Web-based Supplementary Materials for "Bayesian spatial transformation models with applications in neuroimaging data" by Michelle F. Miranda, Hongtu Zhu and Joseph G. Ibrahim

Web Appendix A

In this web appendix, we present a sensitivity analysis of the hyper-parameters a and b of $\lambda = \{\lambda_d; d \in \mathcal{D}\}$ and the hyper-parameters ϕ_k of $\boldsymbol{\beta}$. There are two goals. One is to examine the finite sample performance of STM and its associated parameter estimates under different scenarios. The other is to evaluate MCMC convergence through a diagnostic analysis.

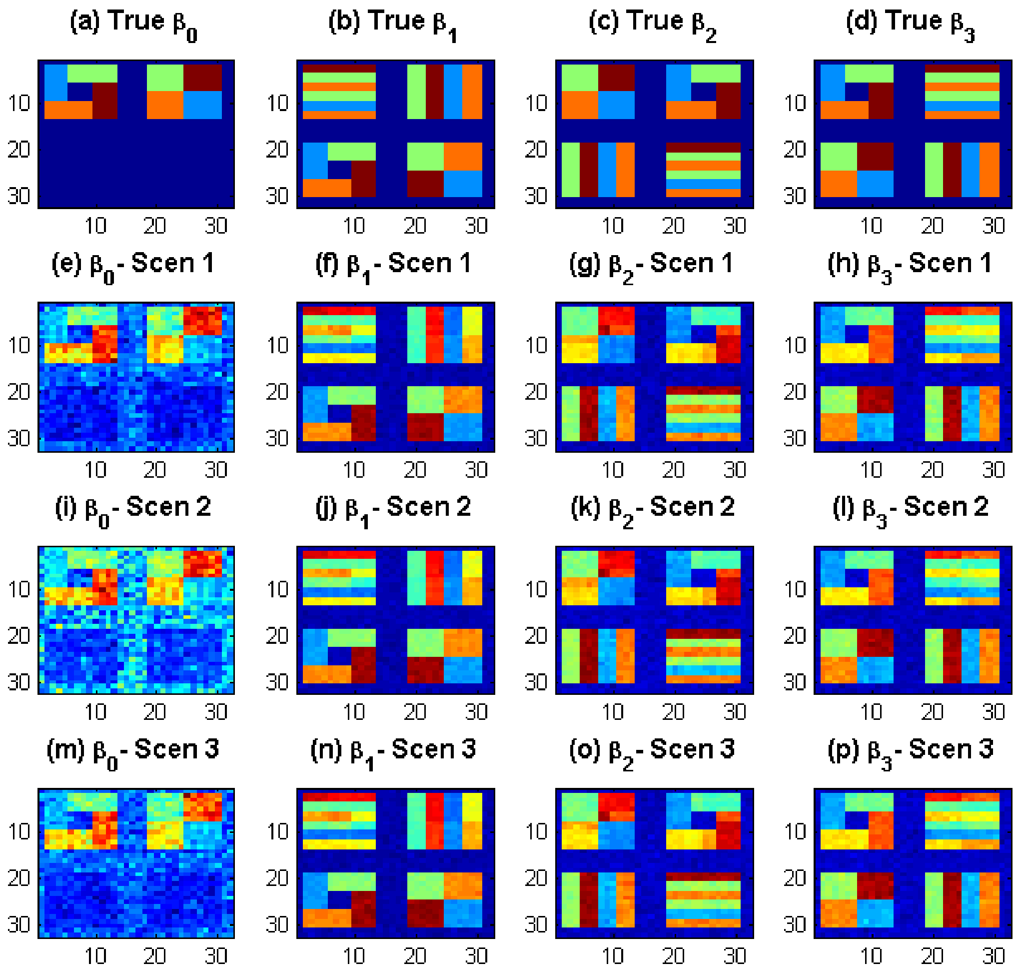
Sensitivity analysis for Λ. We consider three different scenarios for (a, b) including $(-2.0, 2.5)$, $(-3.0, 3.0)$ and $(-3.5, 3.5)$. In most applications, the three scenarios of (a, b) represent a reasonable range of λ. Although it may be desirable to use a wider interval (a, b), very flat priors can lead to slow convergence of the MCMC algorithm. We examine how STM recovers the geometric patterns presented in Section 3. Web Figure 1 reveals that regardless of the different choices of a and b, the STM is able to capture the true underlying pattern. Thus, STM is robust to the choice of the hyperparameters of λ_d.

