}}\right), k = 1, \ldots, R^2 p$, where

$$\nu_{\lambda_{2,k}} = \nu_2 + 1,$$

$$\mu_{\lambda_{2,k}} = \frac{(\nu_2 + 1)\mu_2}{\nu_2 + w_k^2 \mu_2},$$

and w_k is the k-th component of $\boldsymbol{w}$;

10. $g(\Omega) = \text{Wishart}(S_\Omega, \nu_\Omega)$, where

$$\nu_\Omega = (T - p)NS + \nu + 2Rp,$$

$$S_\Omega^{-1} = \left[\sum_{i=1}^{N} \sum_{s=1}^{S} S_{is} + K^{-1} + 2QQ^T\right]^{-1},$$

$$S_{is} = (Y_{is} - W_{is}H_{is})(Y_{is} - W_{is}H_{is})^T$$

is the residual matrix of the i-th subject and the s-th session,

$$Y_{is} = \left(\boldsymbol{y}_{is,p+1,.} \quad \cdots \quad \boldsymbol{y}_{isT.}\right)$$

is a $R \times (T - p)$ matrix,

$$W_{is} = vec^{-1}\left(\boldsymbol{w} + \boldsymbol{\alpha} * \boldsymbol{v}_i + \boldsymbol{\beta} * \boldsymbol{u}_{is}\right),$$

38

where $vec^{-1}(\cdot)$ stacks the R^2p-coefficient vector back into a $R \times Rp$ matrix,

$$Q = \begin{pmatrix} \gamma_1 & \cdots & \gamma_{(Rp-1)R+1} \\ \vdots & \ddots & \vdots \\ \gamma_R & \cdots & \gamma_{R^2p} \end{pmatrix},$$

where $\gamma_1, ..., \gamma_{R^2p}$ are given by

$$\gamma_{R^2p} = \tau_{R^2p}\lambda_{1,R^2p},$$

$$\gamma_k = \tau_k\lambda_{1,k} - \boldsymbol{m}_k \left(\gamma_{k+1}, ..., \gamma_{R^2p}\right)^T, k = 1, ..., R^2p - 1,$$

and

$$\boldsymbol{m}_k = \left(M_{k,k+1}, ..., M_{k,R^2p}\right) \begin{pmatrix} M_{k+1,k+1} & \cdots & M_{k+1,R^2p} \\ \vdots & \ddots & \vdots \\ M_{R^2p,k+1} & \cdots & M_{R^2p,R^2p} \end{pmatrix}^{-1};$$

11. $g(\frac{1}{2\tau_k^2}) = \text{IG}(a_k, b_k), k = 1, ..., R^2p$, where IG denotes the Inverse-Gaussian distribution

with density function $f(x; a, b) = \sqrt{\frac{b}{2\pi}}x^{-\frac{3}{2}} \exp\left\{-\frac{b(x-a)^2}{2a^2x}\right\}, x > 0, a > 0, b > 0$, and

$$a_k = \sqrt{\frac{\lambda_{1,k}^2}{w_k^2\xi_k^2}},$$

$$b_k = \frac{\lambda_{1,k}^2}{\xi_k^2}.$$

Note that the conditional distribution of $2\tau_k^2$ does not have a closed form, but Park and

Casella (2008) showed that $g(2\frac{1}{\tau_k^2})$ is the Inverse-Gaussian distribution. Therefore, we

first draw $\frac{1}{2\tau_k^2}$ from $g(2\frac{1}{\tau_k^2})$ and then transform back to obtain $2\tau_k^2$.

A.3 Error covariance matrices

The Wishart distributions used in the simulation study have 10 degrees of freedom and symmetric Toeplitz scale matrices of the form:

$$
\begin{pmatrix}
\rho & \rho^2 & \cdots & \rho^5 \\
\rho^2 & \rho & \cdots & \rho^4 \\
\vdots & \vdots & \ddots & \vdots \\
\rho^5 & \rho^4 & \cdots & \rho
\end{pmatrix},
$$

where $\rho = 0.3$ and 0.1 for cases 4a and 4b, respectively, and 0.2 for all other cases. The block diagonal covariance matrix for case 4c is:

$$
\begin{pmatrix}
2 & .4 & & & \\
.4 & 2 & & & \\
& & 2 & .4 & .4 \\
& & .4 & 2 & .4 \\
& & .4 & .4 & 2
\end{pmatrix},
$$

and the diagonal covariance matrix for case 4d is $2I_5$.

Chapter 2

Bayesian hierarchical vector autoregressive models for patient-level predictive modeling

2.1 Introduction

Analyses of patient-level observational healthcare databases underpin much of the current evidence base for healthcare practice. Administrative claims, electronic health record (EHR), and patient diary databases in particular have seen increased use in the past decade owing to greater availability at lower costs and technological advances that made computational processing on large-scale data more feasible (Madigan et al. 2014). Because the data reflect healthcare activity within a real-world population, they offer the potential to complement clinical trial results. Administrative claims databases have been the most actively used observational healthcare data source. These databases typically capture data elements used within the reimbursement process, as providers of healthcare services (e.g., physicians, pharmacies, hospitals, and laboratories) must submit encounter information to enable payment for their services (Hennessy 2006).

Neither administrative claims nor EHRs represent the ideal information required to generate reliable evidence. For example, diagnoses recorded on medical claims are used to support justification for the payment for a given visit or procedure; a given diagnosis could represent the condition that the procedure was used to rule out or could be an administrative artifact (e.g., the code used by a medical coder to maximize the reimbursement amount). Some diagnosis codes have been studied through source record verification and have demonstrated adequate performance characteristics (see, for example, Donahue et al. 1997, Hennessy et al. 2010), whereas other conditions and systems provide less certainty (see, for example, Strom 2001, Harrold et al. 2007, Lewis et al. 2007). Limitations exist in EHR systems as well, in which, apart from concerns about incomplete capture, data may be artificially manipulated to serve clinical care (e.g., an incorrect diagnosis recorded to justify a desired medical procedure).

Patient diary databases capture fewer variables and smaller numbers of patients but offer significant advantages. In particular, diaries can capture longitudinal data about activities of daily living, mood, and personal habits and not just when a patient is seeking healthcare.

Patient diaries support significant research activities in chronic disease management (Zettl et al. 2016, Nagai et al. 2016, Coolbrandt et al. 2017). Electronic patient diaries capture data via a web interface or a phone or tablet and represent the most common modality. Participating patients make diary entries either daily or at multiple times per day, or at set intervals.

Major use cases for observational healthcare databases include inferences about the causal effects of healthcare interventions (typically via cohort, case-control, or self-controlled case series designs), healthcare characterization (see, for example, Hripcsak et al. 2016), and patient-level predictive modeling (PLPM), the primary focus of this paper. PLPM lies right at the core of healthcare practice and so-called precision medicine. At least implicitly, every healthcare intervention decision involves a patient-level prediction.

Many researchers have considered the problem of developing predictive models from patient-level data (see, for example, Kansagara et al. 2011, Stiglic et al. 2015). In the vast majority of cases, researchers generated a set of candidate covariates thought to be related to the outcome of interest and employed traditional statistical or machine learning algorithms to predict the outcome using these covariates. He et al. (2014), for example, predict hospital readmissions using logistic regression with a modified forward variable selection algorithm to choose features from the set of all indicator variables for any past occurrence of a medical concept with a diagnostic code. None of He et al.'s many features, however, conveys any information about time. Other recent similar approaches include Sanchez-Morillo et al. (2016), Weng et al. (2017), and Ye et al. (2017).

Some authors have attempted to incorporate explicit temporal information into patient-level predictive models. In the context of stroke prediction, Shahn et al. (2015) proposed a "relational random forests" (RRF) that capture "motifs" in the medical history that are predictive of specific future events. Motifs capture healthcare episodes such as "received drug X or diagnosis Y then suffered condition Z within T days." In their target application, RRF showed modest benefits in predictive performance by reducing predictive errors in the

test set. Wang et al. (2012) consider a patient's health history as the superposition and concatenation of multiple pattern matrices. Each pattern matrix specifies a rigid temporal relationship among health events that repeats over time. They employ a matrix factorization algorithm to learn the pattern matrices, which can then be used to construct features for predictive models. Brzan et al. (2017) show that incorporation of historical patient data improves predictive performance for hospital readmission prediction.

In this paper, we approach PLPM as a multivariate time series modeling problem and consider a temporal generative model for the medical record. We adopt a Bayesian approach and consider an application to patient diary data in the context of drug and alcohol treatment. We explore the predictive performance of this model. Since the model offers advantages in terms of interpretability as compared with alternatives such as random forests and deep learning, we explore insights derived from the model. We apply our model to two patient diary datasets: the first one includes substance use craving, negative affect, and tobacco use (Zheng et al. 2013), and the second one includes functional somatic symptoms (FFS) and psychological discomforts (van Gils et al. 2014).

2.2 Vector autoregressive models for patient diary data

Vector autoregressive (VAR) models represent a mainstay of multivariate time analysis (Lütke-pohl 2005). Previous applications of VAR models to patient diary data include Zheng et al. (2013) and van Gils et al. (2014). However, these studies fit separate models for each individual patient. Here we take a Bayesian approach and simultaneously model *all* patients using a hierarchical model. Bayesian VAR models have been widely applied in various fields of study such as economics (Litterman 2010, Koop and Korobilis 2010, Karlsson 2015) and neuroimaging (Gorrostieta et al. 2013). Here we adopt a flexible Elastic-net prior to induce sparsity in the estimation of the coefficients to aid interpretability (Gefang 2014). The hierarchical structure for the subject-level coefficients enables simultaneous inference about both

the population-level and the subject-level coefficients and, in our target application, improves predictive performance. Our approach also enables estimation of between-subject heterogeneity by contrast with ad-hoc approaches such as the clustering analysis on the subject-specific VAR coefficient estimates deployed by Zheng et al. (2013).

2.2.1 Sparse hierarchical VAR model

Let N denote the number of patients. We assume that each patient records diary measurements at T_n time points, $n = 1, \ldots, N$ on R variables. In what follows we assume daily measurements. Let $\mathbf{y}_{nt} = (y_{nt1}, \ldots, y_{ntR})^T$ be a column vector representing the R measurements at time t, $t = 1, \ldots, T_n$, for patient n, $n = 1, \ldots, N$. Then, for each patient, the sparse hierarchical VAR(p) model assumes

$$\mathbf{y}_{nt} = \sum_{i=1}^{p} A_{ni} \mathbf{y}_{n,t-i} + \boldsymbol{\epsilon}_{nt}, t = p + 1, \ldots, T_n, \tag{2.1}$$

where $A_{ni}, i = 1, \ldots, p$ is a $R \times R$ coefficient matrix representing the lag-i association, and $\boldsymbol{\epsilon}_{nt}$ is the errors, assumed to follow i.i.d. multivariate normal distribution with mean $\mathbf{0}$ and precision matrix $\Omega : \boldsymbol{\epsilon}_{nt} \sim \text{MVN}(\mathbf{0}, \Omega^{-1})$. We write the model in concise matrix form as:

$$\mathbf{y}_n = \left(H_n^T \otimes I_B \right) \mathbf{w}_n + \boldsymbol{\epsilon}_n, \tag{2.2}$$

where

$$\mathbf{y}_n = \begin{pmatrix} \mathbf{y}_{n,p+1} \\ \vdots \\ \mathbf{y}_{nT_n} \end{pmatrix} \tag{2.3}$$

is the response variable and $\otimes$ is the Kronecker product,

$$H_n = \begin{pmatrix} \mathbf{y}_{np} & \cdots & \mathbf{y}_{n,T_n-1} \\ \vdots & \ddots & \vdots \\ \mathbf{y}_{n1} & \cdots & \mathbf{y}_{n,T_n-p} \end{pmatrix} \tag{2.4}$$

represents the regressors,

$$\boldsymbol{w}_n = vec([A_{n1}, ..., A_{np}]) \tag{2.5}$$

is the vectorized coefficients of length $R^2 p$,

$$\boldsymbol{\epsilon}_n = \begin{pmatrix} \epsilon_{n,p+1} \\ \vdots \\ \epsilon_{nT_n} \end{pmatrix} \sim \text{MVN}(\boldsymbol{0}, I_{T_n-p} \otimes \Omega^{-1}) \tag{2.6}$$

is the stacked error term, and I_k represents a $k \times k$ identity matrix.

Here we assume that the precision matrix Ω is the same for all patients but the coefficient vector $\boldsymbol{w}_n$ is unique for patient n. Specifically, we assume that the coefficient vector for the n-th patient comprises two parts:

$$\boldsymbol{w}_n = \boldsymbol{w} + \boldsymbol{v}_n, \tag{2.7}$$

where $\boldsymbol{w}$ is the population-level coefficient and is the same across all patients, $\boldsymbol{v}_n$ is the patient-level deviation for the n-th patient.

For the population-level coefficient, we adopt the doubly adaptive Elastic-net prior (Gefang 2014). Specifically, we assume that the prior for $\boldsymbol{w}$ follows a multivariate normal distribution with mean $\boldsymbol{0}$ and precision matrix D:

$$\boldsymbol{w}|D \sim \text{MVN}\left(\boldsymbol{0}, D^{-1}\right), \tag{2.8}$$

where D is a diagonal matrix and the diagonal elements depend on hyperparameters $2\tau_k^2$ and $\lambda_{2,k} > 0, k = 1, ..., R^2 p$:

$$\text{diag}(D) = \left(\lambda_{2,1} + \frac{1}{2\tau_1^2}, ..., \lambda_{2,R^2 p} + \frac{1}{2\tau_{R^2 p}^2}\right). \tag{2.9}$$

Here $2\tau_k^2, k = 1, ..., R^2 p$, is assumed to follow independent exponential distribution with rate parameter $\frac{\lambda_{1,k}^2}{2\xi_k^2}, k = 1, ..., B^2 p$:

$$2\tau_k^2|2\xi_k^2, \lambda_{1,k} \sim E\left(\frac{2\xi_k^2}{\lambda_{1,k}^2}\right), \tag{2.10}$$

where $\lambda_{1,k} > 0$, $k = 1, ..., B^2p$, and ξ_k^2, $k = 1, ..., R^2p$, depend on the error precision matrix Ω through:

$$\xi_k^2 = M_{k,k} - \left(M_{k,k+1}, ..., M_{k,R^2p} \right) \begin{pmatrix} M_{k+1,k+1} & \cdots & M_{k+1,R^2p} \\ \vdots & \ddots & \vdots \\ M_{R^2p,k+1} & \cdots & M_{R^2p,R^2p} \end{pmatrix}^{-1} \begin{pmatrix} M_{k,k+1} \\ \vdots \\ M_{k,R^2p} \end{pmatrix}, \tag{2.11}$$

$$M = I_{Rp} \otimes \Omega^{-1}. \tag{2.12}$$

We also assume that $\lambda_{1,k}^2$ and $\lambda_{2,k}$, $k = 1, ..., R^2p$ follow two i.i.d. gamma distributions:

$$\lambda_{1,k}^2 \sim \Gamma(\mu_1, \nu_1) \tag{2.13}$$

$$\lambda_{2,k} \sim \Gamma(\mu_2, \nu_2) \tag{2.14}$$

where μ_1, ν_1 and μ_2, ν_2 are the corresponding mean and degree of freedom parameters, assumed to be known. Here, $\lambda_{1,k}$ and $\lambda_{2,k}$ represent the L1 and L2 tuning parameters for the k-th element of w, respectively. It can be shown that the conditional prior for each of the elements of w given the error precision matrix Ω comprise a normal distribution and a Laplace distribution. In other words, this prior provides element-wise shrinkage for the Elastic-net regularization. Note that conditioning on the error precision matrix Ω is important as it ensures unimodal posterior distribution for each of the elements of w (Park and Casella 2008).

For the patient-level deviations, we apply conjugate i.i.d. multivariate normal priors with mean $\mathbf{0}$ and diagonal precision matrix Θ_v:

$$v_n | \Theta_v \sim \text{MVN} \left(\mathbf{0}, \Theta_v^{-1} \right), \tag{2.15}$$

for $n = 1, ..., N$, where the diagonal elements of Θ_v are

$$\text{diag}(\Theta_v) = (\theta_{v1}, ..., \theta_{v,R^2p}), \tag{2.16}$$

for which we apply the gamma hyperpriors on each of the diagonal elements θ_{vk}, $k = 1, ..., R^2 p$, i.i.d.:

$$\theta_{vk} \sim \Gamma(k, s), \tag{2.17}$$

where k and s are the known mean and degree of freedom parameters (see Gefang (2014) for parameterization of the gamma distribution). We assume that the patient-level coefficients are independent of the population-level coefficients given the hyperparameters.

Finally, we use a conjugate Wishart distribution for the error precision matrix Ω:

$$\Omega \sim \text{Wishart}(K, \nu), \tag{2.18}$$

where K is the $R \times R$ scale matrix and ν is the degree of freedom parameter, both assumed to be known.

2.2.2 Posterior inference

We use Markov Chain Monte Carlo (MCMC) to generate draws from the posterior distribution, specifically, a Gibbs sampler with parameter expansion (Gelman et al. 2014). A common computational challenge of Gibbs sampling for hierarchical regression models is that the sampler can be very slow when there is high dependence between the coefficients and their variance parameters. For our model, there is strong dependence between v_n and its precision parameters Θ_v: when the current draw of Θ_v is large, the next draw of v_n will be small, which in turn makes the next draw of Θ_v even larger, and so on, so the sampler can take an impractically long time to explore the entire parameter space.

Using parameter-expansion allows us to reduce inter-parameter correlations and speeds convergence. Specifically, we add an element-wise multiplicative factor α to v_n (i.e., multiply α_k to v_{nk}, $k = 1, ..., R^2 p$) and draw α from a multivariate normal distribution just like the other parameters:

$$\alpha \sim \text{MVN}(\mathbf{0}, a I_{R^2 p}), \tag{2.19}$$

where a is a known constant. In this case, when the current draw of Θ_v is large, the next draw of $\boldsymbol{v}_n$ will be small, but the next draw of $\boldsymbol{\alpha}$ will be large, which results in the next draw of $\boldsymbol{v}_n$ in a normal range. Thus, the expanded parameterization allows the Gibbs sampler to move in more directions and avoid getting stuck. These new parameters have no meaning. However, we are not interested in either $\boldsymbol{\alpha}$ or $\boldsymbol{v}_n$. Rather, the subject-level parameters in this new parameterization become

$$\boldsymbol{v}_n^* = \boldsymbol{\alpha} * \boldsymbol{v}_n, \tag{2.20}$$

and the true precision of these coefficients become

$$\theta_{vk}^* = \alpha_k^{-2}\theta_{vk}, k = 1, ..., R^2 p. \tag{2.21}$$

After obtaining the posterior distributions, we use the posterior modes as the point estimates for the parameters in the model. Using posterior modes as point estimates induces sparsity, and is equivalent to the estimates given by the penalized likelihood methods such as Lasso (Genkin et al. 2007, Park and Casella 2008)). Here we use the values with the highest empirical density as the modes of the posterior distribution. This is a crude approach to find the posterior modes but is computationally efficient for problems with many parameters.

Finally, we use AIC as a criterion to select the optimal order of the VAR model.

2.3 Application 1: predicting substance use craving, negative affect, and tobacco use among young adults in recovery

College and university students experience high rates of substance abuse disorders, with 22.9% meeting the diagnostic criteria versus 8.5% of the general population (National Center on Addiction and Substance Abuse at Columbia University [CASA], 2007). Recognition of this problem has led at least 20 colleges and universities to develop collegiate recovery

communities or CRC (Smock et al. 2011) that provide comprehensive recovery support services and one of these communities provided the diary data for the Zheng et al. (2013) study. Zheng et al. (2013) examined day-to-day associations among substance use cravings, negative affect, and tobacco use among 30 college students in 12-step recovery from drug and alcohol addictions. The intraindividual variability of relevant psychological states combined with the "one day at a time" nature of sustained abstinence warrant a day-to-day investigation of substance use recovery. The authors fit first-order vector autoregression models to each individual predicting daily levels of substance use cravings, negative affect, and tobacco use.

Extensive research has established the relevance of craving as a key criterion in recovery research (e.g., Stalcup et al. 2006). In turn, negative affect is believed to be a predictor of craving, and tobacco use is both an important correlate of substance use and craving and a serious public health risk in its own right. Whether tobacco has any beneficial effects for persons in recovery remains an open question.

Zheng et al. (2013) present a series of conclusions finding overall that the study revealed "substantial person-level heterogeneity in the day-to-day processes that challenge continued abstinence within a college recovery community, providing a picture of accumulated daily recovery risk that could threaten abstinence over both the short and long term."

2.3.1 Data and the preprocessing

As in Zheng et al. (2013), the data consist of 55 adult addict patients from CRC at a Southwestern university with a total number of 1,222 diary reports (Celeveland and Harris 2010). In this study, we included 25 patients from the full sample, excluding 12 of the 55 who are non-smokers, 4 whose diary records exhibit no day-to-day variance, and 14 who have 3 or more missing daily records. Note that our inclusion criteria differ from those in Zheng et al. (2013), leading to the inclusion of a few more patients than were included in their study. All participants were fully anonymized before accessing to the data. All are non-Hispanic White.

All had received professional alcohol/drug dependency treatment. All had received inpatient care, most for 3 months or more. All considered themselves 12-step group members and reported that they read 12-step literature and applied the steps to their lives on a daily basis. Participants provided an average of 25.56 days' worth of data each (ranging from 10 to 33 days), with the average participant missing 0.8 days (ranging from 0 to 2 days).

The following three variables are considered in our application. They are the same as in Zheng et al. (2013) (see Zheng et al. (2013) for descriptive statistics for the three variables and the number of observations for each participant).

Tobacco use. Daily tobacco use was measured with one item asking, "How many cigarettes did you smoke today?" Responses ranged from 0 (no cigarettes), 1 (1 or 2), 2 (2 to 5), 3 (5 to 10), 4 (half pack) to 5 (full pack plus).

Negative affect. Zheng et al. (2013) used the 10-item negative affect scale from the Positive and Negative Affect Scale to assess daily negative affect. Emotions included "Stressed," "Upset," "Scared," "Hostile," and "Irritable." Responses ranged from 1 (very slightly or not at all) to 5 (very much).

Substance use craving. Daily substance use craving was measured with seven items modified from the Desires for Alcohol Scale and the Alcohol Urges Questionnaire to accommodate daily assessment and polydrug use. A sample item reads, "For a moment today I missed the feeling of drinking or drugging." Responses were 1 (strongly disagree) to 5 (strongly agree).

Our pre-processing step includes missing data imputation using the mean of the non-missing consecutive days, log-transformation of all three variables to improve normality, standardization to mean 0 and variance 1 and removal of linear trend for each time series for each patient.

2.3.2 Hyperpriors used in the model

We used weakly informative priors that are conjugate but almost flat on the parameter space in the spirit of Gefang (2014). In particular, we assume $\lambda_{1,k}^2$ and $\lambda_{2,k}$, $k = 1, ..., R^2 p$, follow the $\Gamma(1, 0.001)$ and the $\Gamma(1, 0.01)$ distributions, respectively, the diagonal elements of Θ_v are assumed to follow $\Gamma(1, 0.01)$, and the precision matrix of the error terms Ω is assumed to follow Wishart$((R-1)I_R, 1)$. We assume the multiplicative variables $\boldsymbol{\alpha}$ follow a diffuse normal distribution: MVN$(\mathbf{0}, 100 I_{R^2 p})$.

2.3.3 Criterion to assess model performance

We use prediction accuracy as a criterion to assess the performance of the proposed Bayesian VAR model to compare with the patient-specific VAR models used in Zheng et al. (2013). In particular, all but the last daily record of each individual patient are used to train the model and the last daily record is used as a test set.

We use the posterior predictive distributions to predict the test data (the procedure is described in detail in Karlsson 2015). Specifically, each individual sample of the 1-step ahead forecast for subject n, $n = 1, ..., N$, is generated by the posterior samples of the VAR coefficients and the error precision matrix:

$$\hat{\boldsymbol{y}}_{n,T_{n+1}} = \sum_{i=1}^{p} \hat{A}_{ni} \boldsymbol{y}_{n,T_{n+1-i}} + \hat{\boldsymbol{\epsilon}}_{n,T_{n+1}}, \tag{2.22}$$

where $\hat{\boldsymbol{\epsilon}}_{n,T_{n+1}}$ is drawn from a multivariate normal distribution with mean 0 and precision matrix equal to one posterior sample of Ω:

$$\hat{\boldsymbol{\epsilon}}_{n,T_{n+1}} \sim \text{MVN}(\mathbf{0}, \hat{\Omega}^{-1}). \tag{2.23}$$

Posterior means and the 2.5-th and 97.5-th posterior percentiles of these posterior predictive samples are used as point and interval estimates of the prediction. Note that using posterior modes for parameter estimation enforces sparsity and interpretability while the use of posterior

means minimizes mean squared error (MSE) for forecasts (Karlsson 2015). For the patient-specific VAR model that was utilized in Zheng et al. (2013), patient-specific maximum-likelihood estimates (MLE) are used to predict the test observations for each individual patient. Estimated MSE and coverage probability of the 95% predictive intervals (estimated by the percentage of the intervals containing the true values of the test data points) are used to assess prediction accuracy on the test set for both models.

