Predicting the Brain Response to Treatment using a Bayesian Hierarchical Model

with Application to a Study of Schizophrenia

Ying Guo¹*, F. DuBois Bowman¹ and Clinton Kilts²

¹Department of Biostatistics, The Rollins School of Public Health, Emory University, Atlanta, GA
²Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA

Abstract

In vivo functional neuroimaging, including functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), is becoming increasingly important in defining the pathophysiology of psychiatric disorders such as schizophrenia, major depression and Alzheimer’s disease. Furthermore, recent studies have begun to investigate the possibility of using functional neuroimaging to guide treatment selection for individual patients. By studying the changes between a patient’s pre- and post-treatment brain activity, investigators are gaining insights into the impact of treatment on behavior-related neural processing traits associated with particular psychiatric disorders. Furthermore, these studies may shed light on the neural basis of response and non-response to specific treatments. The practical limitation of such studies is that the post-treatment scans offer little guidance to treatment selection in clinical settings, since treatment decisions precede the availability of post-treatment brain scans. This shortcoming represents the impetus for developing statistical methodology that would provide clinicians with predictive information concerning the effect of treatment on brain activity and, ultimately, symptom-related behaviors. We present a prediction algorithm that uses a patient’s pre-treatment scans, coupled with relevant patient characteristics, to forecast the patient’s brain activity following a specified treatment regimen. We derive our predictive method from a Bayesian hierarchical model constructed on the pre- and post-treatment scans of designated training data. We perform estimation using the expectation-maximization (EM) algorithm. We evaluate the accuracy of our proposed prediction method using K-fold cross-validation, quantifying the error using two new measures that we propose for neuroimaging data. The proposed method is applicable to both PET and fMRI studies. We illustrate its use with a PET study of working memory in patients with schizophrenia and an fMRI data example is also provided.

Key words: Bayesian hierarchical model; fMRI; K-fold cross-validation; PET; Prediction; Post-treatment; Schizophrenia; Treatment decision; Treatment response

INTRODUCTION
Functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) play important roles in defining the neural basis of illness and of risk factors for major psychiatric disorders [Machulda et al., 2001; Eckert et al., 2005; Hirao et al., 2005; Whalley et al., 2006]. However, fMRI and PET study outcomes have had limited clinical translational significance with respect to informing treatment decisions for psychiatric, neurological, or substance abuse disorders. Providing objective individualized information at a neural processing level that aids the selection of an optimal treatment plan is particularly appealing for psychiatric disorders such as schizophrenia. Selecting the best treatment course (from many options) for a patient with schizophrenia is challenging given that the margins of pathophysiology are hard to define on the traditional measures of behavior and self report. Frequently, schizophrenia patients must try multiple medications prior to identifying one that is deemed successful. Another complication of schizophrenia is the high rate of treatment nonadherence, which often occurs because patients experience treatment-related side effects or perceive poor psychiatric symptom responses and thus terminate treatment [Liu-Seifert et al., 2005]. Discontinuing medication is especially problematic for schizophrenia patients because there is an estimated 80% relapse rate among those who stop taking their medications [Hogarty and Ulrich, 1998].

The challenges in selecting appropriate pharmacotherapy treatments for schizophrenia patients were illuminated by a recent large-scale study sponsored by the National Institute of Mental Health (NIMH), called the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) [Lieberman et al., 2005]. The CATIE study revealed small to negligible differences in effectiveness, on average, between several second-generation (atypical) antipsychotics and a first-generation antipsychotic (perphenazine), providing little guidance to the managing clinician regarding treatment choices. fMRI and PET have the potential to make valuable contributions to selecting individualized treatment plans for patients with schizophrenia and other psychiatric disorders. The motivating example for our research comes from a study of working memory among patients with schizophrenia, but other psychiatric disorders, such as
major depression, would benefit from the development of similar statistical approaches for neuroimaging outcomes to aid in treatment decisions.

Numerous studies identify distinct patterns of resting- and task-related neural activity in local brain regions among subjects with psychiatric disorders as compared to a healthy comparison population that support a neuropathophysiology of cognitive/behavioral problems associated with these disorders [Ojeda et al. 2002; Crespo-Facorro et al. 2001; Callicott et al. 2003; Honey et al. 2003; Sabri et al. 2003; Walter et al. 2003; Kameyama et al. 2005; Harvey et al. 2005]. Furthermore, several studies report significant alterations in specific brain regions following psychiatric treatments [Lee et al. 2005; Abdullaev et al. 2002; Suckling and Bullmore 2004]. In many cases, post-treatment follow-up scans reveal the treatment-related reduction or elimination of baseline differences in neural activity between psychiatric patients and normal controls. These collective findings strongly implicate alterations in neural activity in selective brain regions in the pathophysiology of psychiatric disorders and their responses to medical treatment. This provides motivation for and suggests the potential utility of forecasting treatment-related neural responses reflected in PET or fMRI.

Some investigators have sought to link pre-treatment brain activity in specific regions to the eventual clinical response of psychiatric symptoms to treatment [Bruder et al. 2002; Rauch et al. 2001; Brannan et al., 2000; Mayberg, 1997, 2003; Takana et al. 2004; Richardson et al. 2004; Seminowicz et al. 2004]. These studies use various statistical methods such as multiple regression and path analysis to predict the symptom response to treatment based on baseline brain scans. In a related manner, several authors [Moresco et al., 2000; Broady et al., 1999; Lee et al., 2005, Goldapple et al., 2004; Kennedy et al., 2001; Mayberg et al., 2000, 2001] establish associations between treatment response and pre- to post-treatment changes in brain activity. In clinical practice, the insights to be gained from evaluating both baseline and post-treatment scans are offset due to the unavailability of post-treatment scans at the time that a clinician makes treatment decisions for a particular patient. This pragmatic shortcoming suggests
the utility of developing a statistical framework to predict treatment-related brain alterations, which could then combine with baseline scans and patient history to help inform clinical decision-making.

In this paper, we present a framework to predict the (task-specific) brain response to treatment from a patient’s pre-treatment scans, circumventing the aforementioned practical limitation of the unavailability of post-treatment brain scans for clinical decisions regarding treatment selections. Our objective here is distinct from previous work in that we seek to predict post-treatment brain activity, rather than predicting a clinical symptom response to treatment. Our goal is therefore to reduce the wide heterogeneity of response of behavioral symptoms to treatment by use of an intermediate measure of their neural information processing corollaries. Our work described here represents an important first step in attempting to aid treatment decisions by using functional neuroimaging data.

The remainder of this paper is organized as follows. We first describe a PET study of working memory among patients with schizophrenia that serves as the motivating example for our work. In the Methods section, we describe our proposed modeling and prediction frameworks, and we discuss differences in the prediction methods for PET and fMRI data. To accurately evaluate the performance of the proposed prediction algorithm, we use a K-fold cross-validation approach and propose two new measures to quantify prediction errors. The Results section illustrates the application of our methods, using the working memory study, and discusses the performance of our approach. An fMRI data example is also provided. Finally, we summarize our methods and discuss potential extensions for utilizing the predicted post-treatment maps to guide treatment selection in clinical practice.