Geweke diagnostic statistics. Under each scenario, we evaluate convergence at each voxel through the Geweke diagnosis statistics (Geweke, 1992). Web Table 1 presents the percentages of voxels, whose Geweke diagnosis statistics, computed after 1000 iterations of the Markov chain, are smaller than 1.96 (in absolute value). The numbers are shown to be very similar across the three scenarios for all parameters. Compared with other parameters, the β's associated with the indicator variables $\beta_2(d)$ and $\beta_3(d)$ have a smaller proportion of voxels that converge.

Trace plots for ν_k. We present the trace plots for the parameters ν_k associated with each β_k under the three scenarios in Web Figure 2. Web Figure 2 reveals that the MCMC chains converge fast and the posterior estimates of ν_k converge to their true values.

Trace plots of $\boldsymbol{\beta}$, τ_σ and λ. Web Figure 3 presents the trace plots of $\boldsymbol{\beta}$, τ_σ and λ for the scenario $(a, b) = (-3.5, 3.5)$ in some selected voxels. For the sake of space, we omit their trace plots for other scenarios and voxels, since they are essentially similar to each other. Web Figure 3 reveals that the single-site Gibbs sampler algorithm has good convergence properties.

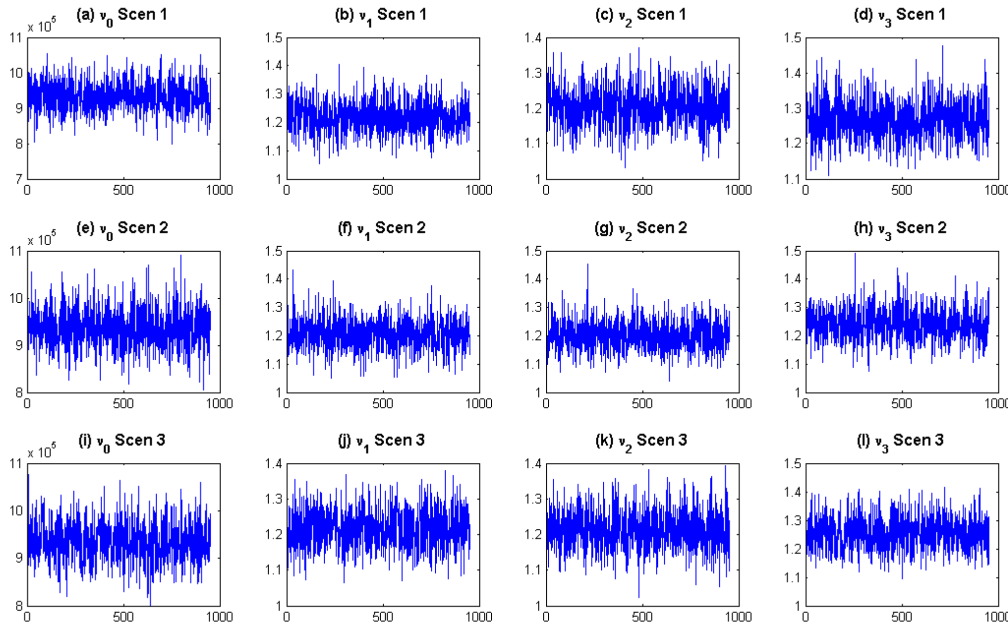
Sensitivity analysis for ϕ_k. There are two strategies of determining ϕ_k. First, for small and moderate N_D, it is possible to integrate ϕ_k into the Gibbs sampler by sampling from



Web Figure 1: Sensitivity analysis of Λ. Panels (a)-(d) represent the true pattern of $\boldsymbol{\beta}$ used to generate the images; (e)-(h): the posterior means of $\boldsymbol{\beta}$ obtained with $(a, b) = (-2.0, 2.5)$; (i)-(l): the posterior means of $\boldsymbol{\beta}$ obtained with $(a, b) = (-3.0, 3.0)$; and (m)-(p): the posterior means of $\boldsymbol{\beta}$ with $(a, b) = (-3.5, 3.5)$.

Web Table 1: Sensitivity analysis for λ indicating the percentages of voxels, whose Geweke diagnosis statistics are smaller than 1.96, according to the Geweke diagnosis statistics (Geweke, 1992) for each scenario considered.