2.3.4 Results

2.3.4.1 Posterior sampling and order selection

We fit two hierarchical VAR models with order one and two, respectively. The posterior sampling procedure was implemented in R. Specifically, we ran 10,000 iterations across 4 chains and used the first half of the iterations as warmup iterations. The second half was thinned by every 20 iterations to break the autocorrelation in the posterior samples, ending up with 250 samples per chain for posterior inference. All 4 chains mixed well and converged to similar stationary distributions for all parameters according to the criterion of $\hat{R} < 1.1$ (Gelman et al. 2014). No serial correlations are found in the chains. The number of parameters for this example is about 300 for VAR(1) and 600 for VAR(2).

For this particular example, the model with order one resulted in a smaller AIC (3027.951) and is favored. Since the second order model underperformed the first order model (AIC=3466.319), we did not pursue higher orders. We note that Zheng et al. (2013) fit VAR(1) models to each individual patient. In fact, since there is one patient who has only 9 days of daily records in the training data, maximum likelihood estimation for models of higher order than one results in singularity problems, requiring a Bayesian and/or regularized approach.

2.3.4.2 Population-level coefficients

Figure 2.1 shows the posterior modes of the population-level coefficients. The three variables of interest are shown in the circles, where "T" represents tobacco use, "N" represents negative affect and "C" represents substance use craving. The arrows represent the lag-1 associations between the variables. In particular, the red arrows show positive associations and the blue arrow shows the negative association. The solid arrows represent the "significant" associations. Specifically, the pink arrow indicates that a 90% interval does not contain 0. The transparent arrows show the associations that are not significant. One can see from Figure 2.1 that tobacco use in the previous day has the strongest positive effect on the tobacco use in the current day, followed by negative affect in the previous day on the negative affect and substance use craving in the current day. Moreover, craving in the previous day has a small negative effect on itself in the current day.

2.3.4.3 Patient-level parameters

Figure 2.2 shows the posterior modes of the patient-level deviations. In these plots, the solid red arrow indicates that a 95% posterior interval does not contain 0. We see that several lag-1 associations between the three variables of interest are zero for most patients, including negative affect and craving in the previous day on themselves in the current day, as well as negative affect in previous days on tobacco use and craving in the current day. For the non zero associations, tobacco use in the previous day has an effect on tobacco use in the current day for all patients. Craving in the previous day also has an effect on tobacco use in the current day for almost all patients. These associations vary across different patients in terms of both sign and magnitude.

Figure 2.3 shows the posterior modes of the standard deviation of the patient-level coefficients (i.e., the square root of the inverse of the diagonal elements of Θ_ν). The solid arrows show the associations with large patient-level heterogeneity: the largest one is the effect of
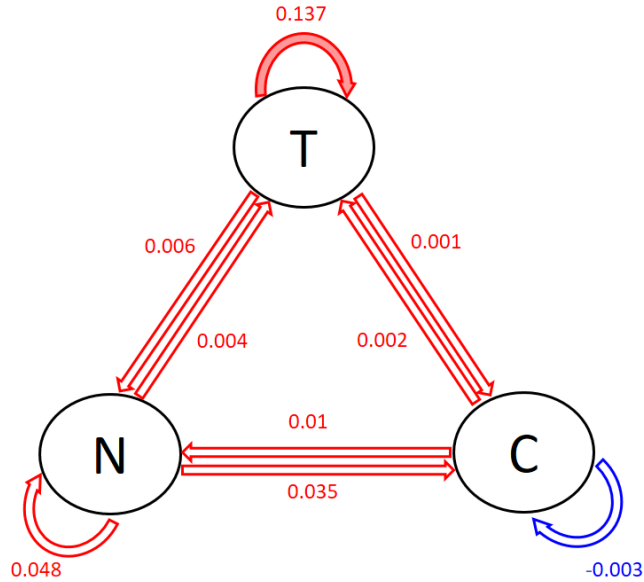
Figure 2.1: Posterior modes of the population-level coefficients. T=tobacco use, N=negative affect, C=cravings. Red=positive, blue=negative. Solid lines=significant coefficients (red: $\alpha = 0.05$; pink: $\alpha = 0.1$), transparent lines=non-significant coefficients.

previous tobacco use on current tobacco use, followed by previous craving on current tobacco use and negative affect. This is consistent with the findings in Figure 2.2.

2.3.4.4   Prediction accuracy

Table 2.1 shows the estimated MSE's of the proposed Bayesian hierarchical model compared with the conventional patient-specific VAR model. We see that the hierarchical model improved the accuracy of the prediction. The overall mean square error averaged over all three variables is reduced by 25%. For the individual variables, predictive accuracy is better using the hierarchical model relatively to the patient-specific VAR model with the mean square errors reduced by 17% for tobacco use, 3% for negative affect and more than 37% for craving, respectively. This shows the advantage in terms of predictive accuracy of the hierarchical model which pools the information across patients.
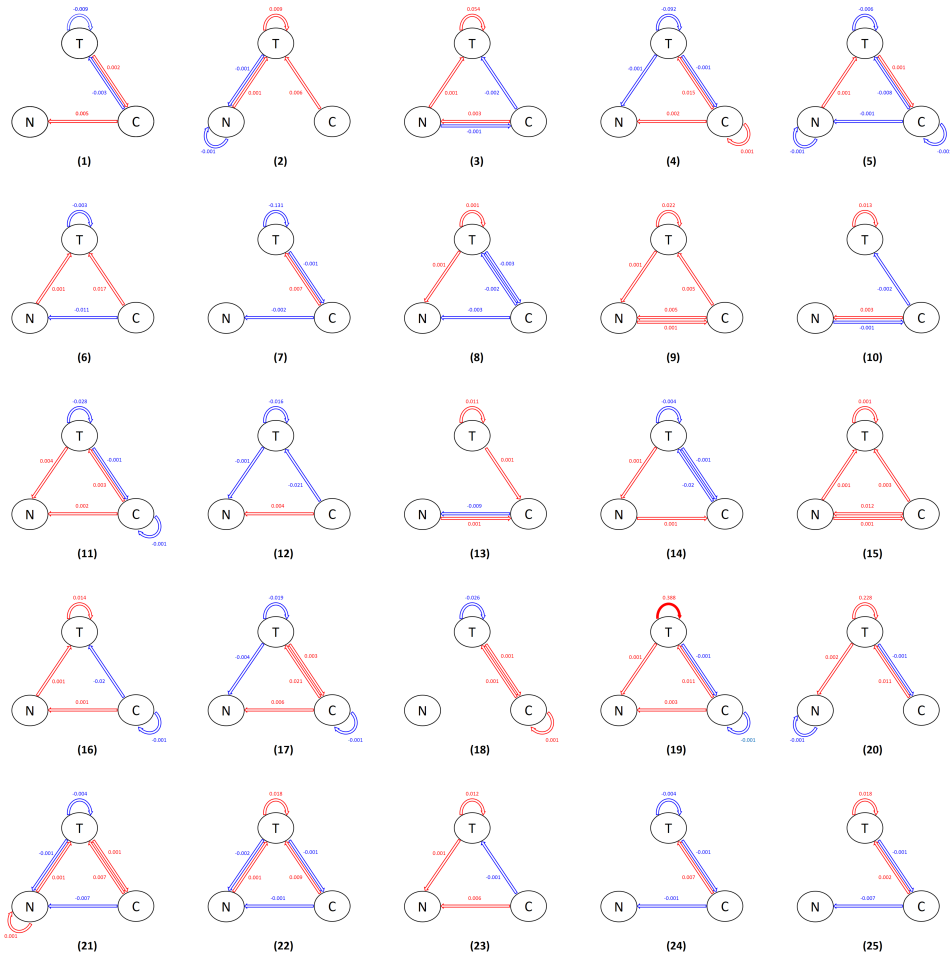
Figure 2.2: Posterior modes of the patient-level deviations. Each panel represents one patient. T=tobacco use, N=negative affect, C=cravings. Red=positive, blue=negative. Solid lines=significant coefficients (red: $\alpha = 0.05$; pink: $\alpha = 0.1$), transparent lines=non-significant coefficients.

Figure 2.4 shows the prediction of the test data along with the 95% intervals for all subjects using the proposed Bayesian model and the maximum-likelihood estimates. The circles represent predictions using the patient-specific VAR models and the triangles represent the Bayesian estimates. The lines going through the circles and the triangles represent the 95% intervals. Red stars show the true test observations. The number of time points in the training data for each subject are shown as well. We see that for the patients who have moderately large
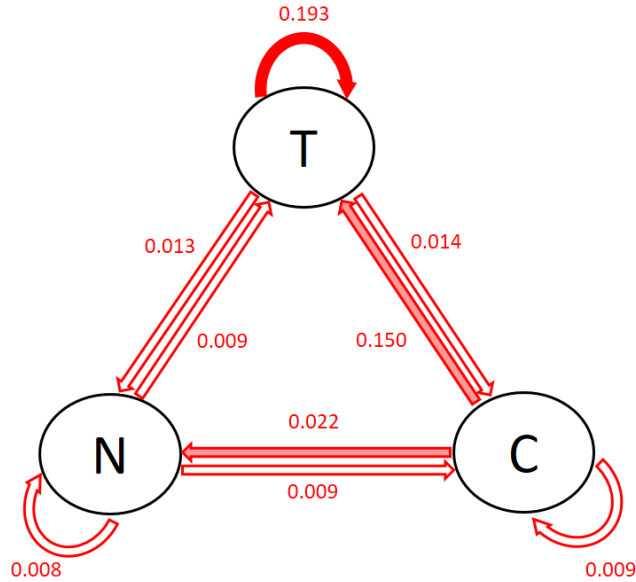
Figure 2.3: Posterior modes of the standard deviation of the patient-level deviations. T=tobacco use, N=negative affect, C=cravings. Red=positive, blue=negative. The first three largest standard deviations are shown in solid arrows.

Table 2.1: Estimated MSE of the hierarchical VAR model and the patient-specific VAR model. % reduction in MSE given by the Bayesian model relative to the patient-specific VAR models are shown. T=tobacco use, N=negative affect, C=craving.

| Model | Overall MSE | MSE by Variable | | |
		T	N	C
Hierarchical VAR	.722	.459	.704	1.002
Patient-specific VAR	.964	.554	.731	1.608
% Reduction	25.1%	17.1%	3.5%	37.7%

number of observations (above 20 time points), the point estimates are quite similar for both methods. However, for the subject with a small number of observations (subject 14 has only 9 training observations), Bayesian prediction is remarkably better than the patient-specific VAR model for cravings.

The estimated coverage probabilities of 95% predictive intervals are, respectively, 97%
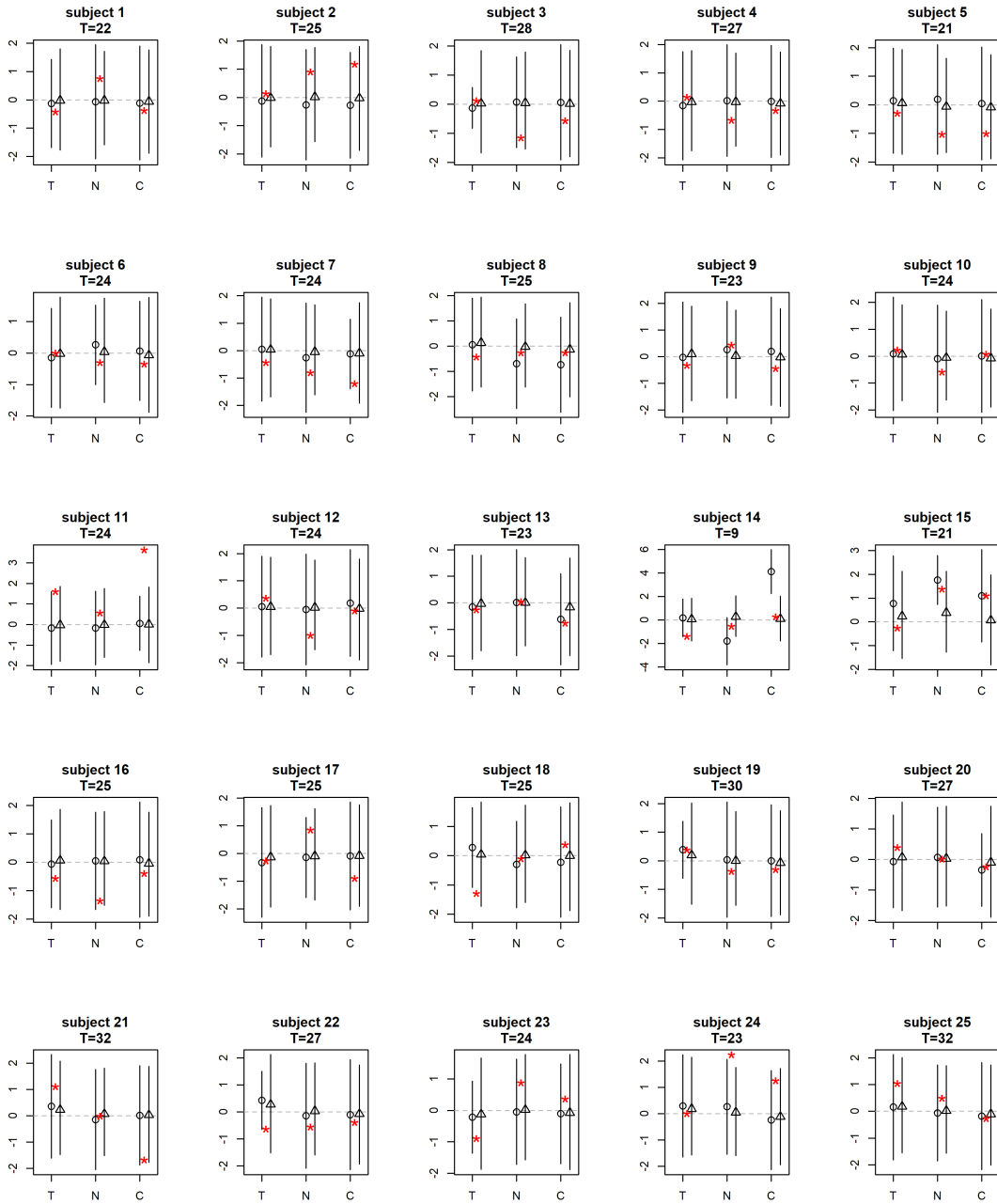
Figure 2.4: Posterior means and the 95% intervals for the 1-step ahead prediction. T=tobacco use, N=negative affect, C=craving. The circles represent the predictions given by the subject-specific MLE's and the triangles represent the Bayesian estimates. The lines show the 95% intervals. Red stars are the true future observations. Number of observations ($T_n, n = 1, \ldots, 25$) for each subject is shown.

and 93% and for the Bayesian model and patient-specific VAR models. For a nominal level of 95%, we would expect about 95% of the predictive intervals to contain the true values. Both

methods are close to the nominal level. However, 68% of the Bayesian intervals are narrower than their patient-specific counterparts. This shows that the Bayesian method can not only provide more accurate point estimates but also a more satisfactory range estimate as well.

2.4 Application 2: predicting functional somatic symptoms (FSS) and psychological discomforts

Approximately 20% to 50% of the physical symptoms presented in primary care and hospital settings cannot be fully explained by organic pathology (Peveler et al. 1997, Fink et al. 1999, de Waal et al. 2004). These symptoms are commonly referred to as functional somatic symptoms (FSS). Previous studies in epidemiology have suggested greater stress (Greene et al. 1985, Robinson et al. 1988, Hatcher and House 2003) and other psychological discomforts (de Benedittis and Lorenzetti 1992, Hesketh et al. 2010) in people who suffer from FSS's than those who do not. Utilizing patient diary databases, a few studies have shown cross-relationship between stress and FSS's in both directions within subjects over time (Dancey et al. 1998, Connelly and Bickel 2011).

To investigate bidirectional relationship and temporal precedence between stress and FSS's within individual patients, van Gils et al. (2014) adopted a VAR modeling framework on data from a diary study with patients who had multiple, persistent FSS's, including external and internal body pain and autonomics (Burton et al. 2009). Granger causality models on stress and a single FSS were performed for each individual patient and each individual FSS. An increase in one or more FSS's were found to be significantly predictable by preceding (mostly first-lag) increase in stress for 30% of the patients, and reverse association in 15% of the patients. Substantial between-patient heterogeneity in the lagged associations between the FSS's and stress was also reported.

2.4.1   Data and the preprocessing

54 patients (age between 21 and 65 years) with persistent FSS's and psychological discomforts were recruited by medical practitioners in a 12-week study between January 2014 and February 2016. Standard handheld personal digital assistant (PDA) computers with the Palm^{TM} operating system are used to record daily diary on 14 FSS's (muscle pain, joint pain, back pain, headache, abdominal pain, pelvic pain, bowel symptoms, dyspepsia, nausea, tight throat, chest pain, weakness, numbness and palpitations) and 5 psychological discomforts (stress, fatigue, anxiety, depression and illness concern). Patients entered 1-3 records on each day, with each data entry consisted of 3 most severe symptoms out of the 14 FSS's all 5 psychological discomfort measures (all patients entered the same 3 symptoms throughout the entire study).

FSS's. Diary questions with respect to FSS's were phrased as follows: "How much have you been bothered by symptom X? Please mark a point on the line between severe symptom X and no symptom X at all." (1-150).

Psychological discomforts. The level of stress and other psychological discomforts were assessed using the following question: "How stressful (fatigue, anxiety-arousing, depressing, illness-concerning) are people and things around you? Please mark a point on the line between very stressful (fatigue, anxious, depressed, illness-concerned) and not stressful (fatigue, anxious, depressed, illness-concerned) at all." (1-150).

Since only 3 out of the 14 FSS's were recorded at each diary entry for each patient, most of the FSS's were recorded by a small number of patients. We therefore use subsets of the data on 4 of the FSS's (headache, joint pain, bowel symptoms, muscle pain) that were recorded by the greatest number of patients (N=12, 12, 10, 15, respectively). For example, 12 patients recorded the symptom headache during the study and are thus included in the first subset. For each of the 4 FSS's (i.e., the 4 subsets), we extend the previous work in van Gils et al. (2014), where the relationship between one FSS and stress was considered, and fit the proposed Bayesian VAR model on the FSS and *all* 5 psychological discomfort variables. This

60

results in 4 models, each of which considers 6 variables. The average (range) length of diary records for the 4 FSS's are, respectively, 90 (64-126), 89 (64-131), 86 (64-109), 90 (83-126) days, with average (range) number of missing daily records 9 (0-37), 9.5 (0-47), 6 (0-24), 4.4 (0-37), respectively.

The preprocessing of the data included: 1) averaging of the variables for each day and each patient to ensure one observation per day per patient, if there are more than one records entered on the same day for the same patient ; 2) missing value imputation using moving average imputation embedded in R package 'impureTS' (`https://cran.r-project.org/web/packages/imputeTS/index.html`); 3) log-transformation of each of the 6 variables to improve normality; 4) standardization to mean 0 and variance 1 and removal of linear trend for each time series for each patient.

We use the same posterior sampling strategy and hyperpriors in Section 2.3 for this application. We use the last 10 observations of each patient as test data and the remainder training data. 1-step ahead forecasting was performed using the same technique in Section 2.3 and h-step ahead forecasting, $h = 2, \ldots, 10$, was performed using the recursive algorithm described in Karlsson (2015). Posterior modes of the parameters are used for statistical inference. Assessment of prediction accuracy (i.e., MSE of the posterior means and coverage probabilities of 95% predictive intervals) and model order selection (i.e., AIC) are the same as Section 2.3.

2.4.2 Results

2.4.2.1 Posterior sampling and order selection

For each of the 4 models, we ran 10,000 iterations across 4 chains and used the second half of the iterations (thinned by every 20 iterations, yielding 250 samples per chain) for posterior inference. All chains mixed well and converged to the stationary distribution. No serial correlations are found in the chains. VAR(1) models outperformed VAR(2) models with

smaller values of AIC across all 4 models (the VAR(1) models for the 4 FSS's have AIC values 5956.6, 6172.1, 4722.8, 7416.8, respectively, and the VAR(2) models 7069.7, 7215.8, 5720.3, 8674.5). We therefore favor VAR(1) models in this application.

2.4.2.2 Population-level coefficients

Figure 2.5 displays the posterior modes of the population-level coefficients for the 4 models (panel (a)-(d) represent FSS=headache, joint pain, bowel symptom and muscle pain, respectively). The coefficient at the r-th row and the c-th column in each panel represents the effect of the c-th variable in the previous day on the r-th variable in the current day. Significant coefficients (at level $\alpha = 0.1$) are in bold font. Purple box on each panel represents the effect of the psychological discomforts in the previous day on the FSS in the current day; blue box represents the reverse effect of FSS in the previous day on the psychological discomforts in the current day.

Inspection of Figure 2.5 indicates that most FSS's and psychological discomfort variables have positive effects on themselves. Most of the cross-relationships between FSS's and psychological discomforts are positive and the magnitudes are smaller than the autocorrelations of the variables. The strongest cross-associations between FSS's and psychological discomforts include the effect of stress in the previous day on bowel symptoms in the current day, the effect of anxiety in the previous day on muscle pain in the current day and the effect of muscle pain in the previous day on illness-concern in the current day. Moreover, there are significant associations between the psychological discomfort variables, including the effect of anxiety on depression for patients who reported headache, fatigue on depression for patients who reported joint pain, stress on anxiety for patients who reported bowel symptoms, and anxiety on fatigue for patients who reported muscle pain.
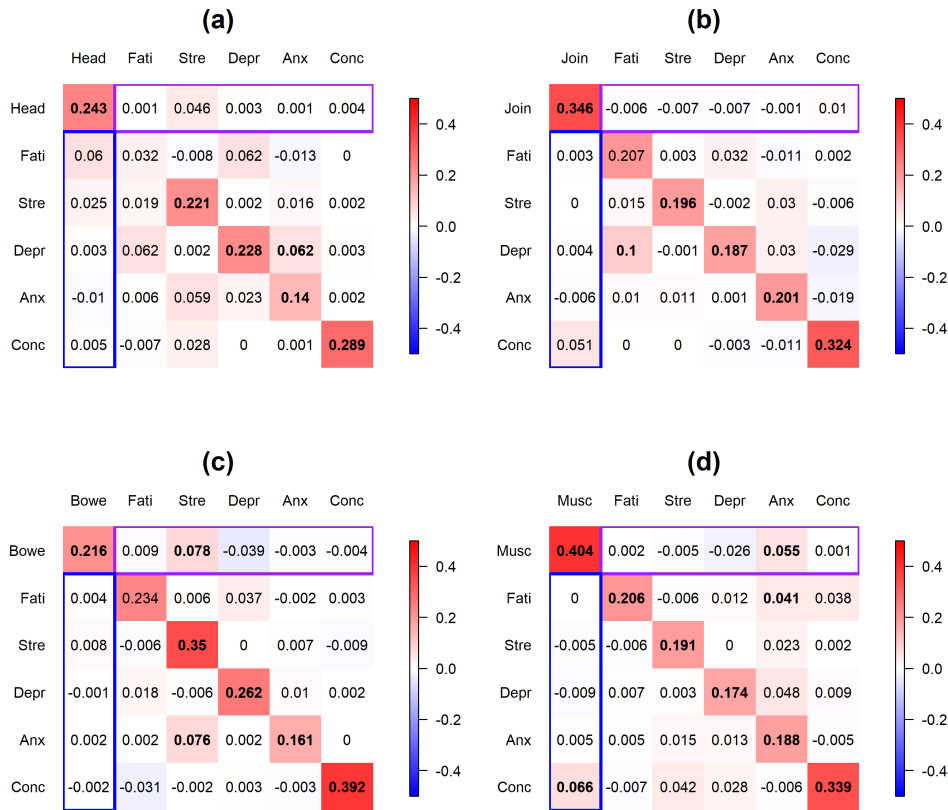
Figure 2.5: Posterior modes of the population-level coefficients for the 4 FSS's. Panel (a)-(d) display FSS=headache, joint pain, bowel symptoms and muscle pain, respectively. The coefficient at the r-th row and the c-th column in each panel represents the effect of the c-th variable in the previous day on the r-th variable in the current day. Significant coefficients (at level $\alpha = 0.1$) are in bold font. Purple box on each panel represents the effect of the psychological discomforts in the previous day on the FSS in the current day; blue box represents the effect of FSS in the previous day on the psychological discomforts in the current day. Head=headache, Join=joint pain, Bowe=bowel symptoms, Musc=muscle pain, Fati=fatigue, Stre=stress, Depr=depression, Anx=anxiety, Conc=illness-concern. Color scale ranges from -0.5 to 0.5

2.4.2.3 Between-patient heterogeneity

Figure 2.6 displays the posterior modes of the standard deviations of the patient-level deviations for the 4 models (Panel (a)-(d) represent FSS=headache, joint pain, bowel symptom and muscle pain, respectively). The highest between-patient heterogeneity lies in the auto-correlation of fatigue. This is consistent across all of the 4 FSS's. For cross-relationships

63

between the FSS's and the psychological discomforts, highest between-patient heterogeneity was present in the effect of headache on illness-concern and depression, the effect of anxiety on headache, the effect of illness-concern and depression on joint pain, and the effect of stress on muscle pain. This is consistent with the previous studies which have found substantial between-subject heterogeneity in the association between stress and FSS's (Burton et al. 2009).