**DESCRIPTION OF EXPERIMENTAL DATA**

To motivate and to illustrate the use our prediction algorithm, we use an $^{15}$O$\text{H}_2$ PET regional cerebral blood flow (rCBF) study examining the neural representations of working memory in schizophrenia patients. Schizophrenia is a serious mental illness associated with a range of disturbances
of cognition, social behavior and affect. One important cognitive function often affected by schizophrenia is working memory, which is characterized by the ability to organize and manipulate information in short-term memory. A recent PET study identified distributed patterns of neural processing underlying working memory deficits in a group of patients and investigated changes in these patterns associated with pharmacotherapy [Kilts et al., 2006]. Our work here focus on 16 male schizophrenia patients from this recent study who were randomized to a double blind, flexible dose antipsychotic drug treatment with either risperidone or olanzapine. The patients were acutely ill and met the DSM-IV criteria for schizophrenia (using the SCID-IP [First et al., 2002]) or schizophrenia affective disorder. The study captured task-related neural activity using PET both prior to pharmacological treatment and following a 12-week treatment regimen. The patients were withdrawn from antipsychotic medications (haloperidol, olanzapine or risperidone) and anticholinergic medications for at least 2 days (3.6 ± 2 days) preceding initial PET imaging.

Working memory was engaged by four load levels of a Paced Auditory Serial Addition Test (PASAT), comprised of a digit shadowing (DS) task and three serial addition tasks. The patients received an auditory presentation of 50 random single digit positive integers via headphones. In the DS condition, the patients were instructed to merely verbally report the digit they heard, essentially requiring no use of working memory. In the three active tasks, subjects were instructed to compute and verbalize the sum of the current and the preceding number, which involved the need to mentally store and manipulate several pieces of information simultaneously. The magnitude and range of the integer values in the serial addition tasks distinguished the three active conditions, with numbers falling between 1 and 3 in the low load condition, between 1 and 5 in the moderate load, and between 1 and 9 in the high load condition. Each of the four load conditions was presented twice in pseudo-random order. The patients who participated in this study provided written informed consent, and the study protocol was approved by the Emory University Investigational Review Board and a Radiation Safety Committee.
Each participant was scanned eight times, twice for each of four load conditions. rCBF measures were determined from 90 second single-frame images acquired following the bolus intravenous infusion of 35 mCi of $[^{15}\text{O}]$H$_2$ presented ten seconds following the onset of a stimulus set [Mazziotta et al., 1985]. Scanner onset was coincident with the detection of cerebral radioactivity. $[^{15}\text{O}]$H$_2$ injections were made at 10-min intervals. Images were reconstructed using a measured attenuation correction (68Ge/68 Ga transmission scans). PET image postprocessing prior to the statistical analysis was conducted using SPM99 (Wellcome Department of Cognitive Neurology, [http://fil.ion.ucl.ac.uk/spm](http://fil.ion.ucl.ac.uk/spm)). PET images for each subject were aligned and resliced [Woods et al., 1998a] and spatially normalized to a population-representative PET atlas [Woods et al., 1998b] centered in Talairach stereotaxic coordinates [Talairach and Tournoux, 1988], smoothed with a 9 mm Gaussian filter, and normalized for global blood flow.

**METHODS**

A subject's post-treatment brain activity depends on various factors including type and duration of treatment, extent and type of treatment response, his/her pre-treatment baseline brain function and relevant subject-specific characteristics such as family medical history and genotype. Our aim is to develop an algorithm that utilizes information related to the above factors to provide accurate predictions for post-treatment brain function. We develop the prediction algorithm by first building a statistical model that describes the interrelationship between a subject's post-treatment and pre-treatment patterns of task-related neural processing. We estimate the model using a training data set that includes both pre-treatment and post-treatment scans. We then derive a prediction algorithm for the post-treatment brain activity based on the estimated statistical model. Once the prediction algorithm is developed, clinical practitioners can apply it to predict the post-treatment brain function for a new patient by inputting information derived from the new patient's pre-treatment scans and relevant
individualized traits. With the predicted post-treatment brain activity, clinical practitioners may find helpful information in deciding whether a treatment is appropriate for a particular subject or which of several treatment choices appears to provide the best probability of therapeutically altering a subject's brain function. We summarize our proposed approach for developing and applying the prediction algorithm in Figure 1. Prior to presenting our prediction algorithm, we describe the underlying statistical model that leads to the formulation of our prediction framework.

**Bayesian hierarchical model**

We derive our predictive framework for treatment-related neural responses from a Bayesian hierarchical model constructed using pre- and post-treatment brain scans. The first level models within-subject scan-specific activation effects associated with various experimental conditions in terms of subject-specific effects, and the second level expresses the subject-specific effects in terms of population parameters. We use the Expectation-Maximization (EM) algorithm to estimate the parameters in the model. We then demonstrate how to derive the prediction algorithm based on the estimated Bayesian hierarchical model.

Our Bayesian hierarchical model is applicable to both PET and fMRI studies, with only slight modifications at the first-stage required to accommodate these modalities. Since our motivating example is a PET study, we present the first stage model for PET data but also comment on the alterations required for applications to fMRI data. To set notation, let \( i = 1, \ldots, N \) index subjects, \( s = 1, \ldots, S \) represent scans, \( j = 1, 2 \) represent treatment period, and \( v = 1, \ldots, V \) index voxels. Also, let \( S_j \) denote the number of scans under each treatment period, with \( S = S_1 + S_2 \). We arrange the rCBF measurements for subject \( i \), measured at voxel \( v \), in the \( S \times 1 \) vector \( Y_i(v) = (Y_{i1}(v), Y_{i2}(v))' \), where \( Y_{i1}(v) \) and \( Y_{i2}(v) \) are pre- and post-treatment rCBF measurements, respectively. The
$S_j \times q$ matrices $X_{ij}^{(l)}$ ($j = 1, 2$) include independent scan-specific variables of interest for prediction such as experimental conditions at each treatment period. Covariates that are not of substantive interest are contained in $Z_{ij}^{(l)}$ ($j = 1, 2$).

The first-stage of the Bayesian hierarchical model resembles a general linear model (GLM) for each individual’s vector of responses at voxel $v$ and is specified as follows

$$
\begin{bmatrix}
Y_{i1}(v) \\
Y_{i2}(v)
\end{bmatrix} = 
\begin{bmatrix}
X_{i1}^{(l)} & 0 \\
0 & X_{i2}^{(l)}
\end{bmatrix} 
\begin{bmatrix}
B_{i1}(v) \\
B_{i2}(v)
\end{bmatrix} + 
\begin{bmatrix}
Z_{i1}^{(l)} & 0 \\
0 & Z_{i2}^{(l)}
\end{bmatrix} 
\begin{bmatrix}
a_{i1}(v) \\
a_{i2}(v)
\end{bmatrix} + 
\begin{bmatrix}
\mathbf{\epsilon}_{i1}(v) \\
\mathbf{\epsilon}_{i2}(v)
\end{bmatrix}
$$

(1)

where $B_{ij}(v)$ ($j = 1, 2$) is the $q \times 1$ vector of subject-specific effects of interest corresponding to $X_{ij}^{(l)}$,

$a_{ij}(v)$ contains effects related to $Z_{ij}^{(l)}$ and hence represents parameters that are not of main interest. We assume that the vector of random errors $\mathbf{\epsilon}_{i}(v) = [\mathbf{\epsilon}_{i1}^T(v), \mathbf{\epsilon}_{i2}^T(v)]^T$ follows a zero-mean multivariate normal distribution $N(0, \Sigma_i^{(l)}(v))$.