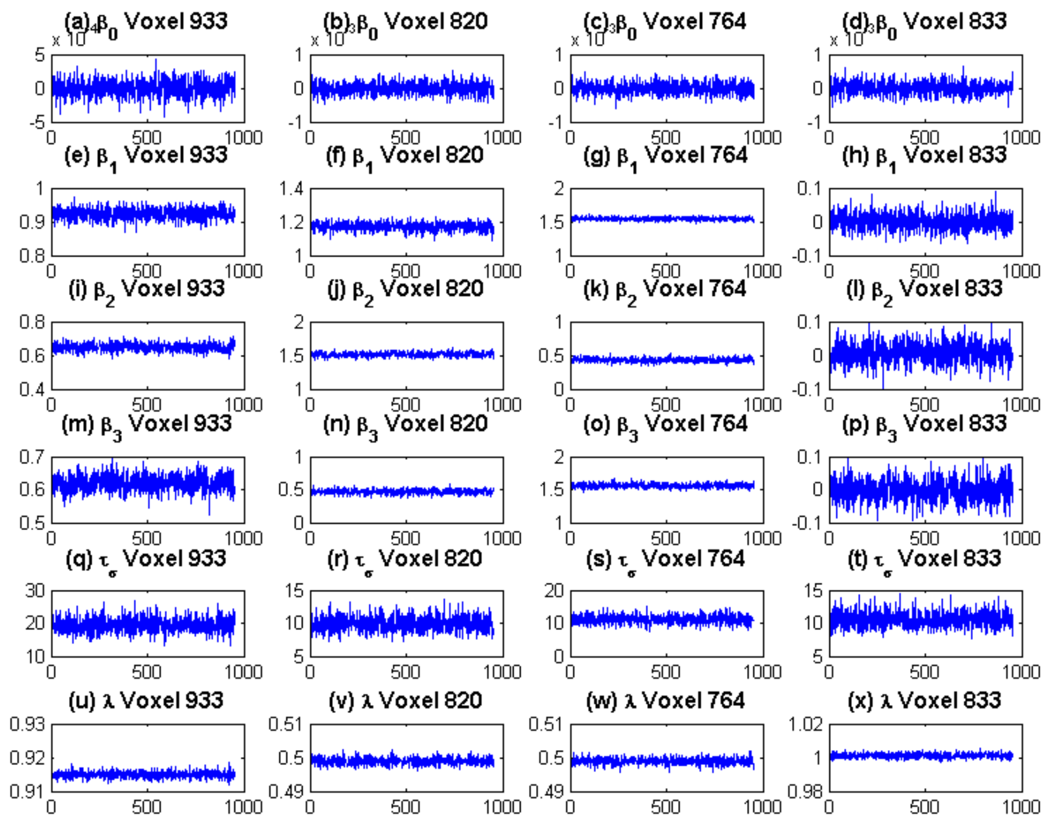
Scenario	$(-2.0, 2.5)$	$(-3.0, 3.0)$	$(-3.5, 3.5)$
λ	96.09	95.21	96.00
τ	94.53	94.63	93.65
β_0	94.73	94.92	94.34
β_1	93.55	94.63	94.92
β_2	89.06	88.48	89.65
β_3	89.06	88.28	89.94