2.4.2.4 Prediction accuracy

Table 2.2 displays the performance of predicting the 10 test data points using the Bayesian VAR model and the conventional patient-specific VAR models for the 4 FSS's. Figure 2.7 displays the h-step forecasting, $h = 1, \ldots, 10$, on headache for all patients who reported this FSS. The black and blue lines display, respectively, the prediction given by the Bayesian model and the patient-specific VAR models. Solid lines represent the point estimates and dashed lines the predictive intervals given by the two methods. Red dots show the true values of the test data points. We can see that both methods performed similarly, with Bayesian model slightly better than the patient-specific VAR models across all 4 FSS's. This is probably due to the fact that the number of observations ($T_n, n = 1, \ldots, N$) for each patient in the training data is much greater than that in Section 2.3, so the Bayesian model utilized much more information contained in the data than in the prior distributions, leading to similar behavior to the maximum-likelihood approach. Nevertheless, the Bayesian model still outperformed the patient-specific models in point estimates due to the pooling among subjects. For the estimated coverage probability, both model tend to slightly overcover (with the Bayesian model slightly more overcovered), but the Bayesian model provides shorter predictive intervals due to the shrinkage imposed by the priors.
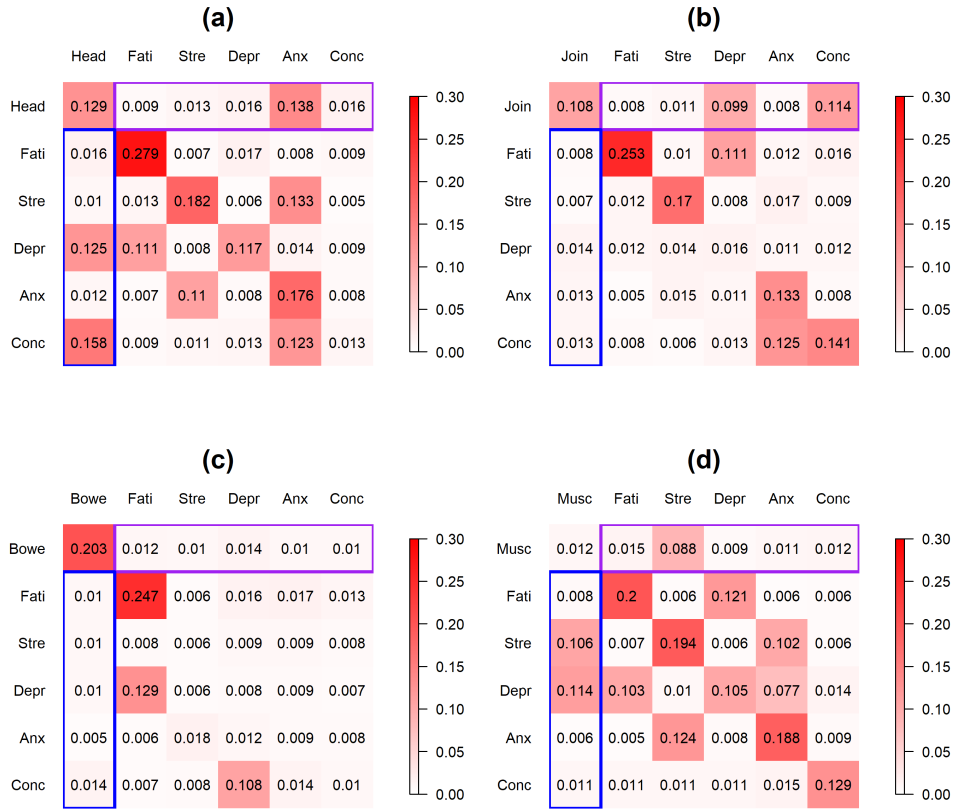
Figure 2.6: Posterior modes of the standard deviations of the patient-level deviations for the 4 FSS's. Panel (a)-(d) display FSS=headache, joint pain, bowel symptoms and muscle pain, respectively. The number at the *r*-th row and the *c*-th column in each panel represents the standard deviation of the effect of the *c*-th variable in the previous day on the *r*-th variable in the current day. Purple box on each panel represents between-patient heterogeneity in the effect of the psychological discomforts in the previous day on the FSS in the current day; blue box represents the between-patient heterogeneity in the effect of FSS in the previous day on the psychological discomforts in the current day. Head=headache, Join=joint pain, Bowe=bowel symptoms, Musc=muscle pain, Fati=fatigue, Stre=stress, Depr=depression, Anx=anxiety, Conc=illness-concern. Color scale ranges from 0 to 0.3.

2.5   Discussion

In summary, we proposed a novel Bayesian hierarchical VAR model for PLPM. Our model simultaneously estimates population-level and patient-level coefficients, as well as estimating between-patient heterogeneity. It provides sparse estimates leading to better interpretability and visualization, as well as higher predictive accuracy.

Table 2.2: Prediction accuracy of the Bayesian VAR model and patient-specific VAR models. Estimated MSE's and coverage probabilities are shown for each single FSS. % reduction in MSE and % shorter predictive intervals given by the Bayesian model relative to the patient-specific VAR models are shown.

	Headache	Joint Pain	Bowel Symptoms	Muscle Pain
			MSE	
Bayesian VAR model	0.743	0.694	0.568	0.698
Subject-specific VAR model	0.744	0.698	0.584	0.704
% Reduction	0.18%	0.58%	2.75%	0.90%
			Coverage probability	
Bayesian VAR model	0.965	0.964	0.97	0.966
Subject-specific VAR model	0.954	0.961	0.968	0.958
% Shorter interval	64%	65%	62%	68%

Application of our model to previous diary studies showed improvement in patient-level prediction in terms of both point and range estimates. The advantage of the partial pooling among subjects in the Bayesian model is more substantial for datasets with a small number of observations for certain patients. When the number of observed time points are sufficiently large for all patients in the data, the Bayesian model behaved similarly to the conventional patient-specific VAR models. Nevertheless, the Bayesian model still outperformed the patient-specific models by providing more information about the population-level coefficients and the between-subject heterogeneity as well as improving prediction accuracy for the future data.

In the examples, we used weakly informative hyperpriors which are conjugate but almost flat on the parameter space. This provides convenient derivation of the conditionals used in the Gibbs sampler while enforcing little prior information. We also tried different values of the hyperparameters but these yielded very similar results.

Further, in the examples, we used posterior modes as parameter estimates for better interprebility and visualization. However, we also tried posterior median. Due to the fact that the posterior distributions are symmetric and unimodal for most of the parameters, the results were very similar in terms of both parameter estimation and future data prediction, although sparsity in the parameter estimation then no longer exists.
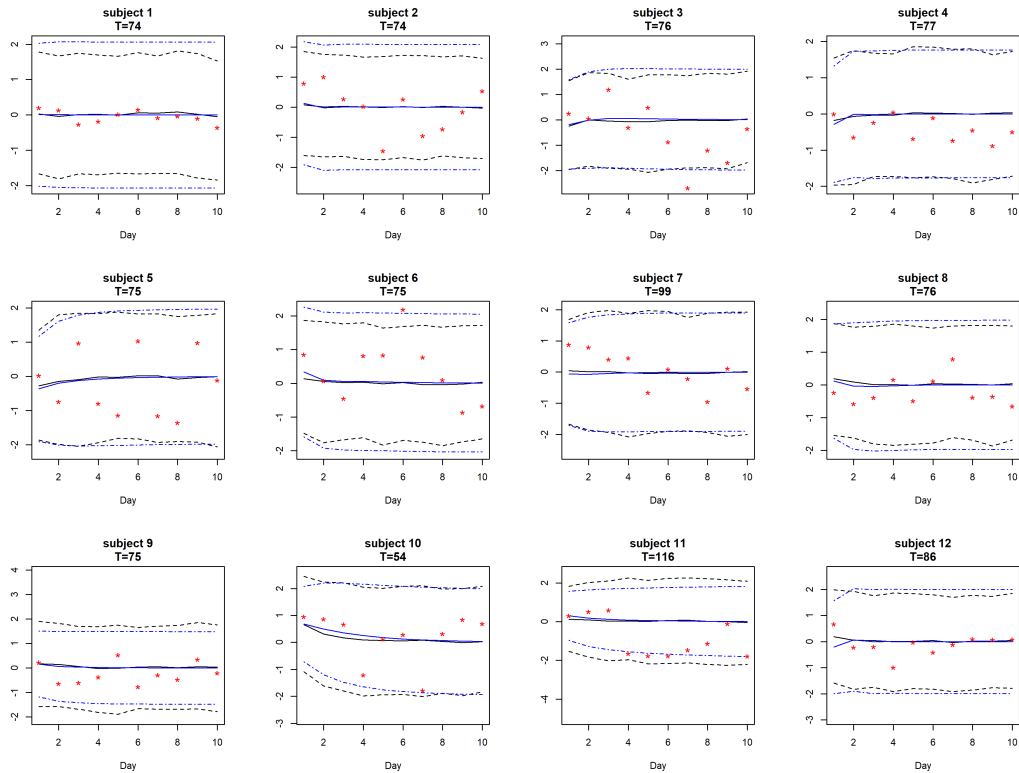
Figure 2.7: Posterior means and 95% intervals for predicting headache for all patients who reported this FSS using the Bayesian VAR model and patient-specific VAR models. Each panel represents one patient (the index shows the h-step forecasting, $h = 1, \ldots, 10$). The black and blue lines display, respectively, the prediction given by the Bayesian model and the patient-specific VAR models. Solid lines represent the point estimates and dashed line the predictive intervals given by the two methods. Red dots show the true values of the test data points. Number of observations (T) for each patient is displayed in each panel.

Finally, in both applications, the prediction of future data is near 0, indicating that neither models suitably describes these data. One possible remedy, to be explored in future work, is the inclusion of covariates such as demographic variables and/or time information such as weekdays versus weekends, which might lead to more precise predictions of future observations.

data were originally presented in Zheng et al. (2013) and van Gils et al. (2014).

Appendix B

Appendix to Chapter 2

B.1 Full conditional distributions

To draw from posterior distribution using Gibbs sampler, one draws each parameter from its conditional distribution given all other parameters at their current values. Let $g(\cdot)$ denote conditional distribution of one parameter given others. We now give the full conditionals of all parameters.

1. $g(\boldsymbol{w}) = \text{MVN}\left(M_w,\ S_w^2\right)$, where

$$S_w^2 = \left[\sum_{n=1}^{N}\left(H_n H_n^T\right) \otimes \Omega + D\right]^{-1},$$

$$M_w = S_w^2\left[\sum_{n=1}^{N}\left(H_n H_n^T\right) \otimes \Omega\left(\hat{\boldsymbol{w}}_n - \boldsymbol{\alpha} * \boldsymbol{v}_n\right)\right],$$

where $\hat{\boldsymbol{w}}_n$ is the MLE of the VAR coefficients for the n-th patient;

2. $g(\boldsymbol{v}_n) = \text{MVN}\left(M_{v_n},\ S_{v_n}^2\right), n = 1, ..., N$, where

$$S_{v_n}^2 = \left[\text{Diag}(\boldsymbol{\alpha})\left(\left(H_n H_n^T\right) \otimes \Omega\right)\text{Diag}(\boldsymbol{\alpha}) + \Theta_v\right]^{-1},$$

$$M_{v_n} = S_{v_n}^2\left[\text{Diag}(\boldsymbol{\alpha})\left(\left(H_n H_n^T\right) \otimes \Omega\right)\left(\hat{\boldsymbol{w}}_n - \boldsymbol{w}\right)\right],$$

where $\text{Diag}(\cdot)$ represents assigning the $R^2 p$ vector on a diagonal matrix;

3. $g(\boldsymbol{\alpha}) = \text{MVN}\left(M_\alpha,\ S_\alpha^2\right)$, where

$$S_\alpha^2 = \left[\sum_{n=1}^{N}\text{Diag}(\boldsymbol{v}_n)\left(\left(H_n H_n^T\right) \otimes \Omega\right)\text{Diag}(\boldsymbol{v}_n) + aI\right]^{-1},$$

$$M_\alpha = S_\alpha^2\left[\sum_{n=1}^{N}\text{Diag}(\boldsymbol{v}_n)\left(\left(H_n H_n^T\right) \otimes \Omega\right)\left(\hat{\boldsymbol{w}}_n - \boldsymbol{w}\right)\right];$$

4. $g(\theta_{vk}) = \Gamma(k_v, s_{vk}), k = 1, ..., R^2 p$, where

$$k_v = \frac{N}{2} + k$$

$$s_{vk} = \left(\frac{1}{2}\sum_{n=1}^{N} v_{nk}^2 + \frac{1}{s}\right)^{-1},$$

where v_{nk} is the k-th element of $\boldsymbol{v}_n$;

5. $g(\lambda_{1,k}^2) = \Gamma\left(\mu_{\lambda_{1,k}^2}, \nu_{\lambda_{1,k}^2}\right), k = 1, ..., R^2 p$, where

$$\nu_{\lambda_{1,k}^2} = \nu_1 + 2,$$

$$\mu_{\lambda_{1,k}^2} = \frac{(\nu_1 + 2)(\mu_1 \xi_k^2)}{\nu_1 \xi_k^2 + 2\tau_k^2 \mu_1};$$

6. $g(\lambda_{2,k}) = \Gamma\left(\mu_{\lambda_{2,k}}, \nu_{\lambda_{2,k}}\right), k = 1, ..., B^2 p$, where

$$\nu_{\lambda_{2,k}} = \nu_2 + 1,$$

$$\mu_{\lambda_{2,k}} = \frac{(\nu_2 + 1)\mu_2}{\nu_2 + w_k^2 \mu_2},$$

where w_k is the k-th element of w;

7. $g(\Omega) = \text{Wishart}(S_\Omega, \nu_\Omega)$, where

$$\nu_\Omega = \sum_{n=1}^{N} (T_n - p) + \nu + 2Rp,$$

$$S_\Omega^{-1} = \left[\sum_{n=1}^{N} S_n + K^{-1} + 2QQ^T\right]^{-1},$$

where

$$S_n = (Y_n - W_n H_n)(Y_n - W_n H_n)^T$$

is the residual matrix of the n-th patient,

$$Y_n = \left(y_{n,p+1} \quad \cdots \quad y_{n,T_n}\right)$$

is a $R \times (T_n - p)$ matrix,

$$W_n = vec^{-1}(w + \alpha * v_n),$$

where $vec^{-1}(\cdot)$ means stacking the $R^2 p$-coefficient vector back into a $R \times Rp$ matrix,

and

$$Q = \begin{pmatrix} \gamma_1 & \cdots & \gamma_{(Rp-1)R+1} \\ \vdots & \ddots & \vdots \\ \gamma_R & \cdots & \gamma_{R^2 p} \end{pmatrix}$$

is a $R \times Rp$ matrix composed of γ_k, $k = 1, ..., R^2 p$, which depend on $2\tau_k^2$, $k = 1, ..., R^2 p$ and Ω through

$$\tau_{R^2 p} \lambda_{1,R^2 p} = \gamma_{R^2 p},$$

$$\tau_k \lambda_{1,k}^2 = \gamma_k + \boldsymbol{m}_k \left(\gamma_{k+1}, ..., \gamma_{B^2 p} \right)^T$$

for $k = 1, ..., R^2 p - 1$, where

$$\boldsymbol{m}_k = \left(M_{k,k+1}, ..., M_{k,R^2 p} \right) \begin{pmatrix} M_{k+1,k+1} & \cdots & M_{k+1,R^2 p} \\ \vdots & \ddots & \vdots \\ M_{R^2 p,k+1} & \cdots & M_{R^2 p,R^2 p} \end{pmatrix}^{-1} ;$$

8. $g(\frac{1}{2\tau_k^2}) = \sqrt{\frac{b_k}{2\pi}} \left(\frac{1}{2\tau_k^2} \right)^{-\frac{3}{2}} \exp \left\{ -\frac{b_k \left(\frac{1}{2\tau_k^2} - a_k \right)^2}{2 a_k^2 \frac{1}{2\tau_k^2}} \right\}$, $k = 1, ..., R^2 p$, where

$$a_k = \sqrt{\frac{\lambda_{1,k}^2}{w_k^2 \xi_k^2}},$$

$$b_k = \frac{\lambda_{1,k}^2}{\xi_k^2}.$$

Note that the conditional distribution of $2\tau_k^2$ is complicated, but Park and Casella (2008) have shown that $g(\frac{1}{2\tau_k^2})$ is Inverse-Gaussian distribution. Therefore, we draw $\frac{1}{2\tau_k^2}$ first from $g(\frac{1}{2\tau_k^2})$ and then transform back to get $2\tau_k^2$.

Chapter 3

A Bayesian hierarchical model for repeat Electrical Stimulation Mappings

3.1 Introduction

Electrical stimulation mapping (ESM) has become a widely used technique to determine localization of brain areas responsible for certain neurophysiological functions, called functional localization, in an individual patient's cortex. Localization of cortex sites related to language commonly utilizes mapping during an object naming task (Penfield and Roberts 1959, Ojemann et al. 1989). Sites where stimulation interferes with naming are often quite focal in regions of frontal and temporal cortex, but with substantial variability between patients in their exact location. It has been shown that those sites are crucial for language, in that a resection that encroaches on them is likely to be associated with a postoperative language deficit, while one that does not is unlikely to have a deficit (Ojemann 1983b, Haglund et al. 1994). The extent to which these crucial sites remain in the same location over time in an individual adult patient is controversial. Individual cases have been described that seem to show evidence for substantial variability (Duffau et al. 2002, Seek et al. 2006, Duffau 2011), and for apparent stability (Ojemann 1983a, Ojemann et al. 2003, Robles et al. 2008, Serafini et al. 2013). In general, studies in patients with intrinsic gliomas have more often shown changes interpreted as evidence of variability, those in patients with epilepsy, changes interpreted as evidence for stability, but there are exceptions in both disease groups. More importantly, previous studies all utilized qualitative approaches (such as simply comparing sites identified during repeat mappings graphically) and none of them has examined the variability in these locations systematically via quantitative measures.

Mapping language sites quantitatively over time presents two particular challenges. First, within a given patient, different numbers of sites might be mapped across repeat mappings. Figure 3.1(a) shows an example of a patient that illustrates this issue. For this patient, five sites are observed during the first mapping but only one site is observed during the second mapping. A second challenge concerns the matching of the sites across mappings. Consider the patient represented in Figure 3.1(b). While we expect sites to exhibit some spatial stability
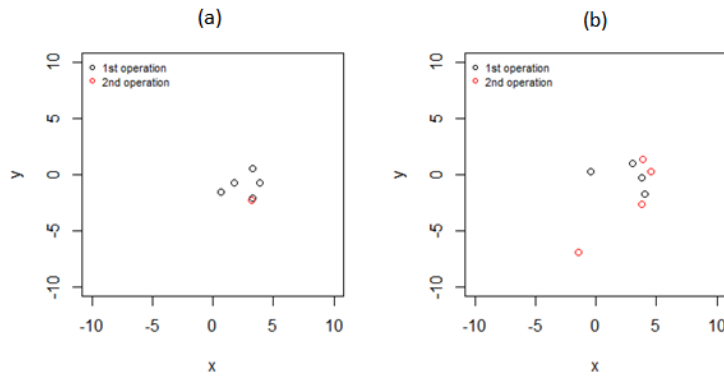
Figure 3.1: An example to illustrate the challenges in modeling repeat Electrical Simulation Mappings. Panel (a): unequal number of sites mapped during the two mappings; Panel (b): matching between the sites.

over time, the site from the second mapping in the lower left quadrant could "belong" to any of the four sites from the first mapping.

We propose a Bayesian model to address these two challenges and provide assessment of variability in language localization quantitatively. Specifically, we assume that the discrepancy in numbers of sites mapped during the 2 mappings is due to missing observations. Take the example in Figure 3.1(a), this means that we assume that the five language sites do exist and treat the four absent locations in the second mapping as missing data. We use augmented Markov Chain Monte Carlo (MCMC) to impute these missing values. Moreover, we model the matching between the sites probabilistically, favoring solutions that respect spatial stability, but incorporating uncertainty in the matching process. We also assume that language sites do not "jump" across different regions in the brain (e.g. from temporal lobe to frontal or parietal lobe), a common assumption in neurophysiology. We adopt a hierarchical model to analyze the variability in the localization across different individual regions. We apply our model to a series of patients with medically refractory epilepsy, who had serial mapping of language with a standardized object naming task over intervals of 1 to more than 20 years, assessing changes in perisylvian localization in language dominant hemisphere over time. These findings provide some further insight into what differentiates these crucial areas from

surrounding cortex, and on the degree of cortical plasticity in the adults.

We proceed as follows. Section 3.2 describe the Bayesian method and the augmented MCMC procedure to impute missing sites and obtain probabilistically optimal matching between the sites. Section 3.3 describes the real data analysis using the proposed approach and provides visualization of the results. Section 3.4 concludes.

3.2 Method

3.2.1 A regional hierarchical model

We propose a Bayesian hierarchical model motivated by the challenges outlined in Section 3.1. Suppose the sites are observed across R regions, where each region is a particular section in the cortex. Denote by $(X_{irj}^s, Y_{irj}^s)^T$ the xy-coordinates of mapped site $j = 1, ..., n_{ir}^s$ for patient $i = 1, ..., N$ at region $r = 1, ..., R$ during mapping $s = 1, 2$, where n_{ir}^s is the total number of sites mapped for patient i at region r during mapping s. We assume that sites arise from independent bivariate normal distributions given the constant true latent location corresponding to each site and the common regional variance:

$$\begin{pmatrix} X_{irj}^s \\ Y_{irj}^s \end{pmatrix} \sim \text{BN}\left(\begin{pmatrix} \mu_{X_{irj}} \\ \mu_{Y_{irj}} \end{pmatrix}, \sigma_r^2 I_2 \right), \tag{3.1}$$

where $(\mu_{X_{irj}}, \mu_{Y_{irj}})^T$ is the mean vector for mapped site j of patient i at region r, and $\sigma_r > 0$ is the standard deviation for region r, assumed to follow an i.i.d. log-normal distribution with hyperparameters a and $b, b > 0$:

$$\sigma_r \sim \text{Lognormal}(a, b^2), r = 1, \ldots, R. \tag{3.2}$$

To address the challenges of unequal number of mapped sites during the two mappings and the matching between the sites, we introduce two sets of augmented variables in the model. Denote $\boldsymbol{M}_{ir} = (M_{ir1}, \ldots, M_{irn_{ir}})^T$ the matching vector of the sites (observed or missing)

across the two mappings for patient i at region r, where $M_{irj} = k$ if the j-th site in the first mapping matches to the k-th site in the second mapping and $n_{ir} = \max(n_{ir}^1, n_{ir}^2)$ is the maximum number of sites mapped over the two mappings. If a site $(X_{irj}^s, Y_{irj}^s)^T$ is missing, we denote the corresponding unobserved value by $\boldsymbol{Z}_{irj} = (Z_{X,irj}, Z_{Y,irj})^T$ (note that we suppress the superscript s here because there can be at most one missing value occurring at either mapping). Table 3.1 shows a synthetic example of mapped sites for a particular patient within a particular region, where 4 sites are mapped at the first mapping (column 1) and only 2 are mapped at the second mapping (column 2). Column 3 lists an example matching between the 4 sites.

Table 3.1: A synthetic example of sites mapped for patient i within region r during the 2 mappings. Column 1: 4 observed sites at the first mapping; Column 2: 2 observed sites and 2 missing sites at the second mapping; Column 3: an example matching vector between the 4 sites; Column 4: description of the matching.

Mapping 1	Mapping 2	$\boldsymbol{M}_{ir}$	Description
$(X_{ir1}^1, Y_{ir1}^1)^T$	$(X_{ir1}^2, Y_{ir1}^2)^T$	3	$(X_{ir1}^1, Y_{ir1}^1)^T$ matches $\boldsymbol{Z}_{ir3}$
$(X_{ir2}^1, Y_{ir2}^1)^T$	$(X_{ir2}^2, Y_{ir2}^2)^T$	4	$(X_{ir2}^1, Y_{ir2}^1)^T$ matches $\boldsymbol{Z}_{ir4}$
$(X_{ir3}^1, Y_{ir3}^1)^T$	$\boldsymbol{Z}_{ir3}$	1	$(X_{ir3}^1, Y_{ir3}^1)^T$ matches $(X_{ir1}^2, Y_{ir1}^2)^T$
$(X_{ir4}^1, Y_{ir4}^1)^T$	$\boldsymbol{Z}_{ir4}$	2	$(X_{ir4}^1, Y_{ir4}^1)^T$ matches $(X_{ir2}^2, Y_{ir2}^2)^T$

We assume non-informative priors for the mean vectors $(\mu_{X_{irj}}, \mu_{Y_{irj}})^T$, the matching vector $\boldsymbol{M}_{ir}$, the missing value $\boldsymbol{Z}_{irj}$ and the hyperparameters a and b. Note that in this model, the variance parameters $\sigma_r, r = 1, \ldots, 4$, and the exponentiated hyperparameter e^a are the primary quantities of interest as they quantify the variability in localization of the language sites within and across the regions mapped during the 2 mappings.