For fMRI studies, the dependent variable $Y_{i}(v)$ usually represents serial blood oxygenation-level dependent (BOLD) responses. The covariate matrix $Z_{ij}^{(l)}$ typically includes the high-pass filtering matrix used to remove unwanted low-frequency trends from the fMRI data. The design matrices of interest $X_{ij}^{(l)}$ are convolved with a hemodynamic response function. We also assume that the fMRI serial responses have been pre-whitened to remove the serial correlations among multiple scans taken across time, but one could easily incorporate serial correlations into the model if they are present.

Following Friston et al. [2002], we parameterize the covariance matrix $\Sigma_i^{(l)}(v)$ in (1) as a linear combination of covariance components, i.e. $\Sigma_i^{(l)} = \sum_k \lambda_k^{(l)} Q_k^{(l)}$, where $\lambda_k^{(l)}$ are covariance hyperparameters and $Q_k^{(l)}$ are basis set matrices that are constructed as constraints for the covariance matrix. The matrices $Q_k^{(l)}$ can be viewed as design matrices that are chosen to model selected forms of
nonsphericity. For our model, we assume that random errors are independent from each other because subject-specific parameters $\mathbf{B}_j(v)$ help account for correlations among within-subject images and the serial correlations present in fMRI data are already removed through pre-whitening. We also assume that the variability of the random errors is different for pre- and post-treatment periods. The covariance structure based on the above two assumptions is modeled with two bases $\mathbf{Q}^{(1)}_1 = \begin{bmatrix} \mathbf{I}_{S_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ and $\mathbf{Q}^{(1)}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{S_2} \end{bmatrix}$, which are associated with pre- and post-treatment variance hyperparameters $\lambda^{(1)}_1$ and $\lambda^{(1)}_2$, respectively. If correlations are present between the repeated scans on an individual, for example in fMRI data that are not pre-whitened, one could include an additional basis set to capture serial correlations in the random errors. Note that the first-stage model specifies different parameters $\mathbf{B}_j(v) (j = 1, 2)$ for the pre- and post-treatment scans, which allows us to capture heterogeneity in these effects across treatment periods. The model also accommodates situations where the treatment effects may vary across experimental conditions. For example, treatment effects may only exist for task-related activation conditions but not for the resting/control condition.

At the second-stage, we focus on modeling the subject-specific effects of interest in terms of population parameters that code effects for treatment assignment and potentially other relevant patient characteristics (e.g., genotype, trauma history). The second-stage model is specified as

$$
\begin{bmatrix}
\mathbf{B}_{n1}(v) \\
\mathbf{B}_{n2}(v)
\end{bmatrix} =
\begin{bmatrix}
\mathbf{X}^{(2)}_{n1} & \mathbf{0} \\
\mathbf{0} & \mathbf{X}^{(2)}_{n2}
\end{bmatrix}
\begin{bmatrix}
\mathbf{\beta}_1(v) \\
\mathbf{\beta}_2(v)
\end{bmatrix} +
\begin{bmatrix}
\mathbf{e}^{(2)}_{n1}(v) \\
\mathbf{e}^{(2)}_{n2}(v)
\end{bmatrix},
$$

(2)

where $\mathbf{X}^{(2)}_{nj}$ contains subject-specific covariates including treatment assignment and other relevant patient characteristics such as medical or family history that may influence the treatment response; $\mathbf{\beta}_j(v)$ consists of the population-level effects associated with $\mathbf{X}^{(2)}_{nj}$.
We use the working memory study to demonstrate how to specify the parameters in the second-stage model to examine the effects of relevant patient characteristics. In the working memory study, patients were randomized to either of two antipsychotic drug treatment arms: A (risperidone) and B (olanzapine). We are interested in the pre- to post-treatment alterations in the brain activity for each of the treatments. Our second-stage model is then specified as

\[
\begin{bmatrix}
    B_{ii}(v) \\
    B_{i2}(v)
\end{bmatrix} = \begin{bmatrix}
    I_{2q} & 0 & 0 \\
    0 & I_{i} & (1-I_{i})I_{q}
\end{bmatrix} \begin{bmatrix}
    \beta_1(v) \\
    \beta_{2A}(v) \\
    \beta_{2B}(v)
\end{bmatrix} + \begin{bmatrix}
    \varepsilon_{i1}^{(2)}(v) \\
    \varepsilon_{i2}^{(2)}(v)
\end{bmatrix},
\]

where \( \beta_1(v) \) contains pre-treatment population-level effects, \( I_{iA} \) is a treatment indicator that equals 1 if subject \( i \) is randomized to treatment A and 0 for treatment B, and \( \beta_{2A}(v) \) and \( \beta_{2B}(v) \) represent post-treatment effects for treatments A and B, respectively. We do not stratify \( \beta_1(v) \) by treatment assignment because subjects are randomized into treatment groups and hence assumed to have similar brain activity at baseline, although one may empirically assess the success of randomization in practice. Our model does stratify post-treatment effects, enabling the detection of differential brain activity across treatments. Similarly, one could stratify the post-treatment effects by other relevant patient characteristics.

The random error \( \varepsilon_i^{(2)}(v) = [\varepsilon_{i1}^{(2)}(v), \varepsilon_{i2}^{(2)}(v)]^T \) represents the deviation of the subject-specific effects from the population mean, and is assumed to follow a zero-mean multivariate normal distribution with the covariance matrix \( \Sigma^{(2)}(v) \), which again can be expressed as a linear combination of covariance components, i.e. \( \Sigma^{(2)}(v) = \sum_k \lambda^{(2)}_k Q^{(2)}_k \). For the working memory study, the covariance structure is specified as follows,

\[
\Sigma^{(2)}(v) = \begin{bmatrix}
    \lambda^{(2)}_{i1}(v)I_q & \lambda^{(2)}_{i2}(v)I_q \\
    \lambda^{(2)}_{i2}(v)I_q & \lambda^{(2)}_{i1}(v)I_q
\end{bmatrix},
\]
where \( \lambda^{(2)}_1(v) \) and \( \lambda^{(2)}_2(v) \) are the variance for the pre- and post-treatment subject-specific effects, and \( \lambda^{(2)}_{12}(v) \) is the covariance between pre- and post-treatment periods. The covariance parameter \( \lambda^{(2)}_{12}(v) \) allows for subject-specific effects to be correlated between the two treatment periods.