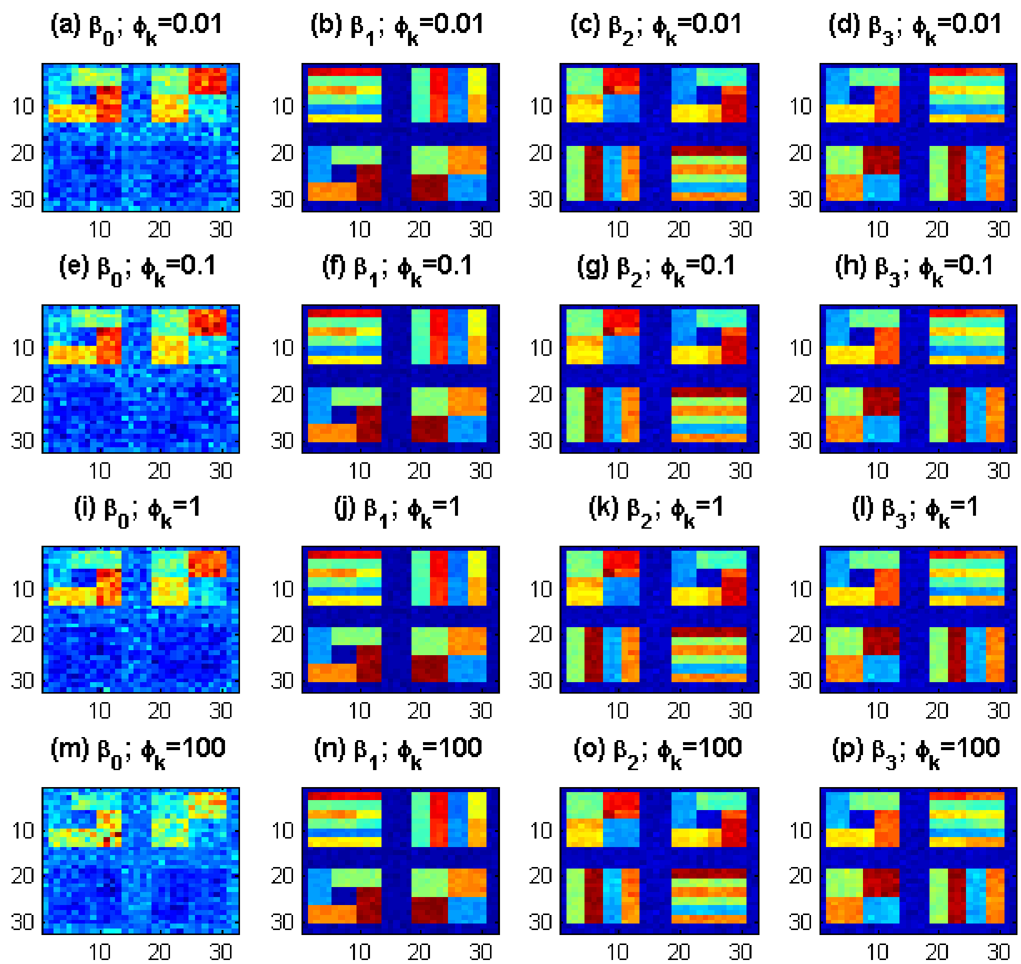
Web Figure 2: Trace plots of ν_k for $k = 0, \ldots, 3$ after a burn-in of 50 iterations and a total of 1000 MCMC iterations under the three scenarios of (a, b). Rows 1-3 correspond to $(a, b) = (-2.0, 2.5)$, $(a, b) = (-3.0, 3.0)$, and $(a, b) = (-3.5, 3.5)$, respectively.

the full conditional distribution of ϕ_k, which is proportional to $p(\beta_k|\nu_k, \phi_k)p(\phi_k)$, where $p(\phi_k)$ is the prior of ϕ_k. Different choices of ϕ_k have been discussed in Ferreira and De Oliveira (2007). The conditional distribution for ϕ_k does not have a simple form, but it can be easily sampled using the slice sampler (Neal, 2003). Sampling ϕ_k requires the computation of the eigenvalues of a sparse $N_D \times N_D$ matrix H_k. For an extremely large N_D, calculating the eigenvalues of H_k can be computationally infeasible. Second, it is common to pre-specify ϕ_k in many applications. Thus, it is important to evaluate the effects of different hyperparameters ϕ_k on parameter estimates.

Inspecting Web Figure 4 reveals that as ϕ_0 increases, the posterior estimates of β_0 get worse, whereas there is no visual difference for other parameter estimates under different ϕ_k. It is expected that the estimation of the model parameters becomes more and more difficult when ϕ_k is large, since the effective sample size decreases as the correlation among observations increases. It results in a decrease of useful information about the parameters of interest, which is contained in the data (Ferreira and De Oliveira, 2007).



Web Figure 3: Trace plots of $\boldsymbol{\beta}$, τ_σ and λ for the scenario $(a, b) = (-3.5, 3.5)$ at 4 random selected voxels. The results show fast convergence of the MCMC chain for all parameters.



Web Figure 4: Posterior estimates of $\boldsymbol{\beta}$ for different values of ϕ_k: (a)-(d): $\phi_k = 0.01$; (e)-(h) $\phi_k = 0.1$; (i)-(l) $\phi_k = 1$; and (m)-(p): $\phi_k = 100$.