3.2.2 Gibbs sampler with Metropolis-Hasting updates

To sample from the posterior distribution, we implement the Gibbs sampler algorithm with Metropolis-Hasting updates for each parameter (Gelman et al. 2014). Specifically, for each

iteration $t = 1, \ldots, T$, we alternately update the parameters one at a time by proposing new values from the following jumping distributions:

1. draw $\sigma_1^{(t)} \sim \text{TN}_0(\sigma_1^{(t-1)}, 0.4^2)$;

2. draw $\sigma_r^{(t)} \sim \text{TN}_0(\sigma_r^{(t-1)}, 0.2^2), r = 2, \ldots, 4$;

3. draw $\mu_{X_{irj}}^{(t)} \sim \text{U}(\mu_{X_{irj}}^{(t-1)} - 4, \mu_{X_{irj}}^{(t-1)} + 4)$;

4. draw $\mu_{Y_{irj}}^{(t)} \sim \text{U}(\mu_{Y_{irj}}^{(t-1)} - 4, \mu_{Y_{irj}}^{(t-1)} + 4)$;

5. draw $Z_{X_{irj}}^{(t)} \sim \text{U}(Z_{X_{irj}}^{(t-1)} - 6, Z_{X_{irj}}^{(t-1)} + 6)$;

6. draw $Z_{Y_{irj}}^{(t)} \sim \text{U}(Z_{Y_{irj}}^{(t-1)} - 6, Z_{Y_{irj}}^{(t-1)} + 6)$;

7. draw $a^{(t)} \sim \text{U}(a^{(t-1)} - 2, a^{(t-1)} + 2)$;

8. draw $b^{(t)} \sim \text{TN}_0(b^{(t-1)}, 2^2)$;

9. draw $\boldsymbol{M}_{ir}^{(t)}$ uniformly from all possible permutations of $\boldsymbol{M}_{ir}$,

where $\text{TN}_0(\cdot)$ stands for a truncated normal distribution on $(0, \infty)$, $\text{U}(\cdot)$ represents the uniform distribution, and the subscript $(t - 1)$ represents the current value of a particular parameter at iteration $t-1$. The jumping sizes of these distributions are chosen such that the acceptance rate is about 0.2, which is believed to be a decent acceptance rate for high-dimensional problems (Gelman et al. 2014). We drew 4 chains using 2 million iterations for each chain, with the first half used as warm-up iterations. We then thinned the remaining 1 million to 200 samples per chain, ending up with 800 posterior samples altogether. All chains mixed well using the criteria of $\hat{R} < 1.1$ (Gelman et al. 2014) and have no serial correlation. The entire posterior sampling procedure is implemented in R.

Notice that within all possible permutations of $\boldsymbol{M}_{ir}$, some of them are redundant due to matching observed sites to missing sites. Taking the example in Table 3.1, two possible permutations are $\boldsymbol{M}_{ir} = (3, 4, 1, 2)$ and $\boldsymbol{M}_{ir} = (4, 3, 1, 2)$. Including both permutations in the pool of proposing values will create a mixture of the posterior samples for the imputed values and cause mixing problem of the chains. We therefore treat these two permutations equivalent and only allow the proposing values for matchings involving missing sites to be in increasing

order. Using, again, the example in Table 3.1, this means that we only keep the permutation $M_{ir} = (3, 4, 1, 2)$ in the proposing pool and eliminate $M_{ir} = (4, 3, 1, 2)$ because the first two values in the former vector (i.e., (3,4)) involve the missing sites and are in increasing order.

3.3 Application

3.3.1 Electrical Stimulation Mapping

We now apply our model to a real dataset to investigate the variability of language sites over time. The data consist of 22 patients, all of whom have received two resections in the language dominant hemisphere, 21 in left brain, 1 in right. Mean age at first operation was 27.7 years (range 10-39). The second operation occurred a mean of 8.35 years (range 1-20.25) after the first. Electrical Stimulation Mappings of language had occurred at both operations, utilizing a standard measure of object naming task (Ojemann et al. 1989). In the entire series, a total number of 246 sites common to the areas of mapping in both operations were mapped, of which 99 sites have positive stimulation interfered with naming and are included in here. These sites are located across 4 brain regions, including anterior temporal, posterior temporal, frontal and parietal, with the number of sites mapped within a certain region during one mapping ranging from 1 to 4. 20 of the patients have at least one site missing during either of the two mappings. Certain sites mapped during the two mapping are closer to each other than the other sites. Figure 3.2 shows the mapped sites for two patients during the two mappings. Yellow and white marks show, respectively, sites mapped at the first and second mapping; blues circles have radius of 1cm in Panel (a) and 5 and 10cm in Panel (b).

3.3.2 Stability in localization of language sites

Figure 3.3 shows the posterior sampling distributions for the variance parameters $\sigma_r, r = 1, \ldots, 4$, and the hyperparameters a and b. Red solid lines and blue dashed lines show,
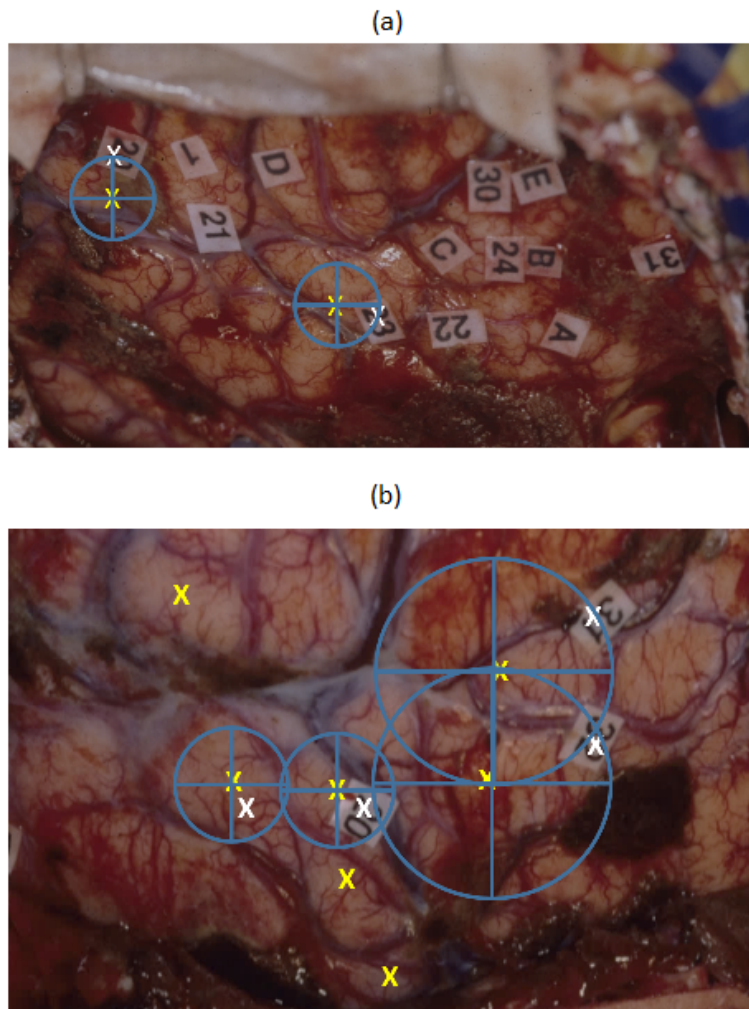
Figure 3.2: Sites mapped for two patients at two Electrical Stimulation Mappings. Yellow and white marks show, respectively, sites mapped at the first and second mapping; blues circles have radius of 1cm in Panel (a) and 5 and 10cm in Panel (b).

respectively, the posterior medians and 95% posterior intervals based on equal-tailed posterior quantiles. These numbers are also shown in Table 3.2 (Column 2). One can see that, the posterior medians are around 0.6cm and the 95% posterior intervals have approximately a range of (0.4,0.9) across all 4 regions. The posterior distribution of σ_1 in the anterior temporal region is slightly more variable and heavy-tailed than the other three regions. The posterior median for e^a is, again, approximately 0.6cm, and the 95% posterior interval has a range of (0.3,1.1). This suggests substantial stability in the localization of language sites within each
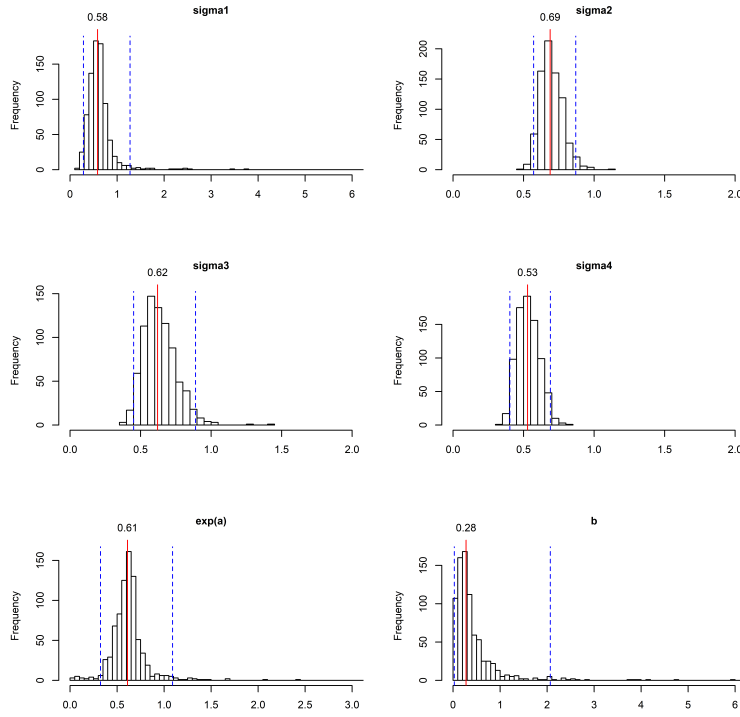
Figure 3.3: Posterior sampling distributions for $\sigma_r, r = 1, \ldots, 4$, and the hyperparameters a and b (e^a is shown for the hyperparameter a). Red solid lines display posterior medians and blue dashed lines display the 95% posterior intervals based on equal-tailed posterior quantiles. The range in the panels is 0-6 for σ_1 and 0-2 for $\sigma_r, r = 2, 3, 4$.

cortex region mapped over the entire study period.

Table 3.2: Posterior median and 95% intervals based on equal-tailed quantiles (shown in brackets) for $\sigma_r, r = 1, \ldots, 4$, and the hyperparameters a and b. Column 2: pooled sample; Column 3: long elapse group; Column 4: short elapse group.

Parameter	All patients	Long elapse group	Short elapse group
σ_1	0.59 (0.29,1.22)	0.65 (0.17,5.5)	0.57 (0.27,1.23)
σ_2	0.69 (0.57,0.87)	0.72 (0.53,1.03)	0.67 (0.52,0.89)
σ_3	0.62 (0.45,0.91)	0.59 (0.27,1.45)	0.63 (0.44,1.03)
σ_4	0.53 (0.4,0.69)	0.58 (0.43,0.81)	0.46 (0.29,0.73)
e^a	0.61 (0.32,1.12)	0.64 (0.17,2.14)	0.58 (0.21,1.17)
b	0.27 (0.02,1.95)	0.41 (0.03,3.43)	0.35 (0.02,2.43)

Figure 3.4 show, respectively, the posterior sampling distributions for the same parameters as in Figure 3.3 for patients who have short (< 7 years) and long (> 7 years) elapses between
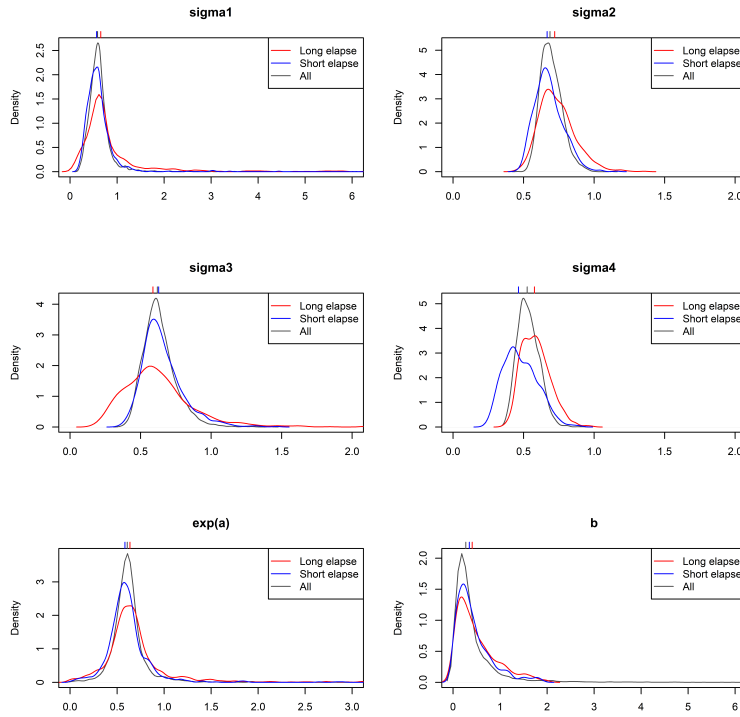
Figure 3.4: Posterior sampling distributions for $\sigma_r, r = 1, \ldots, 4$ and the hyperparameters a and b for different elapse groups. Red lines: patients who had long elapse (> 7 years) between the two mappings; blue lines: patients who had short elapse (< 7 years) between the two mappings; grey line: pooled sample. Posterior medians for the three groups of patients are shown on top of each panel.

the 2 mappings, represented by blue and red lines, respectively. Each group consists of 11 patients. Grey lines show the posterior distribution for the pooled sample. Posterior medians for the 3 groups of patients are shown on top of each plot with the corresponding colors. Table 3.2 shows the posterior median and the 95% intervals for the two groups. We can see that the posterior medians of all three groups are around 0.6cm and the upper bound of 95% intervals around 2cm, confirming that the localization of languages sites are generally stable over time. Nevertheless, the long elapse group resulted in higher posterior medians and wider posterior intervals than the short elapse group for most of the regions. This suggests that over a very long period of time, the movement in language locations does gradually increase in distance and spread.

82

3.3.3 Visualization of imputed missing sites and matching

Figure 3.5-3.26 visualize the imputed missing sites and the matching between the sites mapped during the two mappings for each patient. Here, the colors of the dots represent the regions: grey for anterior temporal, red for posterior temporal, blue for parietal and green for frontal. Round and triangle dots show, respectively, the sites mapped during the first and the second mapping. The solid dots represent observed sites and the transparent dots represent the posterior medians of the imputed values for the missing sites. The imputed values of the missing sites are shown in the clouds behind the corresponding posterior medians. The legends on the top-left corner show the matching between the sites (i.e., $M_{ir}, i = 1, \ldots, 22, r = 1, \ldots, 4$) that appeared most frequently in the posterior samples. One can see that the posterior medians of the imputed sites are very close to the matched observed sites. Nevertheless, the clouds behind the posterior medians suggest some uncertainty in the missing data imputation. For example, for subject 9, the 39-th site is an unobserved site. The posterior median of the imputed values for this site overlaps the matched observed site and the cloud behind it forms a circle centered at the posterior median. Another unobserved site for the same subject (the 42-th site), however, has a shift from the matched observed site. The majority of the cloud centered at the matched observed site with a few exceptions behind other observed sites (i.e., the 40-th and 41-th sites). This reveals how the uncertainty in imputation is affected by all observed sites that could be potential matches with the unobserved site. Moreover, the way the sites matching to each other is intuitive: the sites mapped in the first mapping are most likely to match the closest sites mapped in the second mapping.

3.4 Conclusion

We propose a Bayesian hierarchical model with augmented MCMC posterior sampling procedure to evaluate the variability of language localization for patients with resections in the

language dominant cortex over a wide range of time periods. Our model is capable in handling unequal numbers of sites mapped during the mappings and provides matching between the sites using a probabilistic approach. Our model suggests substantial stability in the location of language sites identified as essential for object naming with Electrical Stimulation Mapping in patients undergoing repeat epilepsy surgery over a mean interval of 8.35 years with a range of 1 to over 20 years. The average variability in these language sites are within 1cm and consistent over different regions in the brain. This finding is similar to the subjectively stable localization previously reported for adolescent epilepsy patients undergoing repeat mapping (Ojemann et al. 2003, Serafini et al. 2013). Our analysis also demonstrates that the variability in language localization is lower for patients who had an elapse less than 7 years than those greater than, suggesting that change in the language localization gradually increases in distance and spread as the elapse between the two consecutive mappings increases. Further, by adding one more set of the matching and missing variables, the model is readily extended to the case where certain subjects have mappings on more than two occasions .

One additional thing to notice from the posterior sampling of the matching is that for patient 21, both sites 91 and 92 in the first mapping match to site 92 in the second mapping with highest frequency, and for patient 22, both sites 95 and 96 in the first mapping are most likely to match to site 96 in the second mapping. This is due to the (almost) symmetric geometric pattern of these sites. For example, for patient 22, there is an equal chance for site 95 and 96 in the first mapping to match site 94 and 96, or 96 and 94 in the mapping. However, when one of them in the first mapping does not match to either 94 or 96, the other always has a higher chance to match to 96 than to 94, and so both of them have higher chance to match to 96 than the other sites. This is the same for patient 21. In order to solve this problem and produce a better visualization of the matching, we eliminated the possibility of site 92 in the first mapping to match site 91 in the second mapping, and site 95 in the fist mapping to match site 96 in the second mapping. By doing so, all sites matched to unique corresponding sites.
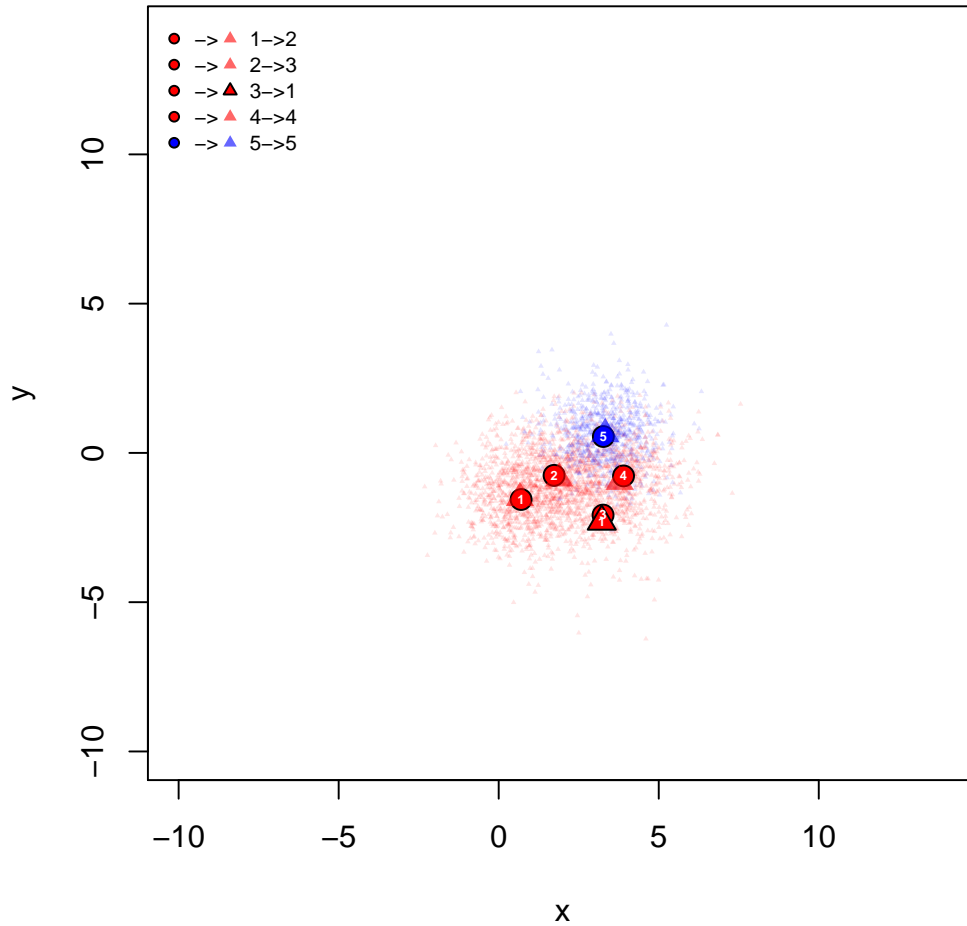
Figure 3.5: Visualization of imputed values and matching for Patient 1. The colors of the dots represent the regions: grey for anterior temporal, red for posterior temporal, blue for parietal and green for frontal. Round and triangle dots represent, respectively, the sites mapped during the first and second mapping. The solid dots represent observed sites and the transparent dots represent the posterior medians of the imputed values for the missing sites. The imputed values of the missing sites are shown in the clouds behind the corresponding posterior medians.

Eliminating these possibilities does not change the posterior distributions of $\sigma_r, r = 1, \ldots, 4$, a and b.
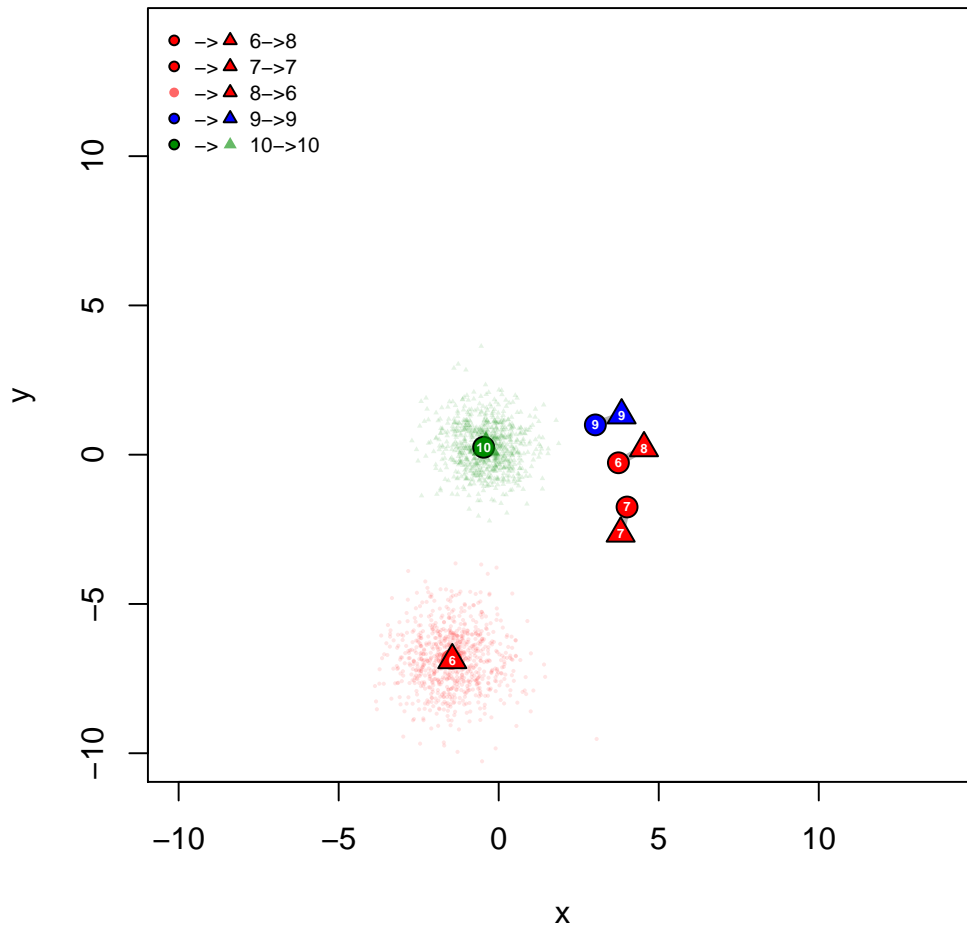
Figure 3.6: Visualization of imputed values and matching for Patient 2.
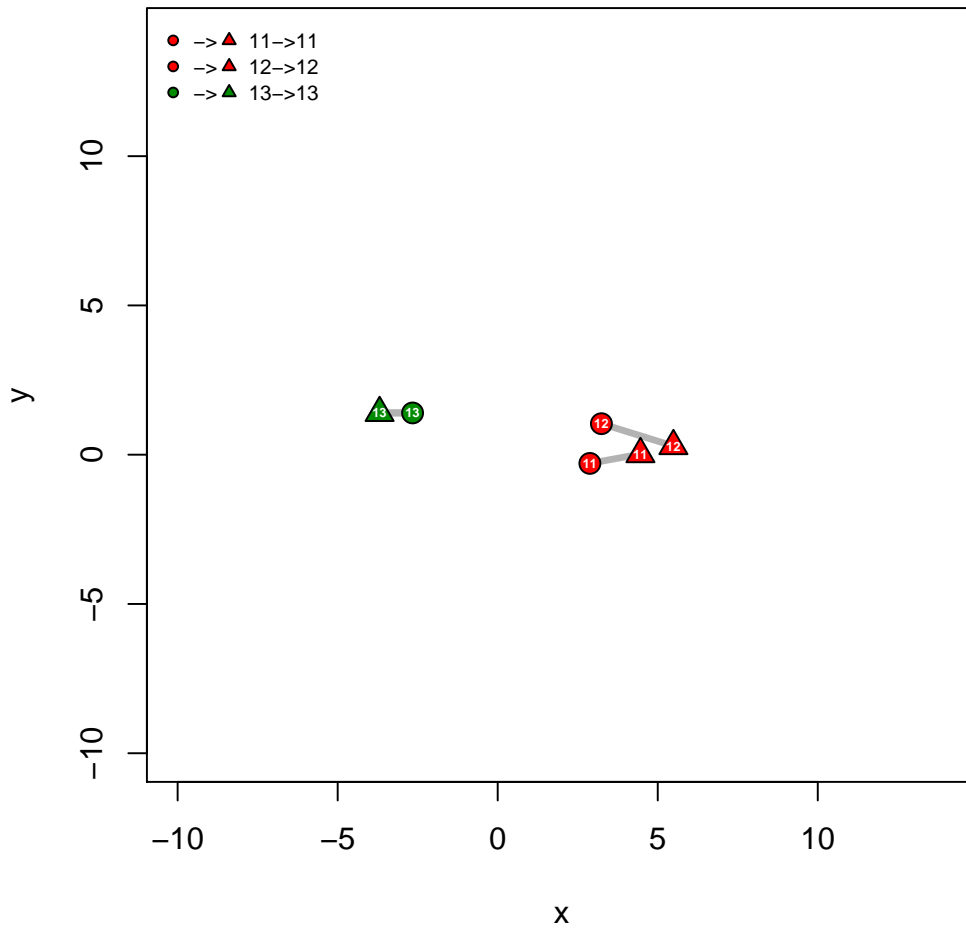
ID = 3

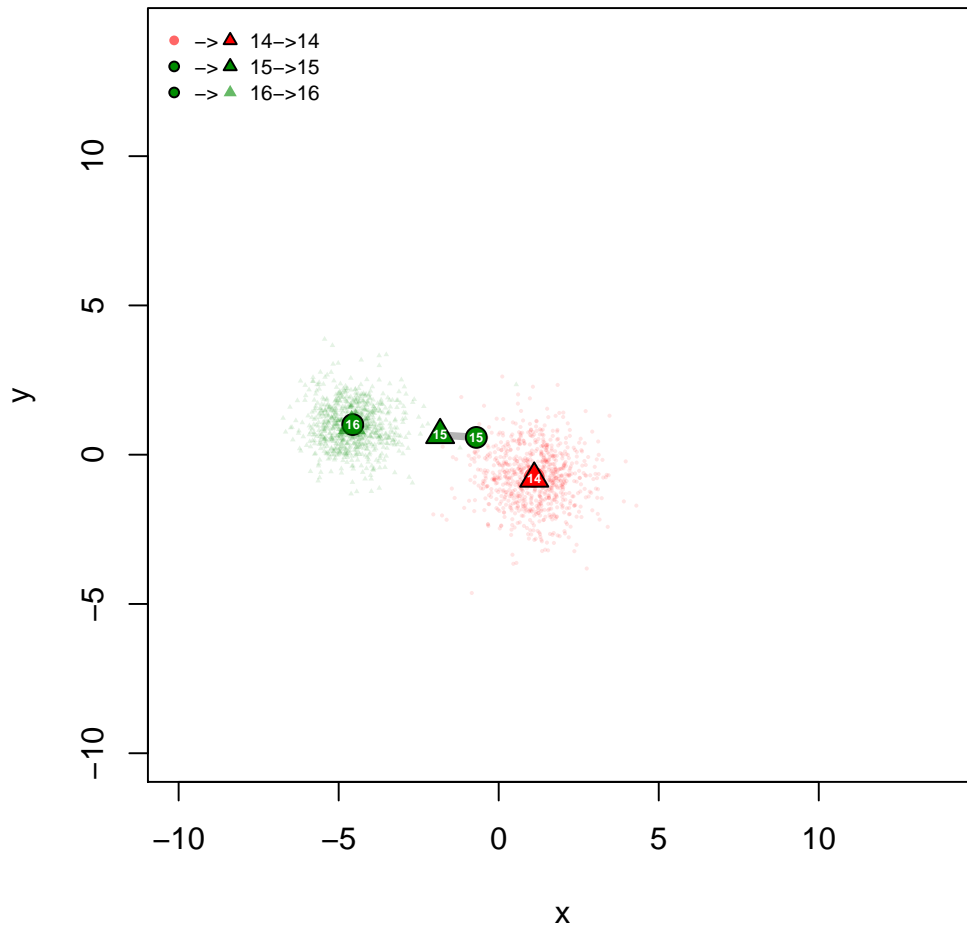**Figure 3.7:** Visualization of imputed values and matching for Patient 3.