One can also specify a second-stage model to model the other subject-specific parameters \( \alpha_i(v) = [\alpha_{i1}(v), \alpha_{i2}(v)]^T \) in the first-stage model (1) in terms of population parameters \( \alpha(v) \) using a similar model as (2). In practice, we are often not interested in making subject-specific inferences for the effects related to \( Z^{(1)}_{ij} \). Hence, we can consider a degenerate second-stage model for \( \alpha_i(v) \) where the error covariance matrix for \( \alpha_i(v) \) is zero. In this case, \( \alpha_i(v) \) for each subject equals its population mean \( \alpha(v) \). Our following presentation will utilize this approach.

For the population level parameters \( \beta(v) \) and \( \alpha(v) \), we specify prior distributions given by 
\[
N(\eta_\beta(v), \Sigma_\beta(v)) \text{ and } N(\eta_\alpha(v), \Sigma_\alpha(v)),
\]
where \( \Sigma_\beta(v) \) and \( \Sigma_\alpha(v) \) are diagonal covariance matrices. If information regarding \( \beta(v) \) or \( \alpha(v) \) is available from previous studies, one can take a full Bayesian approach and treat the population parameters as known, i.e. by setting \( \Sigma_\beta(v) = 0 \) and \( \Sigma_\alpha(v) = 0 \). In practice, we usually have little, if any, prior information for these parameters and a parametric empirical Bayesian scheme with \( \Sigma_\beta(v) = \infty \) and \( \Sigma_\alpha(v) = \infty \) (i.e. the diagonal variance terms are \( \infty \)) [Friston et al., 2002] is appropriate. In this case, the prior expectations \( \eta_\beta(v) \) and \( \eta_\alpha(v) \) do not need to be specified (Appendix A). Therefore, the hyperparameters for our Bayesian hierarchical model only include the parameters in the covariance matrices, i.e. \( \lambda(v) = (\{\lambda^{(1)}_k(v)\}, \{\lambda^{(2)}_k(v)\})^T \), which we estimate from the data under parametric empirical Bayes scheme [Carlin and Louis, 2000; Friston et al., 2002]. To help demonstrate the overall structure of the Bayesian hierarchical model, we present a schematic illustration in Figure 2. The figure shows that the first stage of the hierarchy models brain activity measurements in terms of subject-specific parameters through design matrices containing scan-specific features. The
second stage models the subject-specific effects in terms of population parameters through design matrices containing subject characteristics. The population parameters are modulated by hyperparameters based on prior information.

The expectation-maximization (EM) algorithm [Dempster et al., 1977 and 1981] provides a convenient approach to perform estimation for the Bayesian hierarchical model and offers a way to estimate both the parameters and hyperparameters from the data. For our model, the E-step of EM algorithm involves estimating the conditional distribution of the parameters given the observed data, holding the hyperparameters $\lambda$ fixed. The M-step finds the maximum likelihood estimates of hyperparameters, i.e. $\hat{\lambda}$ that maximizes the marginal distribution $p(Y | \lambda)$, where the marginal distribution is calculated by integrating over the conditional distribution of the parameters obtained from the E-step (Appendix A). For a linear model under Gaussian assumptions such as ours and with the parametric empirical Bayes, the EM algorithm outputs the expectation and the covariance of the Gaussian posterior distribution of the parameters and the marginal maximum likelihood estimates of the hyperparameters (Appendix A). It is worth noting that EM algorithm is not restricted to the parametric empirical Bayes but can also be adopted for full Bayes and maximum likelihood estimation [Carlin and Louis, 2000; Friston et al., 2002].

**Prediction algorithm for post-treatment brain activity**

Based on the proposed Bayesian hierarchical model, we develop a prediction algorithm that can be used to predict task-related brain activity following a specific treatment regimen. Our prediction algorithm provides individualized post-treatment predictions of brain response by incorporating the unique information in each subject’s pre-treatment functional brain scans and other relevant personal characteristics. Due to inherent differences in the mechanism and data structure of PET and fMRI, the
prediction algorithm for these two types of studies has different representations. In the following, we present the models for PET and fMRI studies, separately.

**Prediction for PET**

In PET studies, predictions for the post-treatment rCBF measurements $Y_{i2}(v)$ use information from the pre-treatment rCBF $Y_{i1}(v)$ and subject’s characteristics such as treatment assignment and other relevant risk factors. Based on the proposed two-stage hierarchical model, we derive the following marginal distribution for the rCBF measurements:

$$
(Y'_i(v), Y''_i(v))' \sim N(X'_i(v)\beta(v) + Z_i^{(1)}\alpha(v), \Sigma^{*}(v)),
$$

where $X'_i = X^{(1)}_iX^{(2)}_i$ and $\Sigma^{*}(v) = \Sigma^{(1)}(v) + X^{(1)}_i\Sigma^{(2)}(v)X'^{(1)}_i$, with $X^{(m)}_i = \begin{bmatrix} X^{(m)}_{i1} & 0 \\ 0 & X^{(m)}_{i2} \end{bmatrix}$ for $m = 1,2$.

Similarly, we array the covariates that are not of substantive interest as $Z_i^{(1)} = \begin{bmatrix} Z^{(1)}_{i1} & 0 \\ 0 & Z^{(1)}_{i2} \end{bmatrix}$, and we stratify the parameters as $\beta(v) = \begin{bmatrix} \beta_1(v) \\ \beta_2(v) \end{bmatrix}$ and $\alpha(v) = \begin{bmatrix} \alpha_1(v) \\ \alpha_2(v) \end{bmatrix}$. The marginal covariance matrix $\Sigma^{*}(v)$ for measurements $Y_i$ is the sum of the error covariance at the first-stage model and the error covariance of $B_i$ on the second-stage model projected onto the measurement space by the first-stage design matrix.

Therefore, $\Sigma^{*}(v)$ consists of the variability from both levels of the hierarchical model. Note that if one considers a non-degenerate second-stage model for $\alpha_i$, i.e. with a non-zero error covariance matrix for $\alpha_i$, the marginal covariance matrix $\Sigma^{*}(v)$ will also include the projected error covariance of $\alpha_i$.