Web Appendix B

The goal in this subsection is to examine the effects of different H_k on the parameter estimates of $\boldsymbol{\beta}$. Recall that H_k is given by

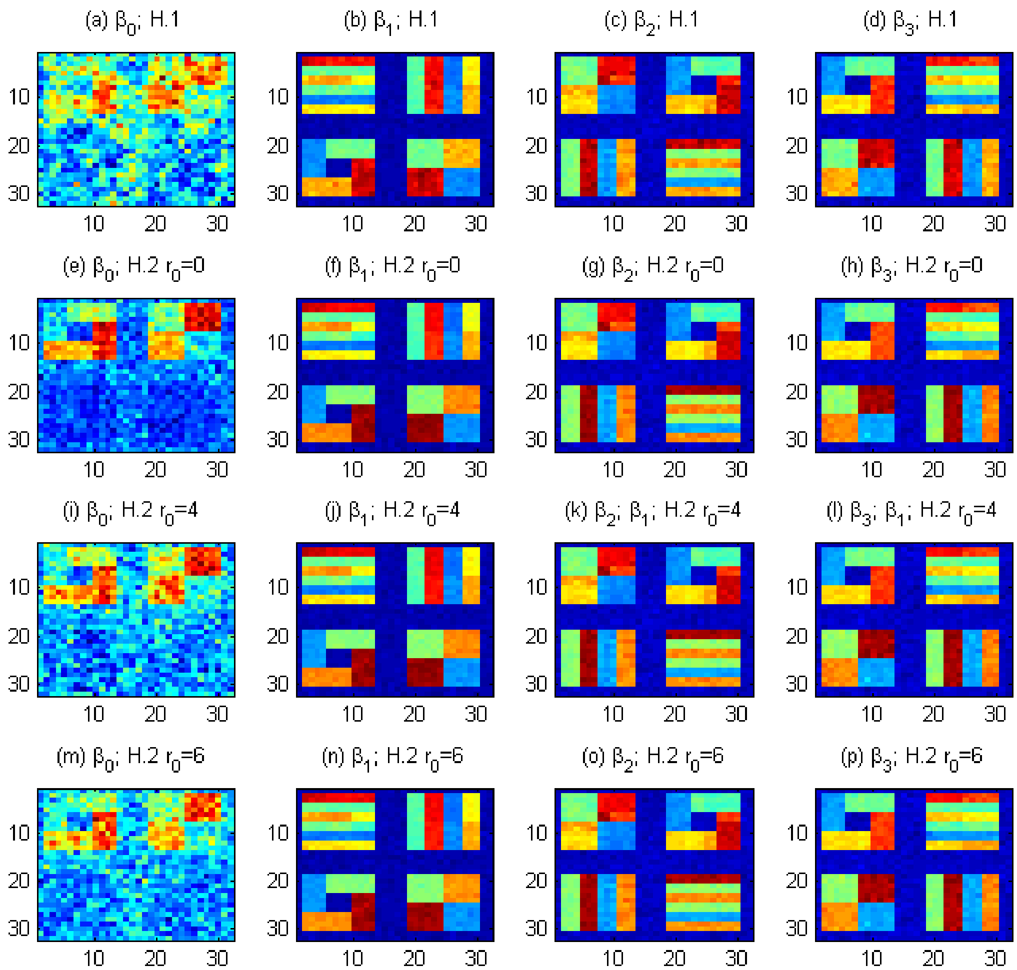
$$h_k(d,d') = \begin{cases} \sum_{d' \in N(d)} \omega_k(d,d')^2, & \text{for } \ d = d', \\ -\omega_k(d,d')^2 \mathbf{1}(d' \in N(d)), & \text{for } \ d \neq d'. \end{cases}$$

Throughout the paper we consider $\omega_k(d,d') = K(||d - d'||_2)$, where $K(u) = \exp\left(-\frac{1}{2}u^2\right)\mathbf{1}(u \leq 2)$. The following possibilities are considered here:

(H.1) A constant kernel $K(u) = \mathbf{1}(u \leq 2)$, meaning all neighbors of the voxel d are given the same weights, $\omega_k(d,d') = \mathbf{1}(||d - d'||_2 \leq 2)$;

(H.2) The Gaussian kernel $K(u) = \exp\left(-\frac{1}{2}u^2\right)\mathbf{1}(u \leq r_0)$, for $r_0 = 0, 4, 6$.