ID = 4

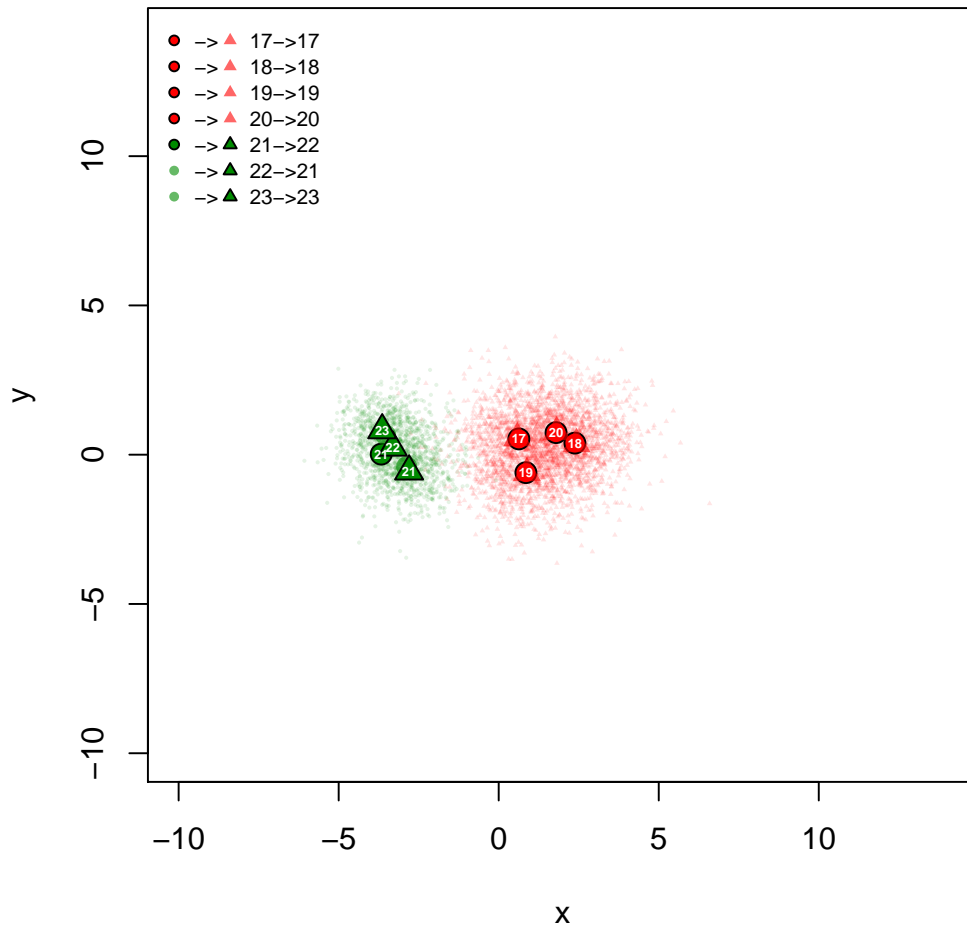**Figure 3.8:** Visualization of imputed values and matching for Patient 4.

Figure 3.9: Visualization of imputed values and matching for Patient 5

Figure 3.10: Visualization of imputed values and matching for Patient 6

Figure 3.11: Visualization of imputed values and matching for Patient 7

Figure 3.12: Visualization of imputed values and matching for Patient 8

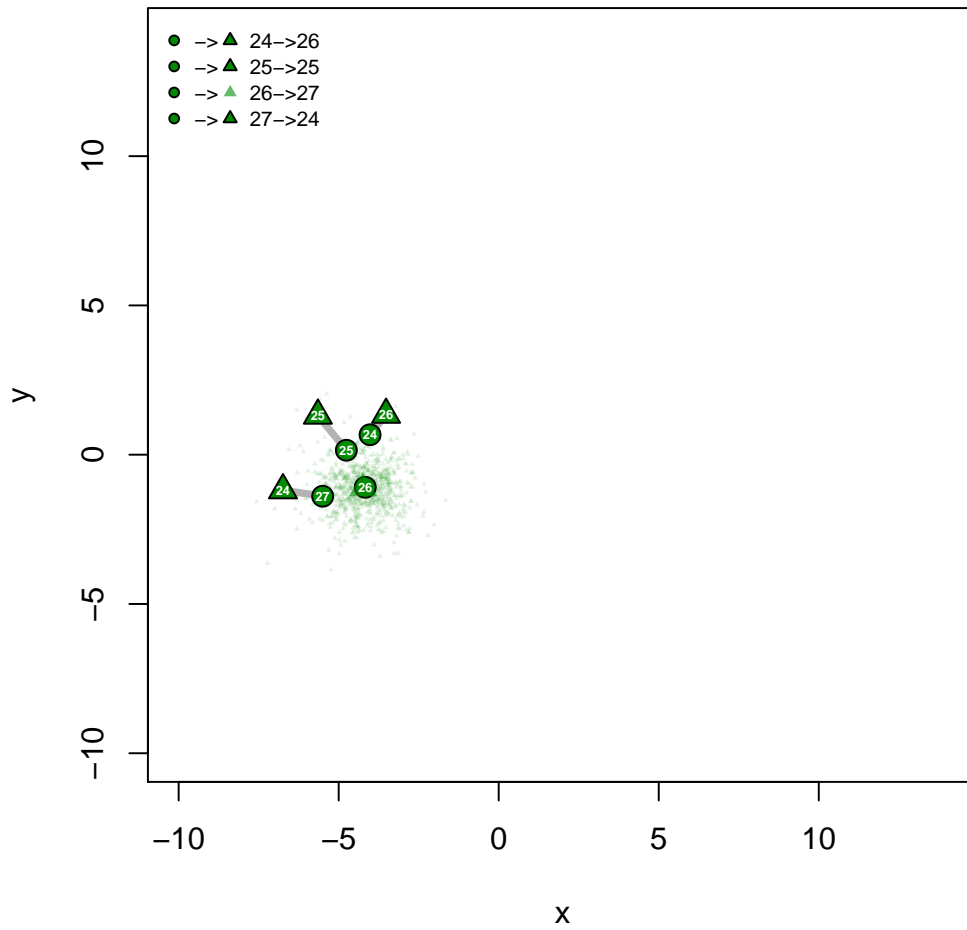**Figure 3.13:** Visualization of imputed values and matching for Patient 9

Figure 3.14: Visualization of imputed values and matching for Patient 10

ID = 11

Figure 3.15: Visualization of imputed values and matching for Patient 11

Figure 3.16: Visualization of imputed values and matching for Patient 12

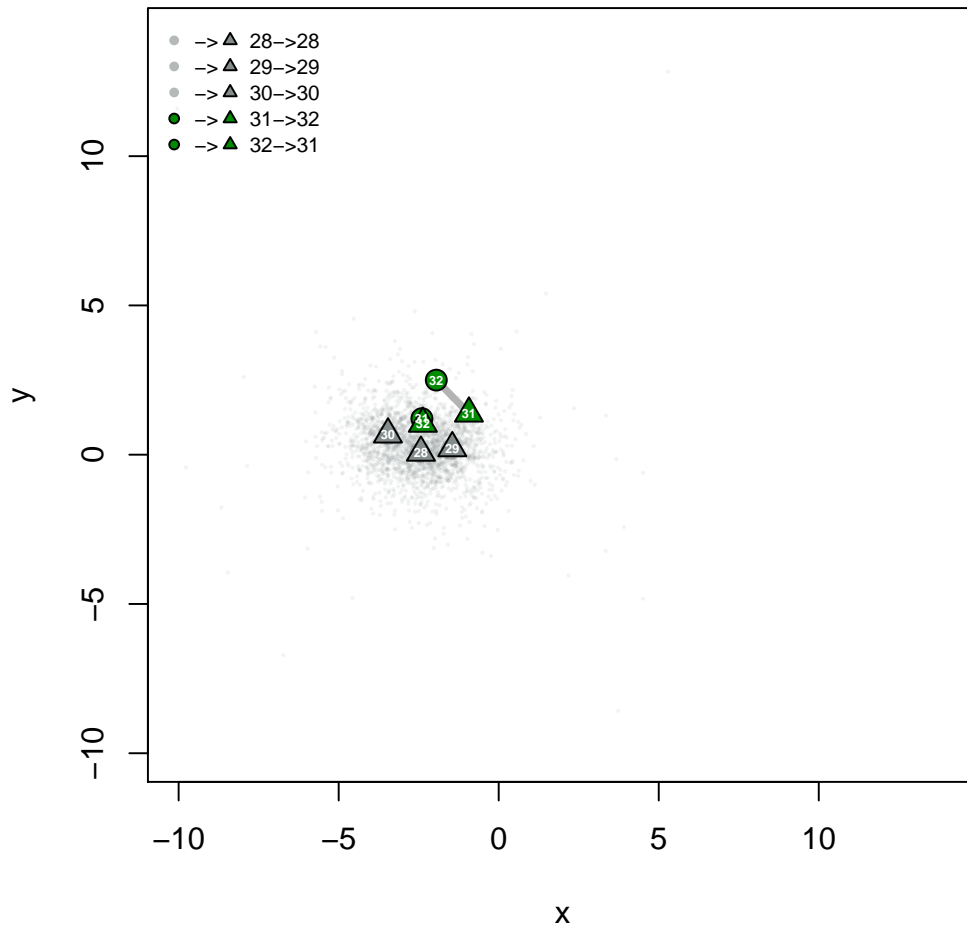**Figure 3.17:** Visualization of imputed values and matching for Patient 13

Figure 3.18: Visualization of imputed values and matching for Patient 14
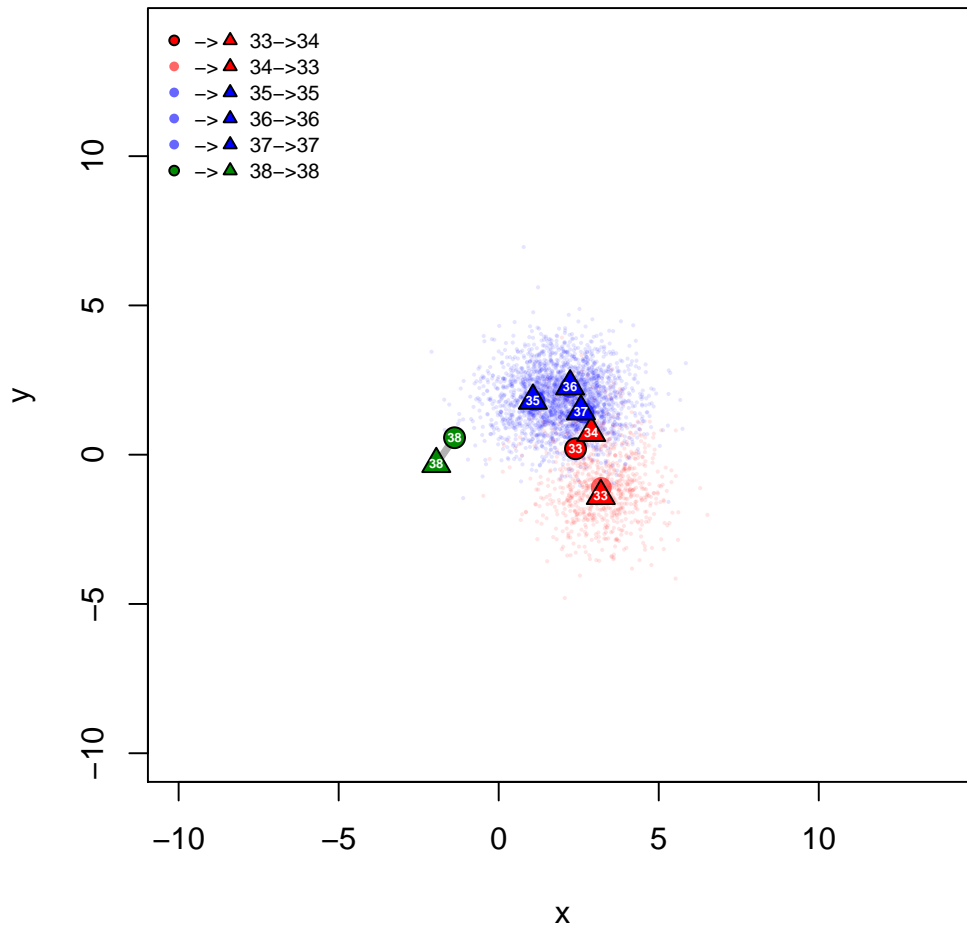
ID = 15

Figure 3.19: Visualization of imputed values and matching for Patient 15

Figure 3.20: Visualization of imputed values and matching for Patient 16

ID = 17

Figure 3.21: Visualization of imputed values and matching for Patient 17

ID = 18

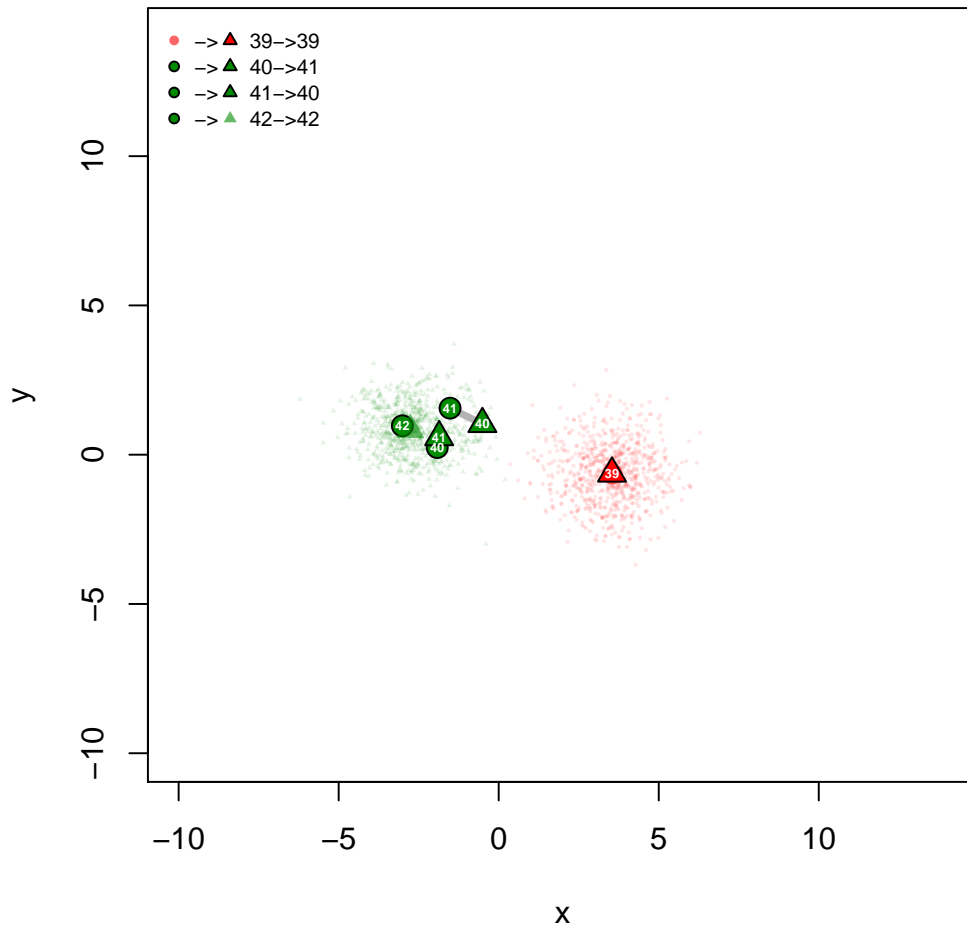**Figure 3.22:** Visualization of imputed values and matching for Patient 18