We then derive the prediction algorithm for the post-treatment rCBF $Y_{i2}(v)$ from the marginal distribution in (5). The prediction algorithm provides the predictive distribution of $Y_{i2}(v)$ given the pre-treatment rCBF responses $Y_{i1}(v)$,
\[
[Y_{i2}(v)| Y_{i1}(v), \beta(v), \alpha(v), \lambda(v)] \sim N(\mu_{21}(v), \Sigma_{21}(v)),
\]

with
\[
\mu_{21}(v) = [X_{r2}^{*} \beta_{2}(v) + Z_{i2}^{(1)} \alpha_{2}(v)] + \Sigma_{21}^{*}(v) \Sigma_{11}^{-1}(v) [Y_{i1}(v) - \{X_{r1}^{*} \beta_{1}(v) + Z_{i1}^{(1)} \alpha_{1}(v)\}],
\]
and
\[
\Sigma_{21}(v) = \Sigma_{22}^{*}(v) - \Sigma_{21}^{*}(v) \Sigma_{11}^{-1}(v) \Sigma_{12}^{*}(v),
\]
where \(\Sigma_{11}^{*}, \Sigma_{22}^{*}, \Sigma_{21}^{*}, \Sigma_{12}^{*}\) are the sub-matrices of \(\Sigma^{*}(v)\) that correspond to the variances and covariances of \(Y_{i1}(v)\) and \(Y_{i2}(v)\), and \(X_{nj}^{*} = X_{nj}^{(1)} X_{nj}^{(2)}\) for \(j = 1, 2\). From (7), one can see the predicted mean of the post-treatment brain activity of a subject is composed of two parts. The first part is the expected post-treatment brain map based on subject’s individual characteristics including treatment assignment and relevant risk factors such as genotype or family history, which is represented by the first additive term in (7). This part represents the population-level expectation for the subject based on his/her characteristics and subjects who share the same characteristics have the same population-level expectation. The second additive term in (7) represents the adjustment to the population-level prediction for this subject using information from his/her pre-treatment scans. More specifically, we adjust the population-level prediction for the post-treatment effect of a subject with this subject’s deviation from his/her pre-treatment population-level expectation, i.e. \([Y_{i1}(v) - \{X_{r1}^{*} \beta_{1}(v) + Z_{i1}^{(1)} \alpha_{1}(v)\}]\). The amount of adjustment depends on the variability of the pre-treatment images and the association between the pre- and post-treatment images: the lower the variability of the pre-treatment images and the higher association they have with the post-treatment images, the larger the effect the pre-treatment images have on the post-treatment prediction.

By inputting the ML estimator of the hyperparameters \(\hat{\lambda}(v)\) and the posterior mean of the parameters provided by the EM algorithm (Appendix A) into (7) and (8), we obtain the estimated conditional mean \(\hat{\mu}_{21}(v)\) and covariance matrix \(\hat{\Sigma}_{21}(v)\). The post-treatment rCBF \(Y_{i2}(v)\) are predicted using the mean of the estimated conditional distribution, i.e. \(\hat{\mu}_{21}(v)\). We can also construct a \(100(1 - \alpha)\%\) prediction.
interval for $Y_{i2}(v)$ based on the estimated conditional variance $\hat{\Sigma}_{21}(v)$. In some PET studies, multiple scans are acquired under each experimental condition, and we have a large number of scans at pre- and post-treatment periods. For these studies, we may choose to predict post-treatment subject-specific effects $B_{i2}(v)$ using a procedure that is analogues to the prediction for fMRI studies which is described in the following section.

**Prediction for fMRI**

In fMRI studies, the observed BOLD response $Y_i(v)$ reflects the combined effects of the experimental conditions and the hemodynamic response function. Therefore, we seek to make predictions for post-treatment subject-specific effects $B_{i2}(v)$ capturing the effects of the experimental stimuli on measured brain function, rather than to make predictions of BOLD responses directly. We derive the conditional distribution of the post-treatment $B_{i2}(v)$ given pre-treatment $B_{i1}(v)$ as the basis for our prediction algorithm. From our Bayesian hierarchical model, the distribution of $B_{i2}(v)$ conditional on $B_{i1}(v)$ is given by

$$
[B_{i2}(v) | B_{i1}(v), \beta(v), \hat{\lambda}^{(2)}(v)] = N(\mu_{2,1}(v), \Sigma_{2,1}(v)) ,
$$

with

$$
\mu_{2,1}(v) = X_{n2}^{(2)} \beta(v) + \Sigma_{21}^{(2)}(v)\Sigma_{11}^{(2)}(v)^{-1}(v)[B_{i1}(v) - X_{n1}^{(2)} \beta_1(v)] , \text{ and }
$$

$$
\Sigma_{2,1}(v) = \Sigma_{22}^{(2)}(v) - \Sigma_{21}^{(2)}(v)\Sigma_{11}^{(2)}(v)^{-1}(v)\Sigma_{12}^{(2)}(v) ,
$$

where $\Sigma_{11}^{(2)}(v)$, $\Sigma_{22}^{(2)}(v)$ and $\Sigma_{21}^{(2)}(v)$ are the sub-matrices in $\Sigma^{(2)}(v)$ that correspond to the variances and covariance of the pre- and post-treatment subject-specific effects, i.e. $\Sigma^{(2)}(v) = \begin{bmatrix} \Sigma_{11}^{(2)}(v) & \Sigma_{12}^{(2)}(v) \\ \Sigma_{21}^{(2)}(v) & \Sigma_{22}^{(2)}(v) \end{bmatrix}$. 


We propose to use the conditional distribution in (9) as the prediction algorithm to make predictions for $\bf{B}_{i2}(v)$ based on the pre-treatment subject-specific effects $\bf{B}_{i1}(v)$. In practice, we can use software such as SPM to fit a standard GLM based on the pre-treatment scans to calculate the OLS estimator $\hat{\bf{B}}_{i1}^{OLS}(v)$. We then input the estimated pre-treatment subject effects $\hat{\bf{B}}_{i1}^{OLS}(v)$ into our prediction algorithm to output the predicted post-treatment maps. Specifically, we predict the post-treatment subject-specific effects $\bf{B}_{i2}(v)$ using the mean of the estimated conditional distribution, i.e. $\hat{\mu}_{2,1}(v)$ and construct a $100(1 - \alpha)\%$ prediction interval for $\bf{B}_{i2}(v)$ using the estimated conditional variance $\hat{\Sigma}_{2,1}(v)$.

Considering a second-stage model as in (3) and a covariance structure as in (4), the proposed predictor $\hat{\bf{B}}_{i2}(v)$ for the $i$th subjects’ post-treatment brain activity, assuming treatment assignment of $t$ where $t = A$ or $B$, is as follows,

$$\hat{\bf{B}}_{i2}(v) = \hat{\beta}_2(v) + \frac{\hat{\lambda}_{12}^{(2)}(v)}{\hat{\lambda}_1^{(2)}(v)} [\hat{\bf{B}}_{i1}(v) - \hat{\beta}_1(v)].$$

The variance for the predicted post-treatment effects $\hat{\bf{B}}_{i2}(v)$ is

$$\hat{\Sigma}_{2,1}(v) = [\hat{\lambda}_{2}^{(2)}(v) - \hat{\lambda}_{12}^{(2)}(v)^2 / \hat{\lambda}_1^{(2)}(v)] \bf{I}_q.$$  

A potential issue that arises for the prediction of fMRI data concerns the extent to which preprocessing steps in fMRI studies such as pre-whitening and high-pass filtering may affect parameter estimates of the Bayesian hierarchical model. To reduce the effect of preprocessing on the prediction algorithm, we suggest that consistent preprocessing steps are performed for both (pre- and post-treatment) scanning sessions, for example, by selecting the same wavelength parameters for the high-pass filtering. Because the prediction algorithm is based on the conditional distribution of post-treatment effects given pre-treatment effects (as shown in equation (9)), consistent preprocessing will make our prediction algorithm more robust (with respect to preprocessing steps) and may serve to increase accuracy.
Model validation: Evaluation of prediction error

We estimate the accuracy of the proposed prediction model, applied to the working memory data, using $K$-fold cross-validation. This approach involves splitting the data into $K$ blocks, with roughly an equal number of subjects in each. At each iteration, we assign $K-1$ blocks as training data and the other block as test data. We fit the proposed Bayesian hierarchical model to the training data and then predict the post-treatment rCBF for patients in the test data using the estimated parameters from the training data. We repeat this procedure for the $K$ different training and test data assignments to compute the predicted post-treatment brain activity for all subjects. In $K$-fold cross-validation, separate data sets are used to construct the prediction algorithm and to evaluate its error. In this way, we obtain a more accurate estimate of the performance of the prediction algorithm in a clinical practice setting where predictions are made for new subjects that the model is not trained on.