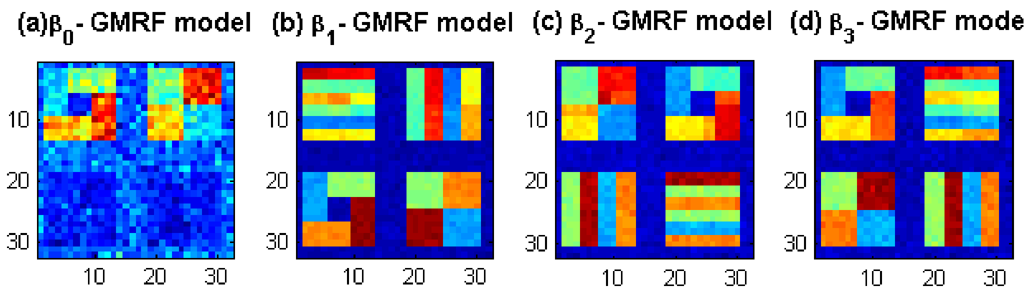
Other hyperparameters were chosen as described in Section 3 of the paper. As shown in Web Figure 5, the corresponding estimates of the intercept in all cases are quite poor, whereas the estimates for the remaining parameters are quite accurate for all cases considered.



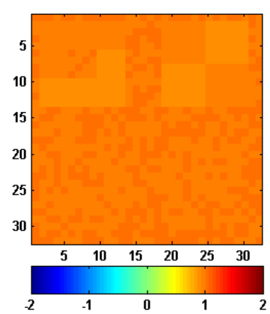
Web Figure 5: Posterior estimates of $\boldsymbol{\beta}$ under different specifications of H_k. Panels (a)-(d): case (H.1); (e)-(p): case (H.2) for $r_0 = 0, 4, 6$.

Web Appendix C

Our goal is to examine the parameter estimates obtained from STM, when the true model corresponds to $\lambda_d = 1$ for all $d \in \mathcal{D}$. Web Figure 6 shows that the STM can reliably recover the true pattern in the $\boldsymbol{\beta}$ images. Web Figure 7 reveals that the estimated λ_d's are close to the true value 1.



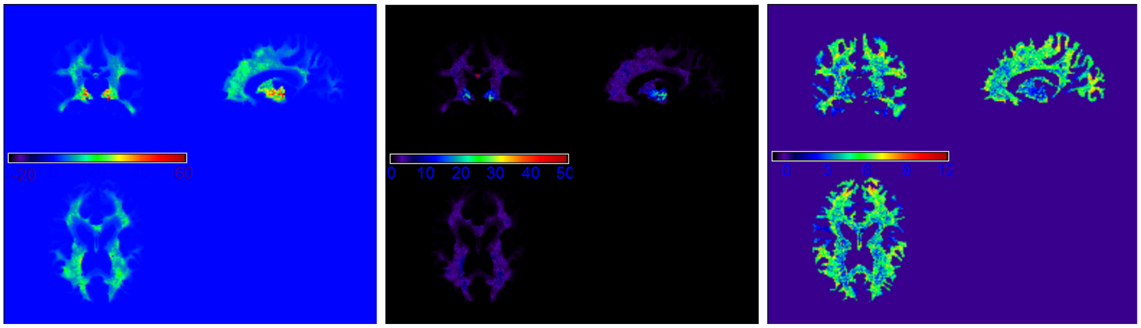
Web Figure 6: The posterior estimates of $\boldsymbol{\beta}$ for the true model with $\lambda_d = 1$ for all $d \in \mathcal{D}$.



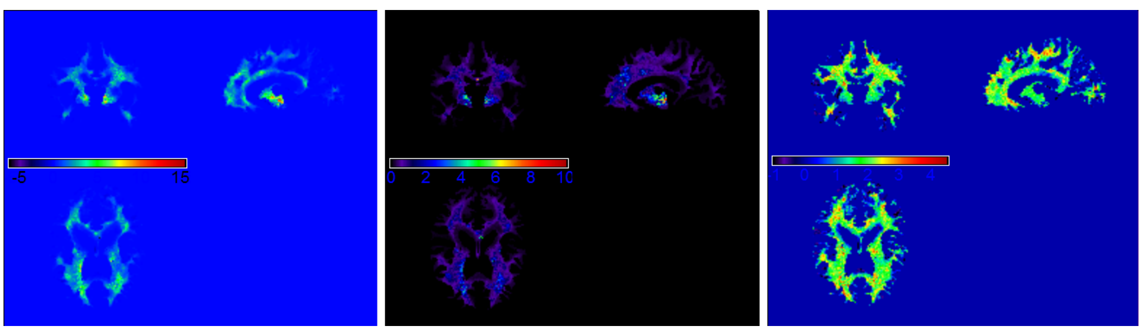
Web Figure 7: The posterior estimated image $\hat{\Lambda} = \{\hat{\lambda}_d : d \in \mathcal{D}\}$ for the true underlying model with $\lambda_d = 1$ for all $d \in \mathcal{D}$.