ID = 19

Figure 3.23: Visualization of imputed values and matching for Patient 19

Figure 3.24: Visualization of imputed values and matching for Patient 20

ID = 21

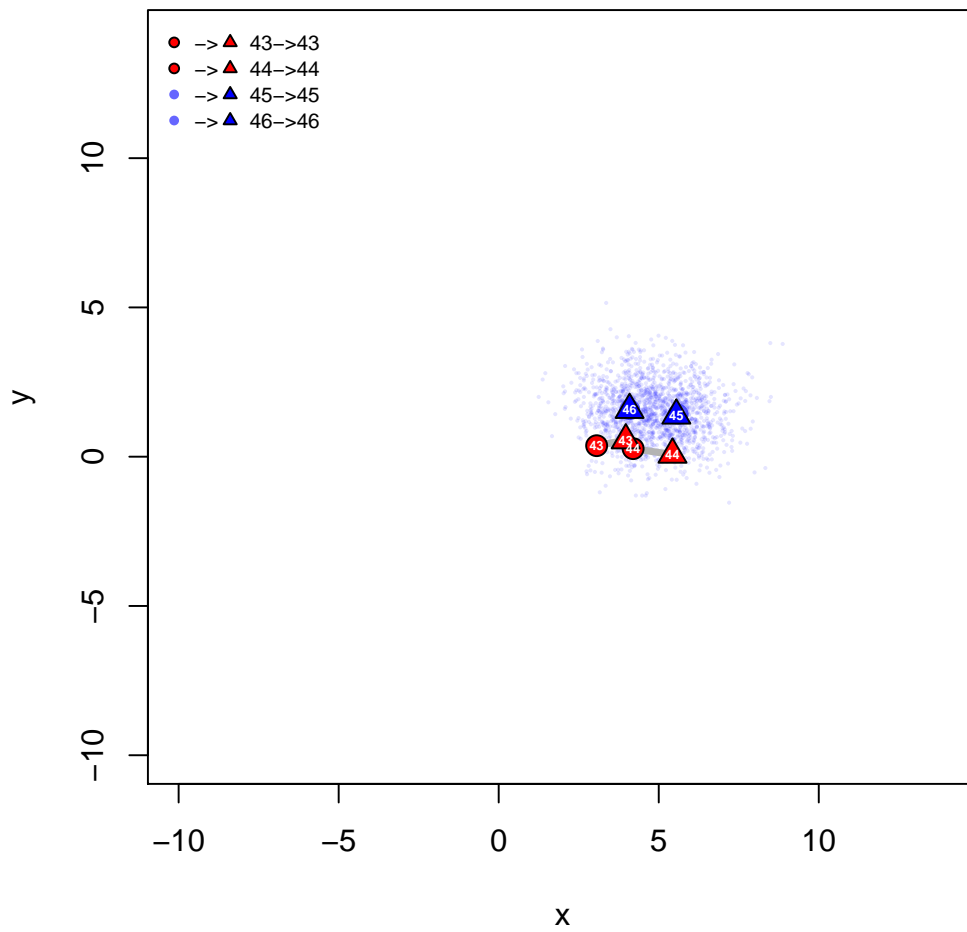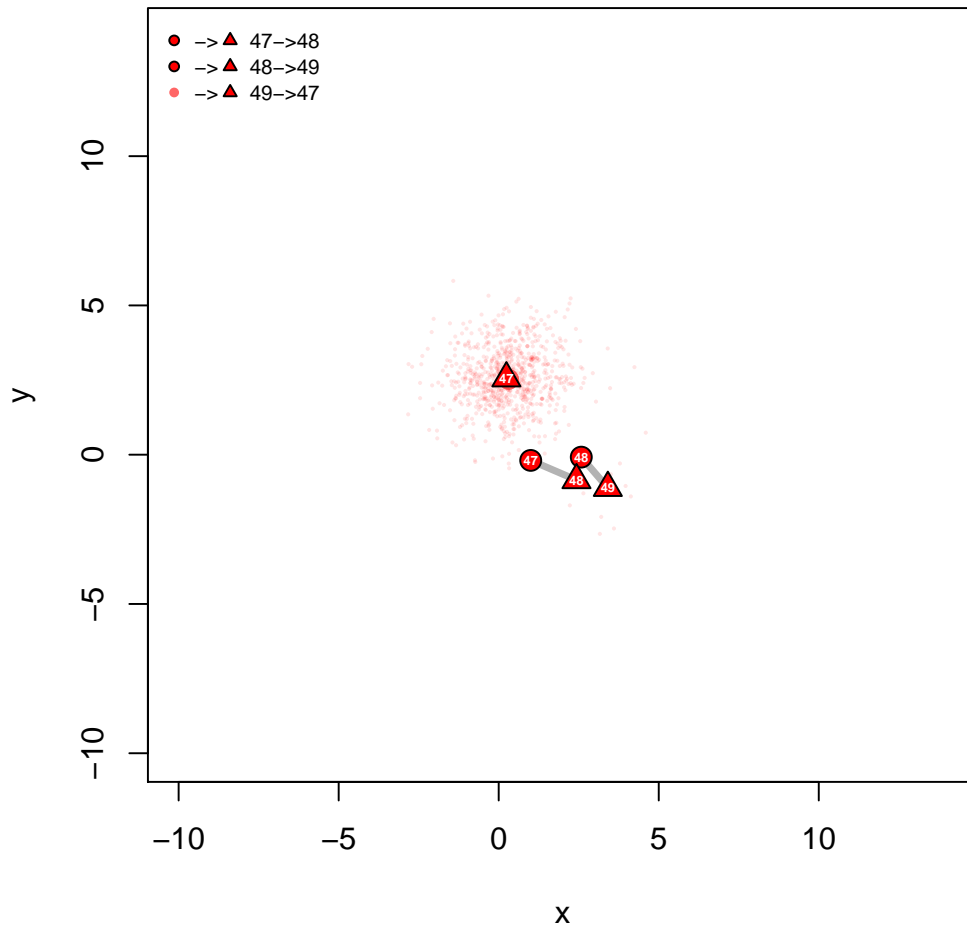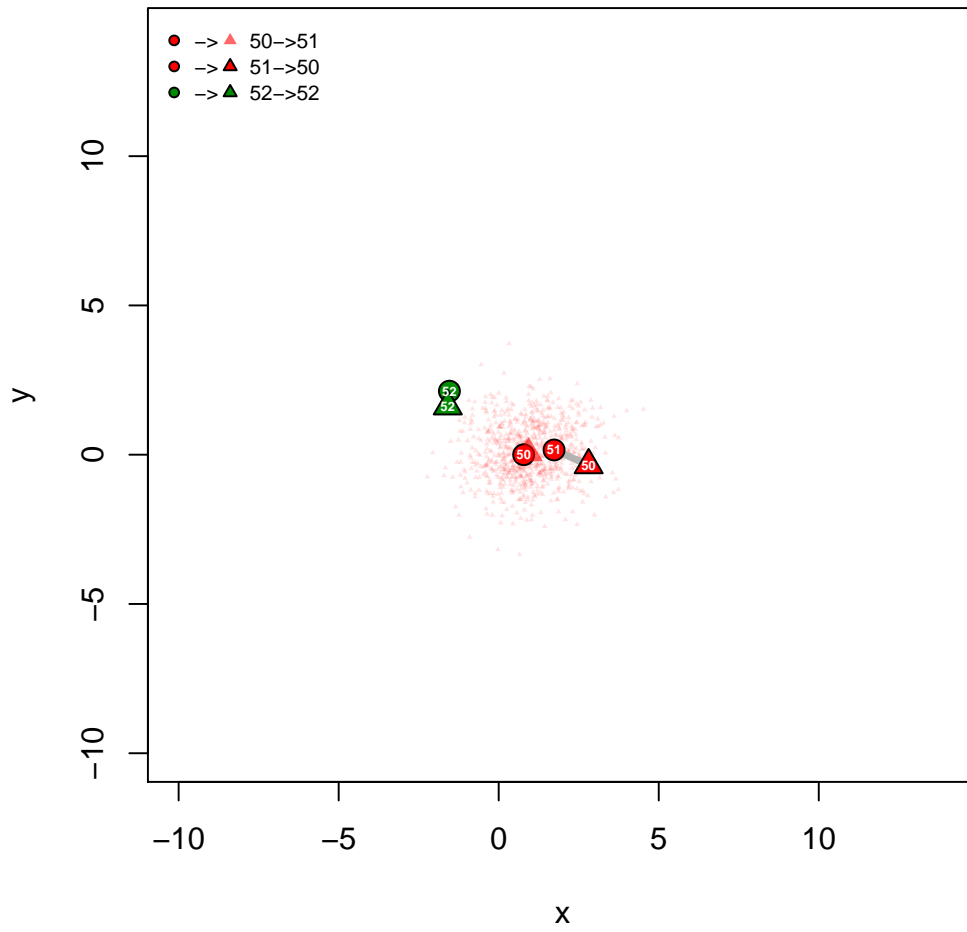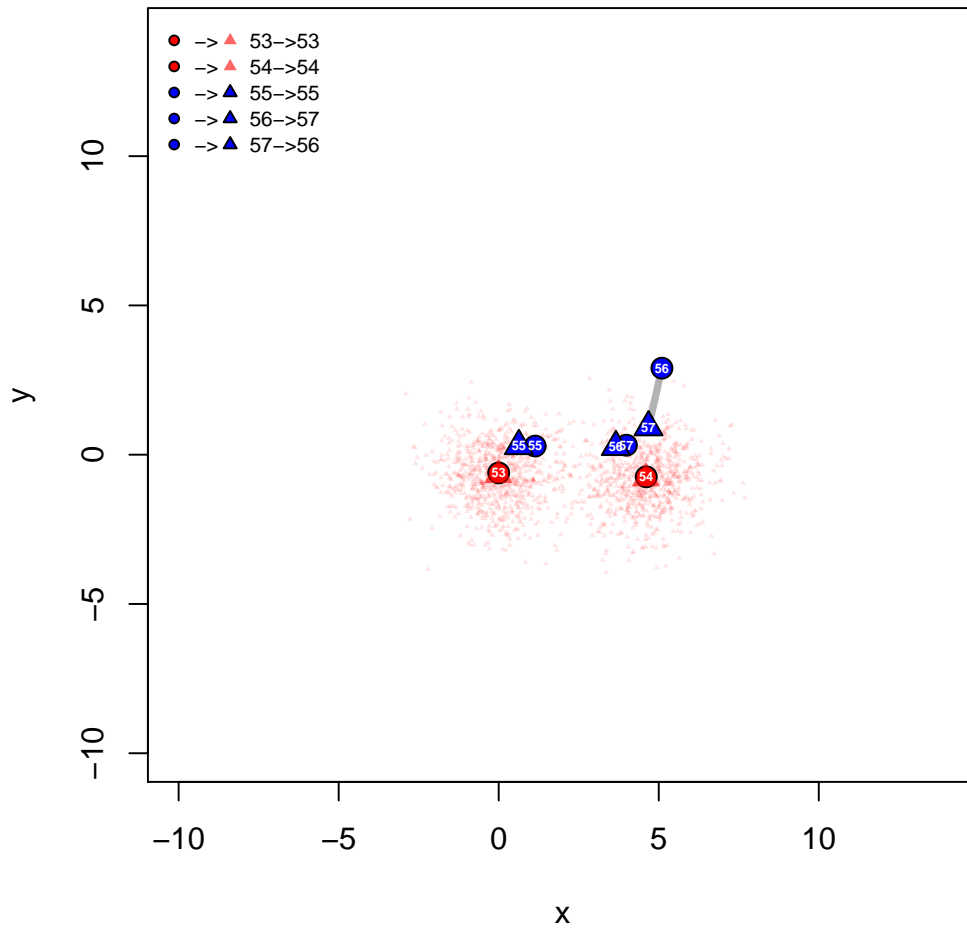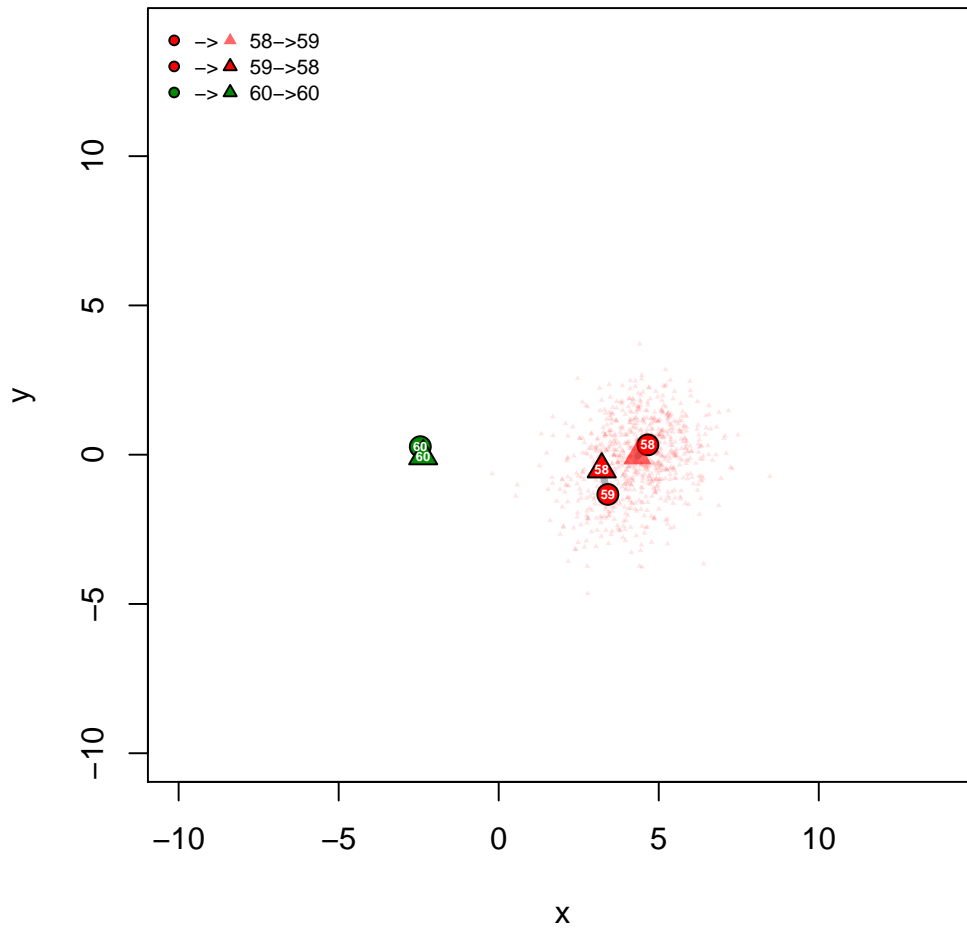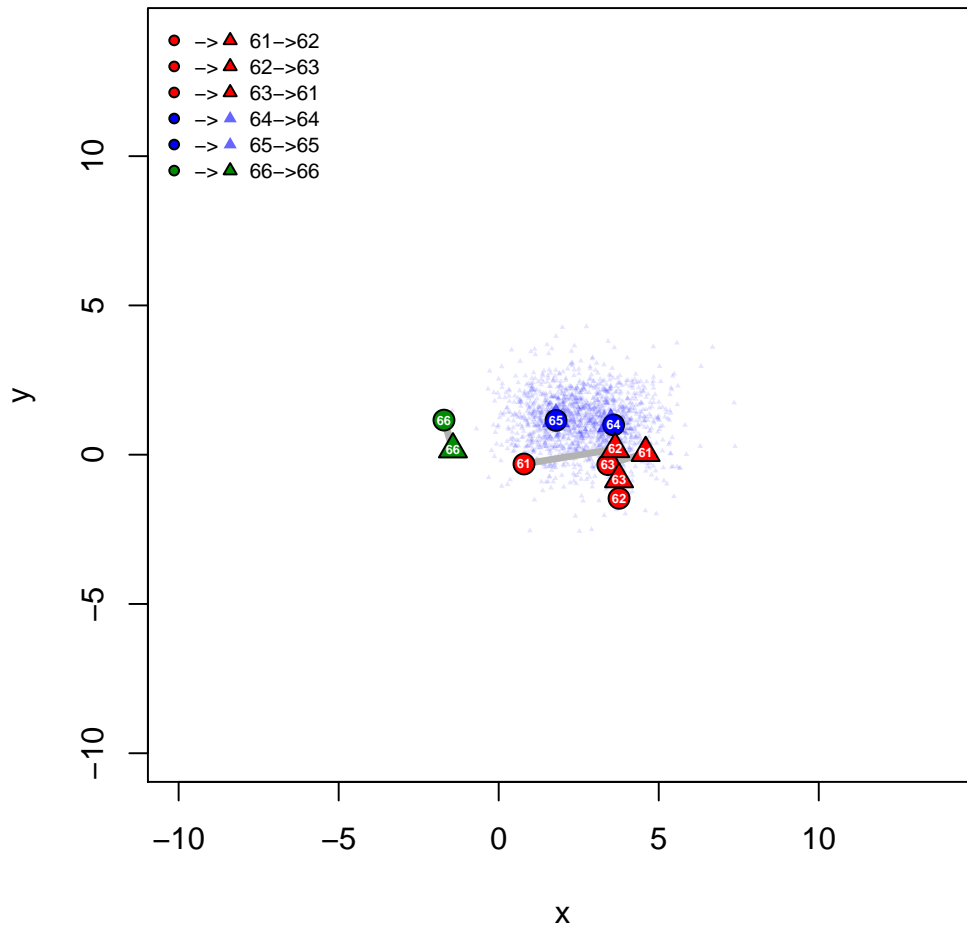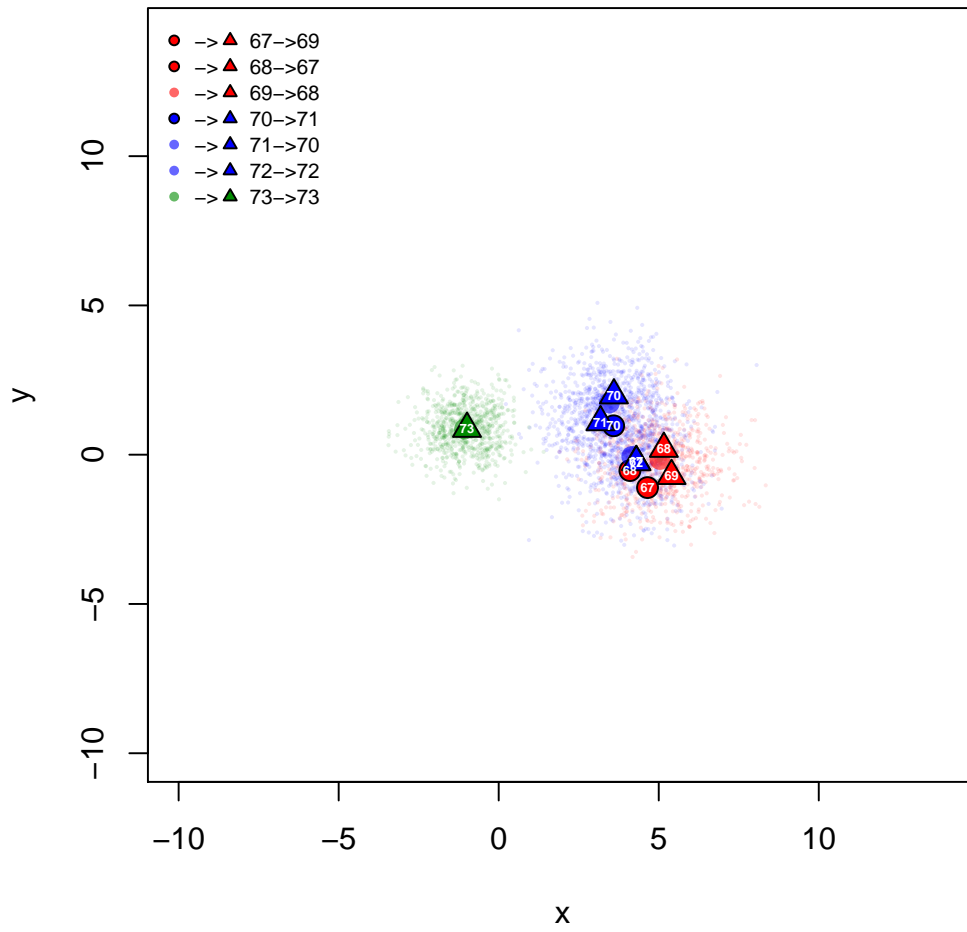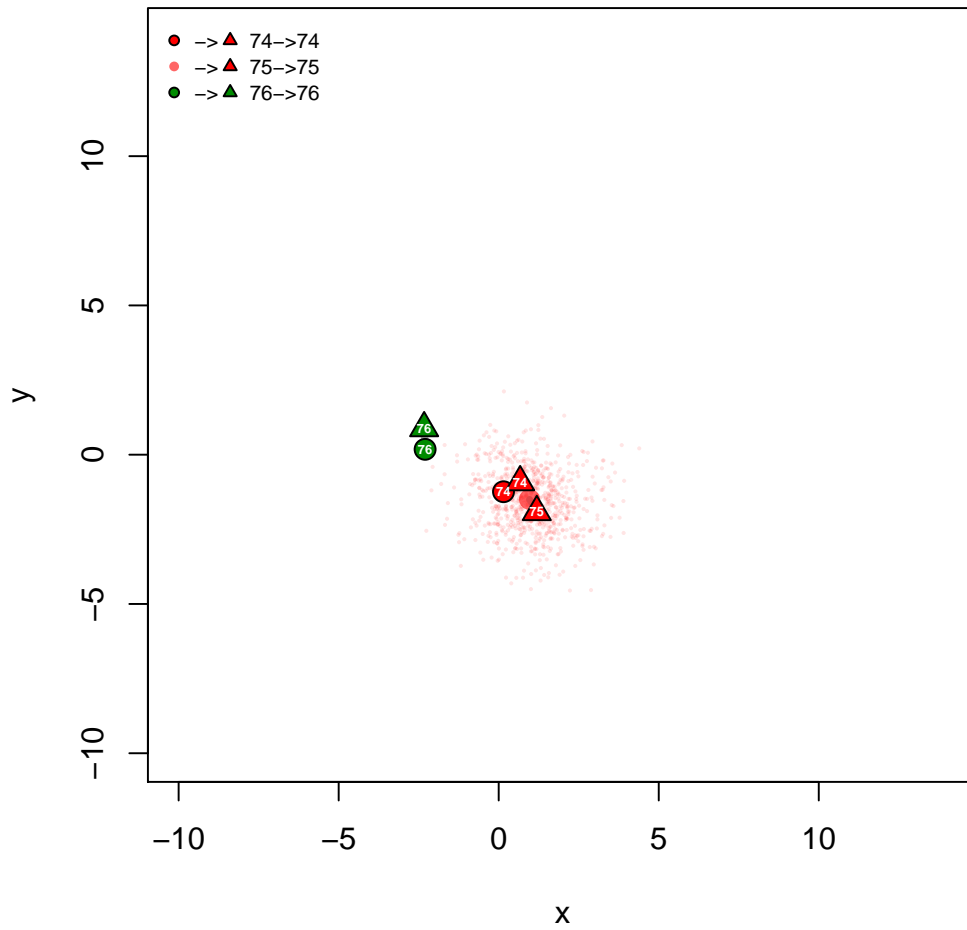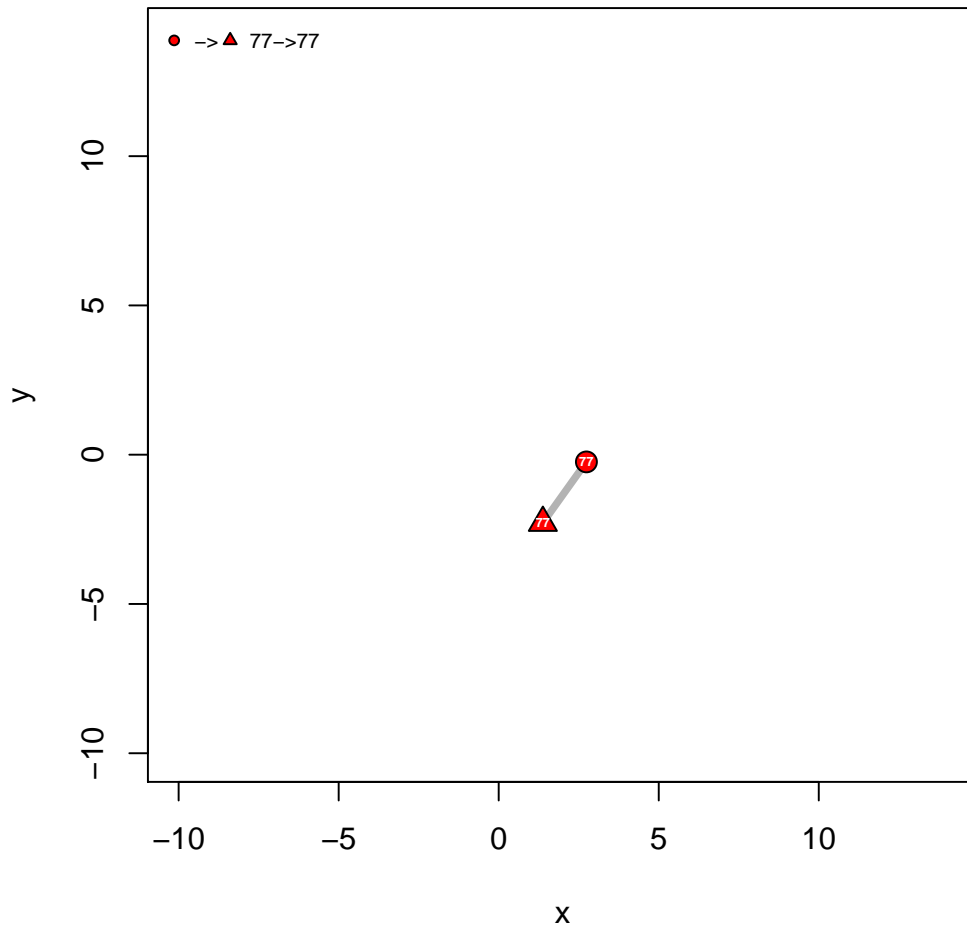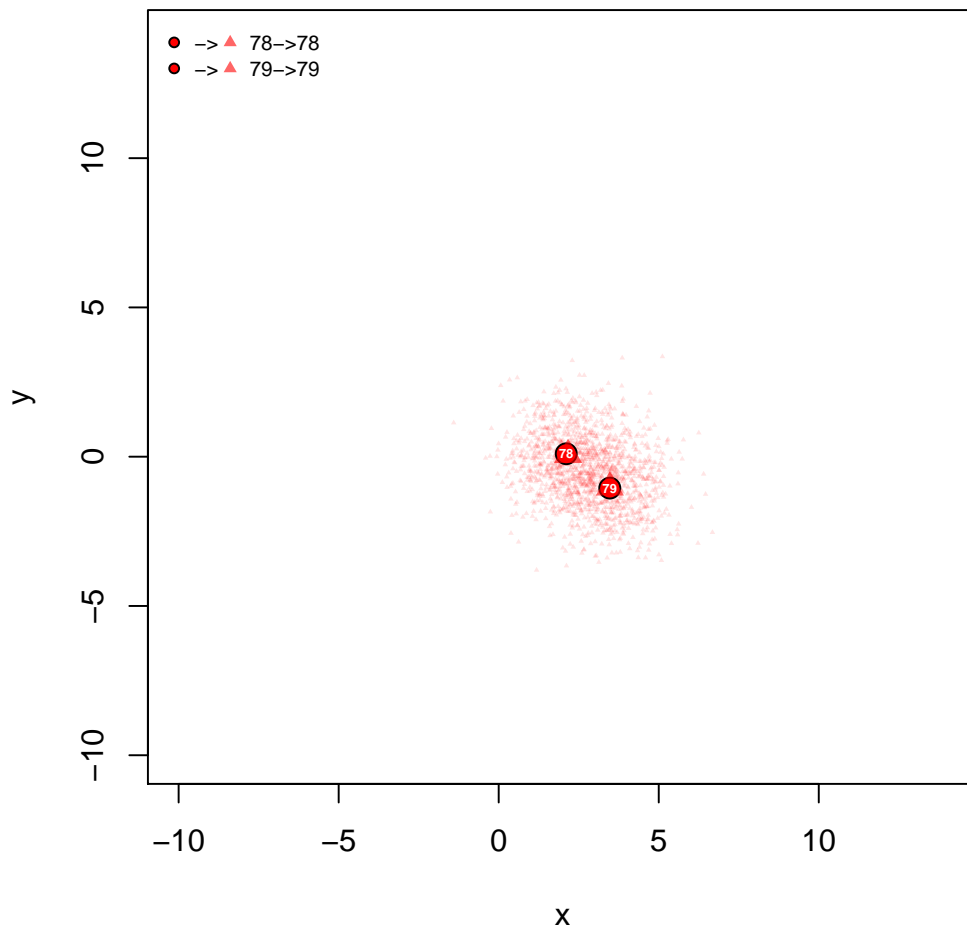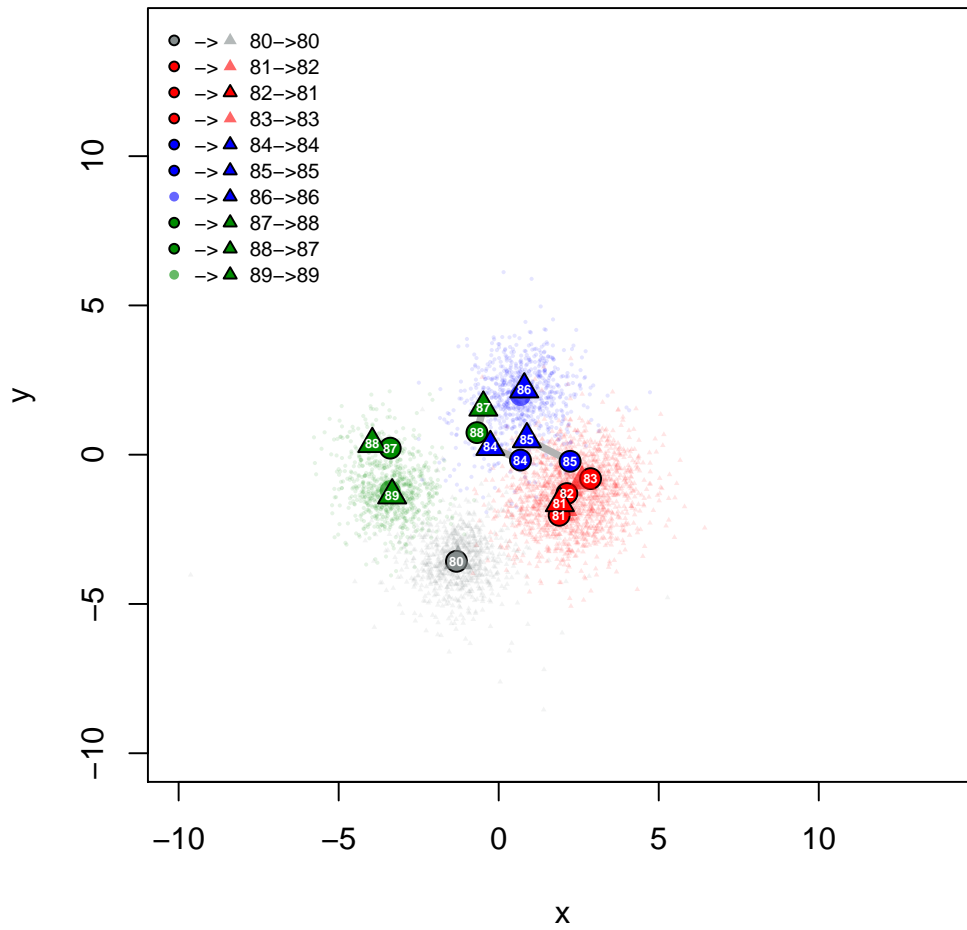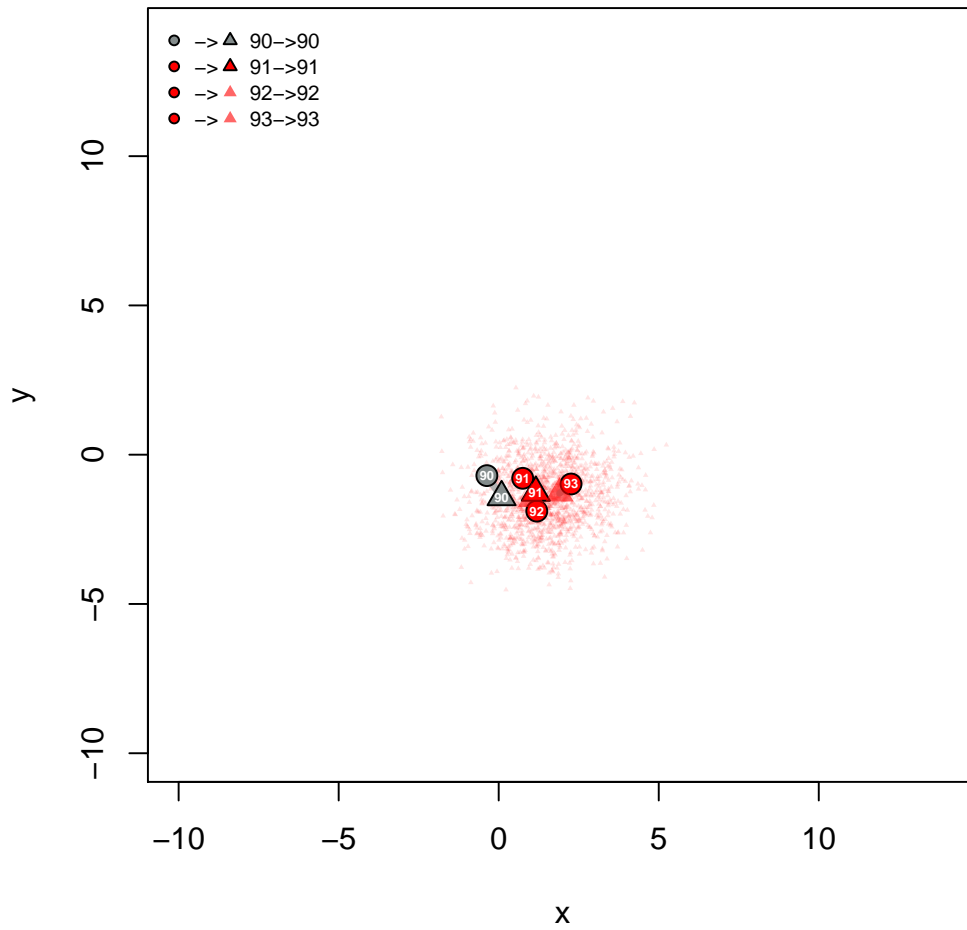**Figure 3.25:** Visualization of imputed values and matching for Patient 21
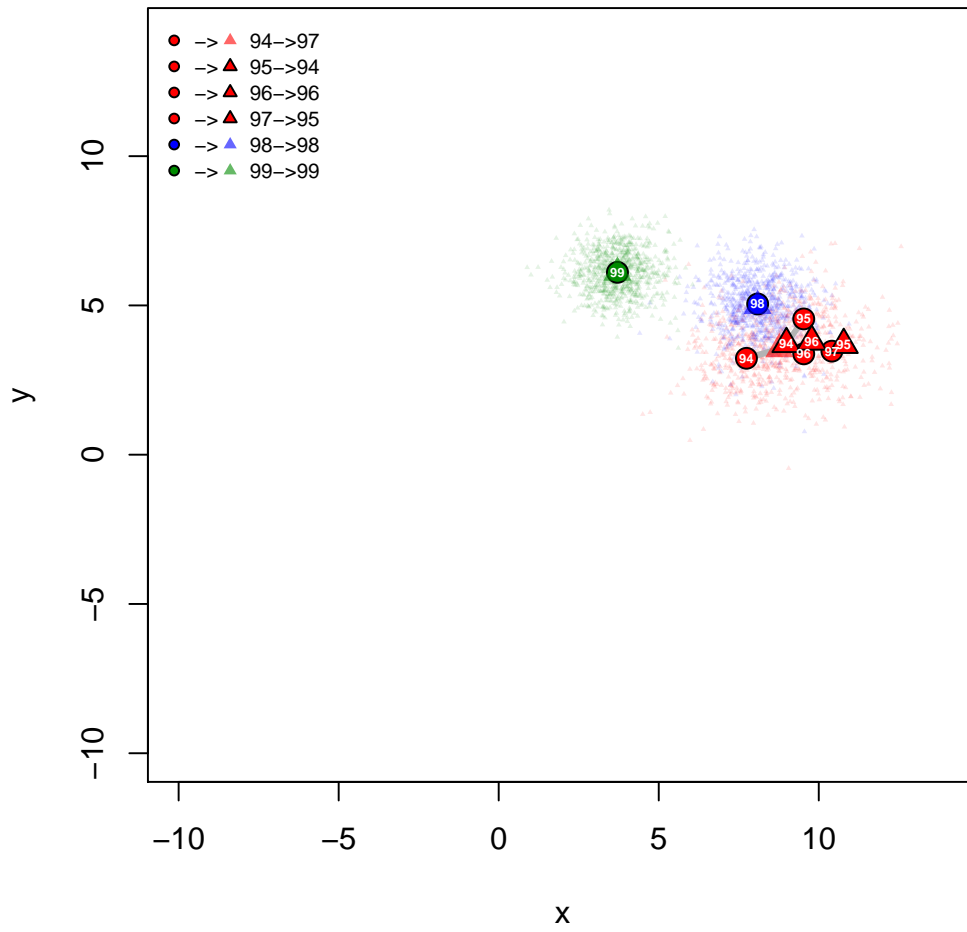
Figure 3.26: Visualization of imputed values and matching for Patient 22