We estimate the prediction error by comparing the observed and predicted post-treatment brain activity. For PET studies, the cross-validation estimate of the prediction error at voxel $v$ is

$$CV(v) = L(\{y_{i2}(v)\}, \{\hat{Y}^{(-i)}_{i2}(v)\}),$$

(12)

where $y_{i2}(v)$ is the observed post-treatment rCBF and $\hat{Y}^{(-i)}_{i2}(v)$ is the predicted post-treatment rCBF computed with the block that includes the $i$th subject removed from the training data sets. The loss function $L$ measures errors between $y_{i2}(v)$ and $\hat{Y}^{(-i)}_{i2}(v)$ for $i = 1, \ldots, N$. We discuss a couple of choices for $L$ below.
We introduce loss functions to accommodate special features of neuroimaging data. Typical choices of the loss function, in other settings, include the squared error and absolute error functions [Hastie et al., 2001]. These commonly applied loss functions are based on the absolute distance between the observed and predicted values and hence depend on the magnitude of the measurements. In functional neuroimaging studies, prediction of the post-treatment brain activity is performed on each voxel, and the prediction error is evaluated across all voxels. Since the brain activity measurements such as rCBF or BOLD have different baseline values across the brain, the squared or absolute error functions are inappropriate. We need scale-free loss functions so that the prediction error is comparable across all the voxels in the brain.

In this paper, we present two loss functions for neuroimaging data. The first loss function is the ratio of the square root of the prediction mean squared error (PMSE) to the average effects at voxel \( v \). It measures the magnitude of the PMSE relative to the average brain activity at each voxel. For PET studies, the loss function is defined as

\[
L_1 \left( \{ y_{i2} (v) \}, \{ \hat{Y}_{i2}^{(-i)} (v) \} \right) = \sqrt{\frac{1}{NS_2} \sum_{i=1}^{N} \left[ \hat{Y}_{i2}^{(-i)} (v) - y_{i2} (v) \right]^2 \left[ \hat{Y}_{i2}^{(-i)} (v) - y_{i2} (v) \right]} \cdot \frac{1}{NS_2} \sum_{i=1}^{N} \hat{V}_{i2}^{(-i)} (v) .
\]  

The second loss function is the proportion of observations whose post-treatment brain activity lies within the \( 100(1 - \alpha)\% \) prediction interval, and is defined as,

\[
L_2 \left( \{ y_{i2} (v) \}, \{ \hat{Y}_{i2}^{(-i)} (v) \} \right) = \frac{1}{NS_2} \sum_{i=1}^{N} \sum_{s=1}^{S_s} I \left[ \hat{Y}_{i2s}^{(-i)} (v) - z_{\alpha/2} \hat{\sigma} \{ \hat{Y}_{i2s}^{(-i)} (v) \} < y_{i2s} (v) < \hat{Y}_{i2s}^{(-i)} (v) + z_{\alpha/2} \hat{\sigma} \{ \hat{Y}_{i2s}^{(-i)} (v) \} \right]
\]  

(14)

where \( y_{i2s} (v) \) and \( \hat{Y}_{i2s}^{(-i)} (v) \) are the \( s \) th element in \( y_{i2} (v) \) and \( \hat{Y}_{i2}^{(-i)} (v) \), respectively, and \( \hat{\sigma} \{ \hat{Y}_{i2s}^{(-i)} (v) \} \) is the estimated standard error of \( \hat{Y}_{i2s}^{(-i)} (v) \) and is calculated as the square root of the \( s \) th diagonal element of \( \hat{\Sigma}_{2s} (v) \).
To evaluate the accuracy of the prediction model applied to fMRI studies, the cross-validation estimate of the prediction error is calculated using (12) where \( y_{12}(v) \) and \( \hat{y}_{12}^{(-i)}(v) \) are replaced with \( \hat{B}_{12}^{OLS}(v) \) and \( \hat{B}_{12}^{(-i)}(v) \). Here, \( \hat{B}_{12}^{OLS}(v) \) is the OLS estimator obtained from a standard GLM based on the post-treatment data and \( \hat{B}_{12}^{(-i)}(v) \) is the predicted post-treatment subject effects computed with the block that includes the \( i \)th subject removed from the training data sets.

In this paper, we have demonstrated how to develop, validate and apply a prediction algorithm using the working memory study of patients with schizophrenia. We derived the prediction algorithm for post-treatment brain activity based on a Bayesian hierarchical model. Once the model has been developed and validated, clinical practitioners can easily apply the algorithm to predict the post-treatment brain function for a new schizophrenia patient by inputting the new patient's pre-treatment images and relevant characteristics. To predict post-treatment maps for studies with different study designs and for patients with other psychiatric disorders such as depression, we propose the following steps:

Step 1: collect both pre-treatment and post-treatment functional neuroimaging data from subjects from a targeted population.

Step 2: follow the proposed modeling framework and inference procedure (Appendix A) to develop a prediction algorithm.

Step 3: validate the prediction algorithm with the cross-validation approach and quantify the error using the proposed measures in (13) and (14).

Step 4: if the prediction algorithm is sufficiently accurate based on clinical practice standard, we can apply it to predict post-treatment brain function maps for a new patient by simply inputting the observed pre-treatment rCBF or estimated pre-treatment subject effects \( \hat{B}_{11}^{OLS}(v) \) into the prediction algorithm. If the accuracy of the algorithm does not meet a minimum standard,
we can either go back to step 1 to collect data from more subjects or incorporate more relevant subject characteristics in the modeling procedure in step 2.

RESULTS

We illustrate the use of our prediction algorithm and the evaluation of its performance using PET rCBF data from the working memory study of patients with schizophrenia. Furthermore, we include the experimental results on an fMRI data from a study of inhibitory control in cocaine dependent men to demonstrate the performance of the proposed algorithm for fMRI.