Web Appendix D

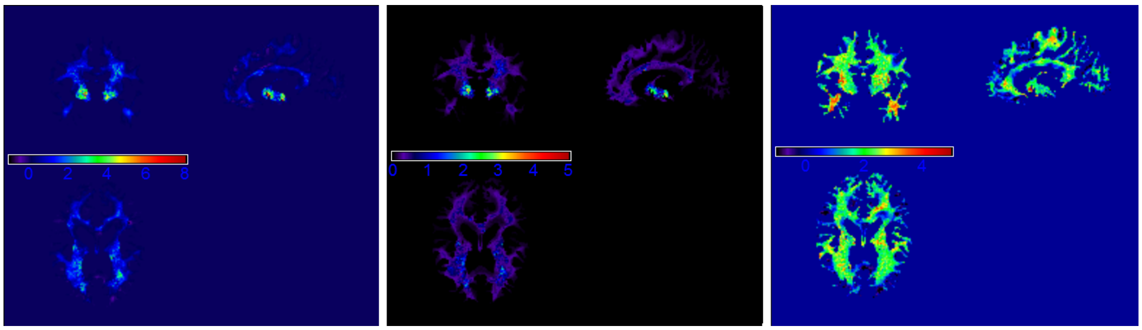
In this subsection we present the estimated $\hat{\beta}_k$ images for the intercept, gender, age, and ADHD status. Web Figures 8, 9, 10 and 11, respectively, show the results. The maps include the posterior mean, the standard deviation and the standardized images given by $\hat{\beta}_k/\widehat{\mathrm{std}}(\hat{\beta}_k)$.



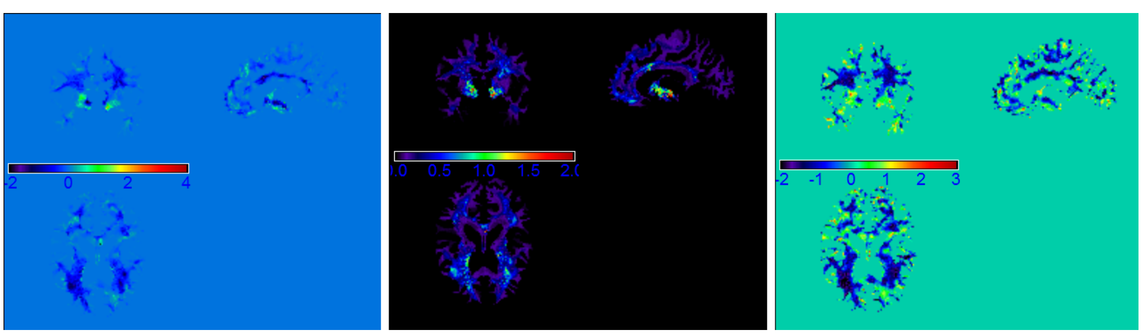
Web Figure 8: The posterior mean, the posterior standard deviation (SD), and the standardized value (mean/SD) images corresponding to the intercept β_0 are shown from the left to the right, respectively.



Web Figure 9: The posterior mean, the posterior standard deviation (SD), and the standardized value (mean/SD) images corresponding to the gender β_1 are shown from the left to the right, respectively.



Web Figure 10: The posterior mean, the posterior standard deviation (SD), and the standardized value (mean/SD) images corresponding to the age β_2 are shown from the left to the right, respectively.



Web Figure 11: The posterior mean, the posterior standard deviation (SD), and the standardized value (mean/SD) images for the ADHD status β_3 are shown from the left to the right, respectively.

References

Ferreira, M. A. R. and De Oliveira, V. (2007). Bayesian reference analysis for gaussian markov random fields. *Journal of Multivariate Analysis* **98,** 789–812.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bernardo, J. M., Berger, J., Dawid, A. P., and Smith, J. F. M., editors, *Bayesian Statistics 4*, pages 169–193. Oxford University Press, Oxford.

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics* **31,** pp. 705–741.