Chapter 4

A note about appropriate background selection in disproportionality analyses of spontaneous drug safety reports

4.1 Introduction

Prior to regulatory authorization, pharmaceutical companies conduct clinical trials of increasing scope and complexity culminating in large-scale randomized, controlled trials. After regulatory authorization, surveillance schemes based on spontaneous reporting system (SRS) databases represent one of the cornerstones for the detection and evaluation of drug hazards. The precise details of each SRS differ in terms of size and scope, statutory reporting mandates, surveillance selectivity or intensity, and organizational structure. Prominent SRSs include the FDA Adverse Event Reporting System (FAERS) of the United States Food and Drug Administration (FDA), the Yellow Card Scheme of the Medicines and Healthcare Products Regulatory Agency (MHRA), and the international pharmacovigilance program of the World Health Organization (the WHO Uppsala Monitoring Center). These systems were created to provide warnings of possible safety problems that would be difficult to detect during clinical drug development due to power limitations, constricted range of demographics, exclusion of patients with extensive co-morbid illnesses and co-medications, and limited duration of follow-up. SRSs are often employed comparatively with other drugs (and without) to assist in characterizing potential risks.

The first step in the SRS process is the submission of case reports of suspected adverse drug reactions and/or adverse events to pharmaceutical companies and health authorities. Although legally required in some countries, there is de facto voluntary reporting for all but pharmaceutical manufacturers. The literature surveying the factors that influence reporting behavior is extensive (Gibbons et al. 2010). Algorithmic methods have emerged to identify signals in SRS databases. These include disproportionality analyses, sequential probability ratio tests, and multiple regression. Most of the practical experience to date has been with so called disproportionality analyses. While the precise operational details of each disproportionality algorithm vary, they all calculate surrogate observed-to-expected ratios in which the reporting experience of each reported Drug-Event Combination (DEC) is compared

to the background reporting experience across all other drugs using an independence model. Head-to-head comparisons of drugs are also not uncommon. In the appropriate clinical context, DECs that stand out statistically against the background reporting experience merit special attention.

4.2 Disproportionality methods

Disproportionality analysis methods for drug safety surveillance comprise the most widely used class of analytic methods for signal detection in SRSs. These methods include the DuMouchel's Bayesian multi-item gamma-Poisson shrinker (MGPS) (DuMouchel 1999, DuMouchel and Pregibon 2001), the Bayesian confidence propagation neural network (BCPNN) (Bate 2007), proportional reporting ratios (PRR) (Evans et al. 2001), and reporting odds ratios (ROR). The methods search SRS databases for potential signals, focusing on low-dimensional projections of the data, specifically 2-dimensional contingency tables. Table 4.1 shows a fictitious table.

Table 4.1: A fictitious 2-dimensional projection of an SRS database onto a particular drug (drug$_i$) and a particular adverse event (AE$_j$). The cells count the numbers of the reports in the SRS database that involve (or do not involve) drug$_i$ and/or AE$_j$. AE=adverse event.

	AE$_j$ = Yes	AE$_j$ = No	Row total
Drug$_i$ = Yes	$w_{00} = 20$	$w_{01} = 100$	$w_{0+} = 120$
Drug$_i$ = No	$w_{10} = 100$	$w_{11} = 980$	$w_{1+} = 1080$
Column total	$w_{+0} = 120$	$w_{+1} = 1080$	$w_{++} = 1200$

The basic task of a disproportionality method is then to rank these tables in order of "interestingness". Different disproportionality methods focus on different statistical measures of association as their measure of interestingness. MGPS focuses on the "reporting ratio" (RR). The observed RR for the drug i-adverse event j combination is the observed number of occurrences of the combination (20 in Table 4.1) divided by the expected number of occurrences under a model of independence (12 in this example). Thus the observed RR

for this example is 20/12 or 1 2/3; this combination occurred about 67% more often than expected.

In the SRS context, however, the count in the w_{00} cell is often small, leading to substantial variability (and hence uncertainty about the true value of the measure of association) despite the often large numbers of reports overall. PRR and ROR do not directly address the variability issue whereas MGPS and BCPNN adopt a Bayesian approach to address the issue.

MGPS and its predecessor GPS are widely used methods and they provide a singular example of large-scale Bayesian shrinkage in routine use by regulators and pharmaceutical manufacturers worldwide. GPS proceeds as follows. Let $w_{00}(i, j)$ denote the w_{00} entry for the 2×2 table for the i-th drug and the j-th adverse event. Assume that each $w_{00}(i, j)$ is a draw from a Poisson distribution with mean $\mu(i, j)$. Let $\lambda(i, j) = \mu(i, j)/E(i, j)$, where $E(i, j) = w_{0+}(i, j)w_{+1}(i, j)/w_{++}(i.j)$, i.e., the expected value of $w_{00}(i, j)$ under an independence assumption, and is assumed to be known. The goal is to estimate these $\lambda(i, j)$'s. GPS assumes that these $\lambda(i, j)$'s arise from a particular 5-parameter prior distribution, namely a mixture of two gamma distributions. GPS adopts an empirical Bayes (EB) approach and chooses the five hyper-parameters to maximize the marginal likelihood. Commonly reported summary statistics include the geometric mean of the posterior distribution for each $\lambda(i, j)$ or the fifth percentile of the posterior (i.e.,"EBGM" and "EB05" respectively). For further details see DuMouchel (1999) and Madigan (1999).

4.3   The choice of background: the Pioglitazone example

Pioglitazone is a thiazolidinedione drug used to treat diabetes. Concerns about a possible causal association between pioglitazone and bladder cancer have existed at least since 2011 (Medical-reference 2011). Table 4.2 shows an actual 2×2 table for the drug pioglitazone and the adverse event of bladder cancer from AERS database as of 2005. Here, the cell counts are the number of reports that mention specific DECs. Due to multiple drugs and/or adverse

events recorded in individual reports, there are altogether 25 million DECs (i.e., the total count w_{++} in Table 4.2), despite the fact that there are 2.7 million distinct reports. As is typically the case, the number of reports included in row 2 of the table (the "background") vastly outnumbers the number of reports in row 1, that is, reports mentioning the drug pioglitazone.

Table 4.2: 2×2 AERS table for pioglitazone and bladder cancer as of 2005.

	Bladder cancer= Yes	Bladder cancer= No	Row total
Pioglitazone= Yes	$w_{00} = 16$	$w_{01} = 33,629$	$w_{0+} = 33,625$
Pioglitazone= No	$w_{10} = 3,512$	$w_{11} = 25,430,705$	$w_{1+} = 25,434,217$
Column total	$w_{+0} = 3,528$	$w_{+1} = 25,464,314$	$w_{++} = 25,467,842$

Table 4.3 shows the actual 2×2 table for the drug metformin and the adverse event of bladder cancer. Like pioglitazone, metformin is indicated for diabetes. Again, the number of reports included in row 2 of the table vastly outnumbers the number of reports in row 1, in this case, reports mentioning the drug metformin.

Table 4.3: 2×2 AERS table for metformin and bladder cancer as of 2005.

	Bladder cancer= Yes	Bladder cancer= No	Row total
Metformin= Yes	$w_{00} = 11$	$w_{01} = 109,406$	$w_{0+} = 109,417$
Metformin= No	$w_{10} = 3,517$	$w_{11} = 25,354,908$	$w_{1+} = 25,358,425$
Column total	$w_{+0} = 3,528$	$w_{+1} = 25,464,314$	$w_{++} = 25,467,842$

Table 4.4 shows the RR, EBGM, and EB05 statistics corresponding to the above pioglitazone and metformin tables (i.e., Table 4.2 and 4.3.) Using the standard signaling threshold of 2, a signal of disproportionate reporting exists for pioglitazone and bladder cancer using EBGM (2.72 in Table 4.5), while no such signal exists for metformin. Table 4.5 shows the same statistics but now using metformin as the "background" for pioglitazone and vice versa (i.e., new 2×2 tables consisting of row 1 only from Table 4.2 and 4.3). Using the standard signaling threshold for EBGM and EB05, remarkably, no signal exists for pioglitazone when metformin is used as the background, despite the fact that the conventional EBGM for pioglitazone and bladder cancer is more than quadruple that of metformin. Thus, even though no

signal exists for metformin and bladder cancer, using metformin as the background makes the EBGM signal for pioglitazone disappear.

Table 4.4: Disproportionality analysis for drug pioglitazone/metformin and adverse event of bladder cancer using all non-pioglitazone/metformin drugs as background from AERS database as of 2005.

Drug	w_{00}	E	RR	EB05	EBGM
Pioglitazone	16	4.65	3.43	1.78	2.72
Metformin	11	15.16	0.73	0.44	0.72

Table 4.5: Disproportionality analysis for drugs pioglitazone/metformin and adverse event of bladder cancer using metformin/pioglitazone as background from AERS database as of 2005.

Drug	w_{00}	E	RR	EB05	EBGM
Pioglitazone	16	6.35	2.52	0.96	1.3
Metformin	11	20.65	0.53	0.42	0.78

Figure 4.1 shows the posterior distributions from the disproportionality analysis for pioglitazone (shown by red curve) and metformin (blue curve) and the adverse event of bladder cancer using all other drugs as background (panel a) versus using metformin and pioglitazone as background respectively (panel b). After changing the background to a single reference drug, the posterior distribution for pioglitazone moved to the left, resulting in smaller values of EBGM and EB05 statistics.

4.4   The choice of background: the general case

In the AERS database as of 2005, for all DECs, 93.9% of such background changes (i.e., using a single drug instead of all drugs as the background) lead to smaller values of the RR. Figure 4.2 shows the distribution of RR pairs calculated using the two kinds of background for a random sample of 100 drug pairs for each adverse event in AERS. The horizontal axis represents the background using all other DEC's in the database and the vertical axis shows the RR using only one drug as the background (both on the log scale). The red-to-pink
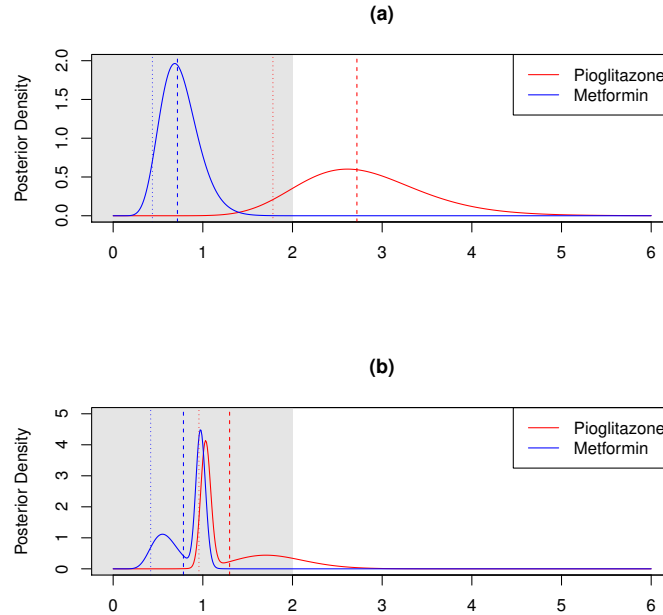
Figure 4.1: Posterior distributions for pioglitazone and metformin using all other drugs as background (panel a) VS only metformin and pioglitazone as background (panel b). Red lines represent pioglitazone and blue lines represent metformin. Dashed lines show EBGM and EB05 statistics. Gray and white delineate the standard signaling threshold of 2.

squares show the proportion of RR pairs lying in each grid on this 2-dimensional space. We can see from Figure 4.2 that the majority of the RR pairs lie below the 45-degree line, such that changing the background to a single drug leads to a smaller value of the RR.

The decreased RR's have two effects on the posterior distribution of the disproportionate analysis: (1) The posterior distribution for a certain DEC moves to the left as its RR decreases, fixing the observed cell count w_{00} and the hyper-parameters fitted by the model; (2) In (M)GPS's empirical Bayes framework, the reduced RR values for the majority of the DECs lead to change in the estimation of the hyper-parameters, which in turn pushes the posterior distribution further to the left. As a result, the EBGM and EB05 statistics become smaller, masking the signal of risk in the drug if the signaling threshold is set to the standard value of 2.

Table 4.6 illustrates the condition for the RR to decrease. Using our example in Section

113
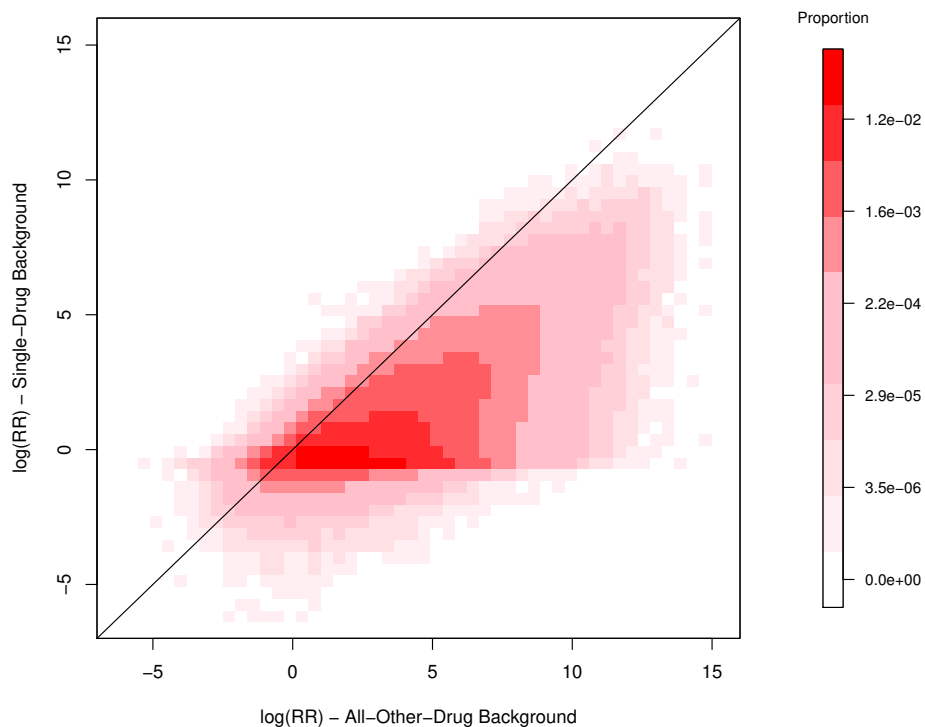
Figure 4.2: Distributions of RR's using the two backgrounds for all DEC's in AERS database as of 2005. The horizontal axis represents the background using all other drugs in the database and the vertical axis shows that using only one reference drug (both on the log scale). The red-to-pink squares show the proportion of RR pairs lying in each grid on this 2-dimensional space. A 45-degree line is also displayed.

4.3, the conventional disproportionality method considers the first row (i.e., pioglitazone) and combines the second and third rows, whereas the method using metformin as background considers only the first 2 rows. Let $A = w_{00} + w_{0'0}$ and $B = w_{01} + w_{0'1}$. One can show that RR decreases if and only if the ratio $(A/B)/(w_{10}/w_{11}) > 1$ (in our example, this ratio equals to 1.37). Since w_{11} is overwhelmingly big compared to other numbers in the table, this ratio is very likely to be greater than 1. Therefore, choosing the background of only one reference drug and carrying out head-to-head disproportionality analysis results in a very high chance to underestimate the true potential hazard of the drugs if such a hazard exists.

Table 4.6: Illustration of the condition under which RR decreases, using pioglitazone and metformin example in ARES database as of 2005.

	Bladder cancer= Yes	Bladder cancer= No	Row total
Pioglitazone	$w_{00} = 16$	$w_{01} = 33,629$	$w_{0+} = 33,625$
Metformin	$w_{0'0} = 11$	$w_{0'1} = 109,406$	$w_{0'+} = 109,417$
All other drugs	$w_{10} = 3,501$	$w_{11} = 25,321,279$	$w_{1+} = 25,324,800$
Column total	$w_{+}0 = 3,528$	$w_{+1} = 25,464,314$	$w_{++} = 25,467,842$

4.5   Discussion

Post-marketing drug safety surveillance based on large-scale spontaneous reporting system (SRS) databases is a useful approach to detect and evaluate potential drug risks after they are approved by regulatory authorization. Disproportionality analysis is a frequently applied technique in such context. Although employing the method in a comparative way to a single existing drug has been reported in the literature (Qizilbash et al. 2012), this note shows that the head-to-head comparison using a single drug as the background has a masking effect on the signaling of adverse events. In order to detect risk signals from large-scale databases such as AERS, it is thus not recommended to use a single reference drug as background in disproportionality analysis, and the standard way of involving all DECs from the entire database represents a more appropriate background selection.

Conclusion

In healthcare practice, the sample size for certain types of patients may be too small to permit accurate conclusions. For example, in personalized treatment, medical decisions are often made using limited historical information. Bayesian methods of inference allow borrowing information from other patients with more historical information, leading to more accurate understanding of the medical condition and more effective treatment of individual patients. In neuroscience, the complex structure of the data necessitates the use of flexible modeling strategies. In this dissertation, we propose Bayesian modeling strategies to address various methodological challenges in the statistical analysis of data in neuroscience and medicine.

We first construct a sparse Bayesian hierarchical VAR model for studying resting state functional connectivity using multi-subject multi-session fMRI data. Our model extends existing multi-subject single subject Bayesian VAR models to allow for the decomposition of functional connectivity into group, subject, and session components. As demonstrated in our simulations, our model improves on two stage procedures for connectivity metrics in which single subject models are estimated in a first step, then pieced together in a second step. The simulations also indicate that our model estimates are quite accurate across a wide variety of conditions, including small samples. Further, our model allows us to systematically address an important question: Is the major portion of variability in resting state neural activity due to differences between subjects or differences within subjects? Analyzing data from 50 subjects in the Human Connectome Project, we find greater variability between subjects than within subjects across sessions separated by a day, suggesting the major source of differences in human brain connectivity is due to differences among subjects. In future work, we intend to increase the computational capacity of the model. Using parallel computing techniques such

as the consensus Monte Carlo algorithm (Scott et al. 2013) in conjunction with our parameter expanded Gibbs sampler will allow us to work with more ROI's and subjects. Increasing the number of ROI's will allow us to 1) subdivide existing regions and assess whether the connections among regions we have discovered require additional refinement, and 2) extend the number of regions under investigation to a whole brain ROI model, allowing us to potentially discover new connections between regions and networks. The ability to handle larger samples will allow us to estimate parameters more reliably. It will also allow us to better differentiate among types of individuals, e.g., healthy controls and autistic individuals, and more generally allow for the exploration of sub-group differences, e.g., men and women.

We then applied the same hierarchical VAR model to data from patient diary databases, focusing on patient-level prediction of medical conditions using posterior predictive sampling. Our reanalysis of the data used in previous diary studies demonstrates the superiority of our model for patient level prediction: the advantage of partial pooling among subjects in the Bayesian model is substantial for datasets where certain patients have only a small number of observations. Our model also provides inference about population-level coefficients and between-subject heterogeneity, something that has not been systematically investigated in previous work.

We also propose a novel Bayesian model for repeat Electrical Stimulation Mappings (ESM) to quantify the change in localization of the brain regions responsible for language function over an extended period of time. We adopt an augmented Markov Chain Monte Carlo (MCMC) posterior sampling procedure, which not only evaluates the variability in language localization, but also imputes unobserved sites and provides a probabilistic solution for matching sites across occasions. Our analysis suggests that language localization is generally stable over time, but the variability in language localization gradually increases in distance and spread as the time between the two consecutive mappings increases.

Finally, we apply Bayesian disproportionality analyses to spontaneous reporting system

(SRS) databases. Using an empirical approach, we found that using the background of a single reference drug, as opposed to the original background of all other drugs in the database, generally results in smaller disproportionality metrics, masking the true safety signaling. We therefore emphasize the appropriate use of Bayesian methods and interpretation of results to avoid unwarranted conclusions and poor decision making.

Bibliography

Agresti, A. and Min, Y. (2005). Frequentist performance of Bayesian confidence intervals for comparing proportions in 2×2 contingency tables. *Biometrics*, 61:515–523.