Prediction for the PET study of working memory

At each voxel, the first-stage of the Bayesian hierarchical model expresses the proportionally scaled rCBF in terms of the subject-specific effects for the four working memory load conditions at the baseline and post-treatment periods. The second-stage models the subject-specific effects in terms of population parameters that are stratified on treatment assignment. We fit the second-stage model and covariance structure specified in (3) and (4).

Posterior probability maps for treatment effects

Using estimates from the Bayesian hierarchical model, we compare alterations in working-memory-related brain activity after treatment with either risperidone or olanzapine to illustrate the differential treatment effects. We first calculate the posterior probabilities for pre- to post-treatment changes in rCBF for each of two treatments. Figures 3(a) (olanzapine) and 3(b) (risperidone) show regions where the posterior probabilities for post-treatment changes are greater than or equal to 0.95 and reveal
differential working-memory-related neural responses for the two treatments. Treatment with olanzapine is associated with decreases in task-related rCBF in the right amygdala/hippocampus, insula, ventral striatum, and inferior frontal cortex, cerebellum, ventromedial and dorsomedial prefrontal cortex, posterior cingulate cortex, brainstem, bilateral lateral orbitofrontal cortex, and left dorsal striatum. Olanzapine-related increases in task-related activity are found in the left insula, posterior superior temporal sulcus (STS), and visual cortex. Treatment with risperidone is associated with decreased working memory-related rCBF in the left posterior hippocampus and inferior temporal gyrus, ventromedial prefrontal cortex, cerebellum, and posterior cingulate gyrus, and with increased activation of left insula, dorsal striatum, inferior frontal cortex, superior temporal gyrus, and thalamus. The posterior probability maps for the direct contrast of the two treatments effects (Figure 3(c)) indicate that the most significant differences in treatment effects are found in visual cortex, lingual gyrus, middle temporal gyrus, medial orbitofrontal cortex, cerebellum, ventral striatum, right insula, rostral anterior cingulate cortex, left thalamus, and right middle frontal gyrus.

**Individualized prediction maps**

We apply the proposed prediction algorithm to forecast the post-treatment rCBF response for each subject in the working memory study. An important advantage of our proposed prediction algorithm is that it provides individualized predictions of the neural responses to treatment based on the unique information in each subject’s baseline scans and relevant personal characteristics. Figure 4(a) illustrates the individualized predictions for the rCBF response to the low working memory condition after 12 weeks of treatment. For brevity, we only show here the predicted maps for 4 subjects in the working memory study (The individual prediction maps for all 16 subjects are presented in Figure S1 in Supplementary Material). Notable differences between subjects exist for the predicted post-treatment distributed patterns of working-memory-related brain activity. Subject 1 shows high task-related activation in the lingual gyrus and moderate activation of the superior temporal gyrus; subject 2
demonstrates high task-related activation in the ventromedial prefrontal cortex, and lingual and superior temporal gyri; subject 3 shows low task-related activation in the ventromedial prefrontal cortex and lingual gyrus and only moderate activation in the left superior temporal gyrus and subject 4 shows moderate activation in the left superior temporal gyrus and lingual gyrus. Our comparisons of the individualized predictions of treatment-related neural responses in Figure 4(a) to the observed data in Figure 4(b) reveal satisfactory agreement between the predicted and observed post-treatment brain activity in these 4 subjects. We observe similar correspondence in most of the 16 subjects (see Figure S1 and S2 in the supplementary materials).

The changes from the baseline rCBF to the predicted post-treatment rCBF response for two selected individuals in the working memory study are illustrated in Figure 5. In practice, these images would indicate to a clinician the potential effect of a treatment on an individual’s distributed neural processing associated with working memory. Subject 12 shows increased predicted task-related activity in the occipital cortex but decreased activity in the medial prefrontal cortex. Subject 14’s predicted responses include decreased activations in the occipital and temporal cortex, and right inferior frontal cortex but increased activation in the left temporal cortex. The individualized prediction maps, such as Figures 4 and 5, highlight individual differences in the effect of antipsychotic drug treatment on the neural responses of patients with schizophrenia to varying working memory load demands.

**Prediction accuracy for the post-treatment brain activity**

We use the $K$-fold cross-validation approach to evaluate the performance of our prediction algorithm in generating projected maps of treatment-related brain responses. Specifically, we split the 16 subjects into 4 equally-sized groups for the $K$-fold cross-validation. The prediction error is quantified by the two loss functions proposed in (13) and (14). Figure 6(a) displays the results reflected by the ratio of the square root of the PMSE to the average brain activity. The predictions prove accurate for most voxels, with the square root of the PMSE generally falling below 10% of the average localized rCBF and in
most voxels below 7%. Figure 6(a) shows that the highest prediction error is observed in the areas of the brainstem, hippocampus, inferior frontal cortex, posterior cingulate cortex, and posterior parietal cortex. The predictions for the medial prefrontal cortex and thalamus are fairly accurate where the errors are no higher than 5% of the average activity. Figure 7 depicts the probability that the 95% prediction interval includes the observed post-treatment rCBF. Across all voxels, 91% of the observed post-treatment rCBF measurements fall within the 95% prediction intervals. For most of the voxels, the coverage probability ranges from 86% to 100%, revealing a reasonable coverage level. The areas that show relatively low coverage probabilities include lateral orbitofrontal cortex, left temporal and occipital cortex. These areas also tend to have a high ratio of PMSE (Figure 6(a)).

Comparison with predictions based on the general linear model

Our novel prediction framework, based on a Bayesian hierarchical model, yields predicted patterns of post-treatment brain activity for a given patient. One could conceptually develop a similar prediction algorithm based on a GLM. The GLM models the brain activity for all subjects using common population parameters and assumes independence and homoscedasticity (sphericity) between scans at pre- and post-treatment periods, i.e.

\[
\begin{bmatrix}
Y_{i1}(v) \\
Y_{i2}(v)
\end{bmatrix} =
\begin{bmatrix}
X_{i1} & 0 \\
0 & X_{i2}
\end{bmatrix}
\begin{bmatrix}
\beta_1(v) \\
\beta_2(v)
\end{bmatrix} +
\begin{bmatrix}
Z_{i1} & 0 \\
0 & Z_{i2}
\end{bmatrix}
\begin{bmatrix}
\alpha_1(v) \\
\alpha_2(v)
\end{bmatrix} +
\begin{bmatrix}
\epsilon_{i1}(v) \\
\epsilon_{i2}(v)
\end{bmatrix},
\] (15)

where \( \epsilon_{i}(v) = [\epsilon_{i1}^T(v), \epsilon_{i2}^T(v)]^T \) follows \( N(0, \sigma^2_{i} I) \). Estimates from the GLM are obtained using OLS. The predicted post-treatment brain activity based on (15) is \( \hat{Y}_{i2}(v) = X_{i2} \hat{\beta}_2(v) + Z_{i2} \hat{\alpha}_2(v) \) which only reflects the population-level expectation and does not take into account of the information from the subject's pre-treatment scans. Figure 6(b) displays the square root of PMSE relative to the average brain activity, based on the GLM. A comparison between Figures 6(a) and 6(b) indicates that prediction errors
based on our Bayesian model are 25% lower than those from the GLM on average and the superiority of our prediction model is consistently observed across the brain.