Aven, T. and Eidesen, K. (2007). A predictive Bayesian approach to risk analysis in health care. *BMC Medical Research Methodology*, 7(38).

Bate, A. (2007). Bayesian confidence propagation neural network. *Drug Safety*, 30(7):623–625.

Beauchamp, M., Nath, A., and Pasalar, S. (2010). fMRI-guided TMS reveals that the STS is a cortical locus of the McGurk effect. *Journal of Neuroscience*, 30(7):2414–2417.

Beckmann, C., DeLuca, M., Devlin, J., and SM, S. (2005). Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):1001–1013.

Biswal, B., Yetkin, F., Haughton, V., and Hyde, J. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34(4):537–541.

Braun, U., Plichta, M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., Mier, D., Mohnke, S., Heinz, A., Erk, S., Walter, H., Seiferth, N., Kirsch, P., and Meyer-Lindenberg, A. (2012). Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *Neuroimage*, 16(2):1404–1412.

Brzan, P., Obradovic, Z., and Stiglic, G. (2017). Contribution of temporal data to predictive performance in 30-day readmission of morbidly obese patients. *PeerJ*, 5(e3230).

Buckholtz, J. and Meyer-Lindenberg, A. (2012). Psychopathology and the human connectome: toward a transdiagnostic model of risk for mental illness. *Neuron*, 74(6):990–1004.

Buckner, R., Andrews-Hanna, J., and Schacter, D. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy Sciences*, 1124:1–38.

Buckner, R. and Vicent, J. (2007). Unrest at rest: default activity and spontaneous network correlations. *Neuroimage*, 34(4):1091–1096.

Burton, C., Weller, D., and Sharpe, M. (2009). Functional somatic symptoms and psychological states: an electronic diary study. *Psychosomatic Medicine*, 71(1):77–83.

Celeveland, H. and Harris, K. (2010). The role of coping in moderating within-day associations between negative triggers and substance use cravings: A daily diary investigation. *Addictive Behaviors*, 35:60–63.

Chiang, S., Guindani, M., Yeh, H., Haneef, Z., Stern, J., and Vannucci, M. (2016). Bayesian vector autoregressive model for multi-subject effective connectivity inference using multi-modal neuroimaging data. *Human Brain Mapping*, 38(3):1311–1332.

Cohen, M. and Bookheimer, S. (1994). Localization of brain function using magnetic resonance imaging. *Trends in Neurosciences*, 17(7):268–277.

Connelly, M. and Bickel, J. (2011). An electronic daily diary process study of stress and health behavior triggers of primary headaches in children. *Journal of Pediatric Psychology*, 36(8):852–862.

Coolbrandt, A., Bruyninckx, E., Verslype, C., Steffens, E., Vanhove, E., Wildiers, H., and Milisen, K. (2017). Implementation and use of a patient symptom diary during chemotherapy: a mixed-methods evaluation of the nurse perspective. *Oncology Nursing Forum*, 44(5):e213–e222.

Cordes, D., Haughton, V., Arfanakis, K., Carew, J., Turski, P., Moritz, C., Quigley, M., and Meyerand, M. (2000). Mapping functionally related regions of brain with functional connectivity MR imaging. *American Journal of Neuroradiology*, 21(9):1636–1644.

Cribben, I. and Fiecas, M. (2016). Functional connectivity analyses for fMRI data. In Ombao, H., Lindquist, M., Thompson, W., and J, A., editors, *Handbook of Neuroimaging Data Analysis*, chapter 14. CRC Press.

Damoiseaux, J., Rombouts, S., Barkhof, F., Scheltens, P., Stam, C., Smith, S., and Beckmann, C. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences of the United States of America*, 103(37):13848–13853.

Dancey, C., Taghavi, M., and Fox, R. (1998). The relationship between daily stress and symptoms of irritable bowel: a time-series approach. *Journal of Psychosomatic Research*, 44(5):537–545.

Davis, R., Zang, P., and Zheng, T. (2016). Sparse vector autoregressive modeling. *Journal of Computational and Graphical Statistics*, 25(4):1077–1096.

Dawson, D., Lam, J., Lewis, L., Carbonell, F., Mendola, J., and Shmuel, A. (2016). Partial correlation-based retinotopically organized resting-state functional connectivity within and between areas of the visual cortex reflects more than cortical distance. *Brain Connectivity*, 6(1):57–75.

de Benedittis, G. and Lorenzetti, A. (1992). Minor stressful life events (daily hassles) in chronic primary headache: relationship with mmpi personality patterns. *Headache: The Journal of Head and Face Pain*, 32(7):330–334.

de Waal, M., Arnold, I., Eekhof, J., and van Hemert, A. (2004). Somatoform disorders in general practice: prevalence, functional impairment and comorbidity with anxiety and depressive disorders. *The British Journal of Psychiatry*, 184:470–476.

Deshpande, G., LaConte, S., James, G., Peltier, S., and Hu, X. (2009). Multivariate Granger causality analysis of fMRI data. *Human Brain Mapping*, 30(4):1361–1373.

Di, X., Fu, Z., Chan, S., Hung, Y., Biswal, B., and Zhang, Z. (2015). Task-related functional connectivity dynamics in a block-designed visual experiment. *Frontiers in Human Neuroscience*, 9(543).

Donahue, J., Weiss, S., Goetsch, M., Livingston, J., Greineder, D., and Piatt, R. (1997). Assessment of asthma using automated and full-text medical records. *The Journal of Asthma*, 34(4):273–281.

Downing, P., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293(5539):2470–2473.

D'Souza, A., Abidin, A., Leistritz, L., and Wismüller, A. (2017). Exploring connectivity with large-scale Granger causality on resting-state functional mri. *Journal of Neuroscience Methods*, 1(287):68–79.

Duffau, H. (2011). Brain plasticity: a new concept in neuroscience, a new tool in neurosurgery. In H, D., editor, *Brain mapping: from neural basis of cognition to surgical applications.*, pages 349–360. Springer.

Duffau, H., Denvil, D., and Capelle, L. (2002). Long term reshaping of language, sensory and motor maps after glioma resection: a new parameter to intergrate in the surgical strategy. *Journal of Neurology, Neurosurgery, and Psychiatry*, 72(4):511–516.

DuMouchel, W. (1999). Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *The American Statistician*, 53(3):170–190.

DuMouchel, W. and Pregibon, D. (2001). Empirical Bayes screening for multi-item associations. *Proceedings of the seventh international conference on knowledge discovery and data Mining SIGKDD*, pages 67–76.

Evans, S., Waller, P., and Davis, S. (2001). Use of proportional reporting ratios (prrs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Safety*, 10(6):483–486.

Fink, P., Sorensen, L., Engberg, M., Holm, M., and Munk-Jorgensen, P. (1999). Somatization in primary care. prevalence, health care utilization, and general practitioner recognition. *Psychosomatics*, 40(4):330–338.

Fox, M., Snyder, A., Vincent, J., Corbetta, M., van Essen, D., and Raichle, M. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9673–9678.

Fraser, D. (2011). Is Bayes posterior just quick and dirty confidence? *Statistical Science*, 26(3):299–316.

Friston, K. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2:56–78.

Ge, T., Holmes, A., Buckner, R., Smoller, J., and Sabuncu, M. (2017). Heritability analysis with repeat measurements and its application to resting-state functional connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, 114(21):5521–5526.

Gefang, D. (2014). Bayesian doubly adaptive Elastic-net Lasso for VAR shrinkage. *International Journal of Forecasting*, 30(1):1–11.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2014). *Bayesian Data Analysis*. Chapman & Hall CRC.

Genkin, A., Lewis, D., and Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics*, 49(3):291–304.

Gibbons, D., Amatya, A., Brown, C., Hur, K., Marcus, S., Bhaumik, D., and Mann, J. (2010). Post-approval drug safety surveillance. *Annual Review of Public Health*, 31:419–437.

Goghari, V. and Lawlor-Savage, L. (2017). Comparison of cognitive change after working memory training and logic and planning training in healthy older adults. *Frontiers in Aging Neuroscience*, 9(39).

Gorrostieta, C., Fiecas, M., Ombao, H., Burke, E., and Cramer, S. (2013). Hierarchical vector auto-regressive models and their applications to multi-subject effective connectivity. *Frontiers in Computational Neuroscience*, 7(159).

Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438.

Green, P., Łatuszyński, K., Pereyra, M., and Robert, C. P. (2015). Bayesian computation: a summary of the current state, and samples backwards and forwards. *Statistics and Computing*, 25(4):835–862.

Greenberg, E. (2012). *Introduction to Bayesian Econometrics (2 ed.)*. Cambridge University Press.

Greene, J., Walker, L., Hickson, G., and Thompson, J. (1985). Stressful life events and somatic complaints in adolescents. *Pediatrics*, 75(1):19–22.

Greicius, M., Srivastava, G., Reiss, A., and Menon, V. (2004). Default-mode network activity distinguishes alzheimer's disease from healthy aging: evidence from functional mri. *Proceedings of the National Academy of Sciences of the United States of America*, 101(13):4637–4642.

Gusnard, D. and Raichle, M. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nature Reviews Neuroscience*, 2(10):685–649.

Haglund, M., Berger, M., Shamseldin, M., Lettich, E., and Ojemann, G. (1994). Cortical localization of temporal lobe language sites in patients with gliomas. *Neurosurgery*, 34(4):567–576.

Harrold, L., Saag, K., Yood, R., Mikuls, T., Andrade, S., Fouayzi, H., Davis, J., Chan, K., Raebel, M., Von Worley, A., and Platt, R. (2007). Validity of gout diagnoses in administrative data. *Arthritis and Rheumatism*, 57(1):103–108.

Hatcher, S. and House, A. (2003). Life events, difficulties and dilemmas in the onset of chronic fatigue syndrome: a case-control study. *Psychological Medicine*, 33(7):1185–1192.

He, A., Tan, L., Tang, Y., James, G., Wright, P., Eckert, M., Fox, P., and Liu, Y. (2003). Modulation of neural connectivity during tongue movement and reading. *Human Brain Mapping*, 18(3):222–232.

He, D., Mathews, S., Kalloo, A., and Hutfless, S. (2014). Mining high-dimensional administrative claims data to predict early hospital readmissions. *Journal of the American Medical Informatics Association*, 21(2):272–279.

He, H. and Garcia, E. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21:1263–1284.

Hennessy, S. (2006). Use of health care databases in pharmacoepidemiology. *Basic and Clinical Pharmacology and Toxicology*, 98(3):311–313.

Hennessy, S., Leonard, C., Freeman, C., Deo, R., Newcomb, C., Kimmel, S., Strom, B., and Bilker, W. (2010). Validation of diagnostic codes for outpatient originating sudden cardiac death and ventricular arrhythmia in medicaid and medicare claims data. *Pharmacoepidemiology and Drug Safety*, 19(6):555–562.

Hesketh, T., Zhen, Y., Lu, L., Dong, Z., Jun, Y., and Xing, Z. (2010). Stress and psychosomatic symptoms in chinese school children: cross-sectional survey. *Archives of Disease in Childhood*, 95(2):136–140.

Higdon, D. (1998). Auxiliary variable methods for markov chain Monte Carlo with applications. *Journal of the American Statistical Association*, 93(442):585–595.

Hripcsak, G., Patrick, R., Duke, J., Shah, N., Park, R., Huserh, V., Suchard, M., Schuemie, M., DeFalco, F., Perotte, A., Banda, J., Reich, C., Schilling, L., Matheny, M., Meeker, D., Pratt, N., and Madigan, D. (2016). Addressing clinical questions at scale: Ohdsi characterization of treatment pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27):7329–7336.

Iturria-Medina, Y. and Evans, A. (2015). On the central role of brain connectivity in neurodegenerative disease progression. *Frontiers in Aging Neuroscience*, 7(90).

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. Wiley.

Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., and Kripalani, S. (2011). Risk prediction models for hospital readmission: A systematic review. *The Journal of the American Medical Association*, 306(15):1688–1698.

Karlsson, S. (2015). Forecasting with Bayesian vector autoregression. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, pages 791–987. North Holland.

Kennedy, D., Redcay, E., and Courchesne, E. (2006). Failing to deactivate: resting functional abnormalities in autism. *Proceedings of the National Academy of Sciences of the United States of America*, 103(21):8275–8280.

Kepinska, O., Pereda, E., Caspers, J., and Schiller, N. (2017). Neural oscillatory mechanisms during novel grammar learning underlying language analytical abilities. *Brain and Language*, 175:99–110.

Kiviniemi, V., Kantola, J., Jauhiainen, J., Hyvärinen, A., and Tervonen, O. (2003). Independent component analysis of nondeterministic fMRI signal sources. *Neuroimage*, 19(2 pt 1):253–260.

Kohno, M., Morales, A., Ghahremani, D., Hellemann, G., and London, E. (2015). Risky decision-making: prefrontal function and mesocorticolimbic resting-state connectivity in methamphetamine users. *JAMA Psychiatry*, 71(7):812–820.

Koop, G. and Korobilis, D. (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3(4):267–358.

Lee, M., Smyser, C., and Shimony, J. (2013). Resting-state fMRI: a review of methods and clinical applications. *American Journal of Neuroradiology*, 34(10):1866–1872.

Lewis, J., Schinnar, R., Bilker, W., Wang, X., and Strom, B. (2007). Validation studies of the health improvement network (thin) database for pharmacoepidemiology research. *Pharmacoepidemiology and Drug Safety*, 16(4):393–401.

Li, Q., Lin, N., and Xi, R. (2010). Bayesian regularized quantile regression. *Bayesian Analysis*, 5(3):533–556.

Lindquist, M. and Sobel, M. (2016). Effective connectivity and causal inference in neuroimaging. In Ombao, H., Lindquist, M., Thompson, W., and J, A., editors, *Handbook of Neuroimaging Data Analysis*, chapter 14. CRC Press.

Litterman, R. (2010). Techniques of forecasting using vector autoregressions. *Federal Reserve Bank of Minneapolis Working Paper*, 15.

Liu, D., Yan, C., Ren, J., Yao, L., Kiviniemi, V., and Zang, Y. (2010). Using coherence to measure regional homogeneity of resting-state fMRI signal. *Frontiers in Systems Neuroscience*, 4(24).

Lütkepohl, H. (2005). *SNew Introduction to Multiple Time Series Analysis*. Springer Science & Business Media.

Madigan, D. (1999). Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system: discussion. *The American Statistician*, 53(3):198–200.

Madigan, D., Stang, P., Berlin, J., Schuemie, M., Overhage, J., Suchard, M., Dumouchel, W., Hartzema, A., and Ryan, P. (2014). A systematic statistical approach to evaluating evidence from observational studies. *Annual Review of Statistics and Its Application*, 1:11–39.

Margulies, D., Böettger, J., Long, X., Lv, Y., Kelly, C., Schäefer, A., Goldhahn, D., Abbushi, A., Milham, M., Lohmann, G., and Villringer, A. (2010). Resting developments: a review of fMRI post-processing methodologies for spontaneous brain activity. *Magma*, 23(5-6):289–307.

Martínez-Montes, E., Valdés-Sosa, P., Miwakeichi, F., Goldman, R., and Cohen, M. (2004). Concurrent EEG/fMRI analysis by multiway partial least squares. *Neuroimage*, 22(3):1023.

McCarthy, M. (2007). *Bayesian Methods for Ecology (1st Ed.)*. Cambridge University Press.

Medical-reference (2011). France and Germany suspended use of Actos for bladder cancer risk. Accessed: 2011.

Nagai, M., Nishikawa, N., Iwaki, H., Ando, R., Yabe, H., and Nomoto, M. (2016). A novel patient diary for assessment of motor fluctuations and drug efficacy in parkinson's disease. *Movement Disorders*, 31(S53).

Ogawa, S., Lee, T., Key, A., and Tank, D. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 84(24):9868–9872.

Ojemann, G. (1983a). Brain organization for language from the perspective of electrical stimulation mapping. *Behavioral and Brain Sciences*, 6(2):189–206.

Ojemann, G. (1983b). Electrical stimulation and neurobiology of language. *Behavioral and Brain Sciences*, 6(2):221–230.

Ojemann, G., Ojemann, J., Lettich, E., and Berger, M. (1989). Cortical language localization in left, dominant hemisphere: An electrical stimulation mapping investigation in 117 patients. *Journal of Neurosurgery*, 71(3):316–326.

Ojemann, S., Berger, M., Lettich, E., and Ojemann, G. (2003). Localization of language function in children: results of electrical stimulation sapping. *Journal of Neurosurgery*, 98(3):465–740.

Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482):681–686.

Penfield, W. and Roberts, L. (1959). *Speech and Brain Mechanisms*. Princeton University Press.

Peng, C., Chen, Y., Cui, Y., Zhao, D., Jiao, Y., Tang, T., Ju, S., and Teng, G. (2016). Regional coherence alterations revealed by resting-state fMRI in post-stroke patients with cognitive dysfunction. *PLOS One*, 11(7).

Petrella, J., Sheldon, F., Prince, S., Calhoun, V., and Doraiswamy, P. (2011). Default mode network connectivity in stable vs progressive mild cognitive impairment. *Neurology*, 76(6):511–517.

Peveler, R., Kilkenny, L., and Kinmonth, A. (1997). Medically unexplained physical symptoms in primary care: a comparison of self-report screening questionnaires and clinical opinio. *Journal of Psychosomatic Research*, 42(3):245–252.

Pievani, M., Filippini, N., van den Heuvel, M., Cappa, S., and Frisoni, G. (2014). Brain connectivity in neurodegenerative diseases–from phenotype to proteinopathy. *Nature Reviews Neurology*, 10(11):620–633.

Powers, D. (2011). Evaluation: from precision, recall and f-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1):37–63.

Qizilbash, N., Mendez, I., and Sanchez-de la Rosa, R. (2012). Benefit-risk analysis of glatiramer acetate for relapsing-remitting and clinically isolated syndrome multiple sclerosis. *Clinical Therapeutics*, 34(1):159–176.

Raichle, M. and Mintun, M. (2006). Brain work and brain imaging. *Annual Review of Neuroscience*, 29:449–476.

Robinson, D., Greene, J., and Walker, L. (1988). Functional somatic complaints in adolescents: Relationship to negative life events, self-concept, and family characteristics. *The Journal of Pediatrics*, 113(3):588–593.

Robles, S., Gatignol, P., Lehericy, S., and Duffau, H. (2008). Long-term brain plasticity allowing a multistage surgical approach to world health organization grade ii gliomas in eloquent areas. *Journal of Neurosurgery*, 109(1):615–624.

Roebroeck, A., Formisano, E., and Goebel, R. (2005). Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage*, 25(1):449–476.

Saito, T. and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluation binary classifiers on imbalanced datasets. *PLOS One*, 10(3):e0118432.

Salvador, R., Suckling, J., Schwarzbauer, C., and Bullmore, E. (2005). Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):937–946.

Sanchez-Morillo, D., Fernandez-Granero, M., and Leon-Jimenez, A. (2016). Use of predictive algorithms in-home monitoring of chronic obstructive pulmonary disease and asthma: a systematic review. *Chronic Respiratory Disease*, 13(3):264–283.

Schippers, M., Renken, R., and Keysers, C. (2011). The effect of intra- and inter-subject variability of hemodynamic responses on group level Granger causality analyses. *Neuroimage*, 57(1):22–36.

Scott, S., Blocker, A., Bonassi, F., Chipman, H., George, E., and McCulloch, R. (2013). Bayes and big data: the consensus Monte Carlo algorithm. *International Journal of Management Science and Engineering Management*, 11:78–88.

Seek, M., Pegna, A., Ortigue, S., Spinelli, L., Dessibourg, C., Delavelle, J., Blanke, O., Michel, C., Landis, T., and Villemure, J. (2006). Speech arrest with stimulation may not reliably predict language deficit after epilepsy surgery. *Neurology*, 66(4):592–594.

Senden, M., Reuter, N., van den Heuvel, M., Goebel, R., Deco, G., and Gilson, M. (2017). Task-related effective connectivity reveals that the cortical rich club gates cortex-wide communication. *Human Brain Mapping*, 39(3):1246–1262.

Serafini, S., Komisarow, J., Gallentine, W., Mikati, M., Bonner, M., Kranz, P., Haglund, M., and Grant, G. (2013). Reorganization and stability for motor and language areas using cortical stimulation: Case example and review of the literature. *Brain Sciences*, 3(4):1597–1614.

Shahn, Z., Ryan, P., , and Madigan, D. (2015). Predicting health outcomes from high dimensional longitudinal health histories using relational random forests. *Statistical Analysis and Data Mining*, 8:128–136.

Shehzad, Z., Kelly, A., Reiss, P., Gee, D., Gotimer, K., Uddin, L., Lee, S., Margulies, D., Roy, A., Biswal, B., Petkova, E., Castellanos, F., and Milham, M. (2009). The resting brain: unconstrained yet reliable. *Cerebral Cortex*, 19(10):2209–2229.

Smock, S. A., Baker, A., Harris, K., and D'Sauza, C. (2011). The role of social support in collegiate recovery communities: A review of the literature. *Alcoholism Treatment Quarterly*, 29:35–44.

Spiegelhalter, D., Abrams, K., and Myles, J. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Son, Ltd.

Sporns, O., Chialvo, D., Kaiser, M., and Hilgetag, C. (2003). Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, 8(9):418–425.

Stalcup, S., Christian, D., Stalcup, J., Brown, M., and Galloway, G. (2006). A treatment model for craving identification and management. *Journal of Psychoactive Drugs*, 38:189–202.

Stiglic, G., Brzan, P., Fijacko, N., Wang, F., Delibasic, B., Kalousis, A., and Obradovic, Z. (2015). Comprehensible predictive modeling using regularized logistic regression and comorbidity based features. *PloS One*, 19(12):e0144439.

Strom, B. (2001). Data validity issues in using claims data. *Pharmacoepidemiology and Drug Safety*, 10(5):389–392.

Sun, F., Miller, L., and D'Esposito, M. (2004). Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. *Neuroimage*, 21(2):657–658.

Thomas Yeo, B., Krienen, F., Sepulcre, J., Sabuncu, M., Lashkari, D., Hollinshead, M., Roffman, J., Smoller, J., Zöllei, L., Polimeni, J., Fischl, B., Liu, H., and Buckner, R. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3):1125–1165.

Tian, L., Jiang, T., Wang, Y., Zang, Y., He, Y., Liang, M., Sui, M., Cao, Q., Hu, S., Peng, M., and Zhuo, Y. (2006). Altered resting-state functional connectivity patterns of anterior cingulate cortex in adolescents with attention deficit hyperactivity disorder. *Neuroscience Letters*, 400(1-2):39–43.

Valdés-Sosa, P., Sánchez-Bornot, J., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., and Canales-Rodríguez, E. (2005). Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):969–981.

van Gils, A., Burton, C., Bos, E., Janssens, K., Schoevers, R., and Rosmalen, J. (2014). Individual variation in temporal relationships between stress and functional somatic symptoms. *Journal of Psychosomatic Research*, 77(1):34–39.

Wang, F., Lee, N., Hu, J., Sun, J., and Ebadollahi, S. (2012). Towards heterogeneous temporal clinical event pattern discovery: a convolutional approach. *In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 6(7):453–461.

Wang, J., Zuo, X., Gohel, S., Milham, M., Biswal, B., and He, Y. (2011). Graph theoretical analysis of functional brain networks: test-retest evaluation on short- and long-term resting-state functional MRI data. *PLOS One*, 6(7).

Webb, J., Ferguson, M., Nielsen, J., and Anderson, J. (2013). BOLD Granger causality reflects vascular anatomy. *PLOS One*, 8(12).

Weng, S., Reps, J., Kai, J., Garibaldi, J., and Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLOS One*, 12(4):e0174944.

Ye, C., Fu, T., Hao, S., Zhang, Y., Wang, O., Jin, B., Xia, M., Liu, M., Zhou, X., Wu, Q., Guo, Y., Zhu, C., Li, Y., Culver, D. S., Alfreds, S., Strearns, F., Sylvester, K., Widen, E., McElhinney, D., and Ling, X. (2017). Prospective prediction of incident hypertension within the next year using statewide electronic health records and machine learning. *JMIR Preprints*.

Zettl, U., Bauer-Steinhusen, U., Glaser, T., Czekalla, J., Hechenbichler, K., Limmroth, V., and Hecker, M. (2016). Adherence to long-term interferon beta-1b injection therapy in patients with multiple sclerosis using an electronic diary. *Advances in Therapy*, 33(5):834–847.

Zhang, H., Duan, L., Zhang, Y., Lu, C., Liu, H., and Zhu, C. (2011). Test-retest assessment of independent component analysis-derived resting-state functional connectivity based on functional near-infrared spectroscopy. *Neuroimage*, 55(2):607–615.

Zhang, T. (2006). Existence of the signal in the signal plus background model. *Recent Developments in Nonparametric Inference and Probability*, 50:144–155.

Zhen, Z., Tian, J., Qin, W., and Zhang, H. (2007). Partial correlation mapping of brain functional connectivity with resting state fMRI. *Proceedings of the SPIE: Physiology, Function, and Structure from Medical Images*, 6511.

Zheng, Y., Wiebe, R., Cleveland, H., Molenaar, P., and Harris, K. (2013). An idiographic examination of day-to-day patterns of substance use craving, negative affect, and tobacco use among young adults in recovery. *Multivariate Behavioral Research*, 48(2):241–266.

Zuo, X. and Xing, X. (2014). Test-retest reliabilities of resting-state fMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neuroscience and Biobehavioral Reviews*, 45(12):100–118.