The improved accuracy of the proposed prediction algorithm based on the Bayesian hierarchical model over one based on the GLM is mainly attributable to the fact that our model takes advantage of the subject-specific information from each individual’s baseline scans in the prediction of their treatment-related brain responses. In comparison, GLM-based predictions do not consider the baseline information due to the independence between scans from the two treatment periods. Consequently, our prediction algorithm borrows strength from the baseline scans to provide individualized predictions that reflect the unique traits of a subject, whereas predicted values from the GLM are more homogenous across subjects.

**Prediction accuracy for the fMRI study of inhibitory control**

We illustrate the use of our prediction algorithm for fMRI data from a study of inhibitory control among cocaine dependent men. The study investigates the effects of cocaine dependence on the neural representations of experimental tasks targeting inhibitory control, which is commonly impaired by drug addiction. We use the fMRI data collected from 9 cocaine-dependent subjects and 11 control subjects. For each subject, both pre- and post-treatment scans were obtained, with the follow-up scans occurring roughly 24 days following the initial image acquisitions. To illustrate the proposed method, we consider a Stop task, designed to evaluate the ability to inhibit a prepotent response. The task involves repeated presentations of visual Go stimuli (uppercase alphabetical letters), appearing for 500 msec with an interstimulus interval of 2.3 seconds. Subjects execute the speeded Go response by pressing a button as quickly as possible. For the Stop signal, an auditory tone, lasting 500 msec, occurs at random in 16% of the trials. The presentation of the Stop signal following the Go stimulus indicates that the subject should attempt to refrain from executing the Go response.
We seek to make predictions for post-treatment subject-specific effects corresponding to the Stop task. At each voxel, the first-stage of the Bayesian hierarchical model expresses the observed fMRI BOLD responses in terms of the subject-specific effect for the Stop task at the baseline and post-treatment periods. The second-stage models the subject-specific effects in terms of population parameters that are stratified on the cocaine-dependent status. We fit the second-stage model and covariance structure specified in (3) and (4).

We use $K$-fold cross-validation to evaluate the performance of our prediction algorithm in generating projected maps of post-treatment subject-specific BOLD effects for the Stop task. Specifically, we split the 20 subjects into 4 equally-sized groups for the $K$-fold cross-validation. The prediction error is quantified by calculating two loss functions proposed in (13) and (14) based on $\hat{B}_{i2}^{OLS}(v)$ and $\hat{B}_{i2}^{(-i)}(v)$. Here, $\hat{B}_{i2}^{OLS}(v)$ is the OLS estimator for the subject effect of the Stop task obtained from a standard GLM based on the post-treatment fMRI data and $\hat{B}_{i2}^{(-i)}(v)$ is the predicted post-treatment subject effect computed with the block that includes the $i$th subject removed from the training data sets. Figure 8(a) displays the results reflected by the square root of the PMSE standardized by the average subject effects. The predictions prove reasonably accurate, with the standardized PMSE falling below 15% for the majority of voxels. The highest prediction error is observed in the areas of medial frontal cortex, superior frontal cortex, occipital cortex and superior parietal cortex. Figure 8(b) depicts the probability that the 95% prediction interval includes the post-treatment subject effects estimated from the observed post-treatment fMRI data, i.e. $\hat{B}_{i2}^{OLS}(v)$. Across all voxels, 96% of the observed post-treatment subject effects fall within the 95% prediction intervals. For most of the voxels, the coverage probability is above 90%, revealing a satisfactory coverage level.
DISCUSSION

This research aimed to develop a prediction mechanism to forecast an individual patient’s brain response to an indicated treatment, focusing here on a core cognitive deficit in patients with schizophrenia. The results suggest that the neural response varies between two members of the class of newer atypical antipsychotics and between individuals within a treatment condition. An accurate prediction model would provide a quantifiable and objective framework to assist the currently symptom-driven, subjective treatment decision-making process for schizophrenia and these potential advantages extend to many other psychiatric disorders.

We construct our predictive method based on a Bayesian hierarchical model. Several authors have commented on the data analysis advantages of hierarchical models in neuroimaging studies [Worsley et al., 2002; Beckmann et al., 2003; Woolrich et al., 2004]. Our Bayesian hierarchical model offers a general framework for estimating the model parameters when prior knowledge is either present or absent. When there is no prior knowledge and an empirical Bayesian scheme is adopted, Bayesian estimation is akin to the classical frequentist approach [Friston et al., 2002]. In the presence of prior knowledge, such as the information regarding brain connectivity and activation, Bayesian methods have the advantage of being able to incorporate that information in both estimation and inference procedures. Our Bayesian framework, therefore, provides the ability to make connections to corresponding frequentist approaches, if needed, but also grants the ability to include prior information when it is available. Finally, the Bayesian framework offers greater potential to extend the current model to better inform the underlying complex neurophysiology.

The proposed Bayesian hierarchical model allows us to account for the heterogeneity in the task-related brain activity by modeling the within-subject repeated scans on the first level with subject-specific effects that are modeled subsequently on the second level in terms of population parameters. Moreover, our Bayesian hierarchical model provides a unified framework for considering a variety of
covariance structures that best describe the variances of, and the correlations between, each individual’s repeated scans.

We focus here on estimating the proposed hierarchical model using a parametric empirical Bayesian approach carried out with an EM algorithm. There is a close relationship between parametric empirical Bayes and classical frequentist inference procedures [Friston et al., 2002]. In particular, for a hierarchical linear model under Gaussian assumptions, the conditional means of the population level parameters are equivalent to the maximum likelihood estimates, and the EM estimates for the hyperparameters of the covariance components are the same as those based on restricted maximum likelihood (ReML). Therefore, one may use standard software such as PROC MIXED to make inference. However, this standard software can only be used for frequentist inference. Compared to these standard estimation tools, the advantages of the EM algorithm lie in it generalizability: it provides a uniform framework through which full Bayes, parametric empirical Bayes and Maximum likelihood estimation can be performed [Friston et al., 2002], and it offers the flexibility to estimate more complex nonspherical covariance structures that are needed to capture the heterogeneity and correlations among images taken across different conditions or times.

We have demonstrated that our prediction framework accurately forecasts post-treatment neural processing in a PET study of working memory among individuals with schizophrenia and an fMRI study of inhibitory control among cocaine dependent subjects. Since the performance of our prediction algorithm depends on the degree of association between pre- and post-treatment neural processing, which will vary for different tasks and psychiatric disorders, higher accuracy is obtainable. When applying our prediction framework to predict neural processing in either PET or fMRI studies involving other tasks or psychiatric disorders, it is important to follow the recommended steps (in the "Model validation: Evaluation of prediction error " section) to properly train and validate the model prior to use to ensure accurate predictions.
The estimation and validation of the proposed prediction algorithm involves heavy computation because the EM estimation procedure and the $K$-fold cross-validation involve time-consuming computations. Furthermore, the variance hyperparameters in our Bayesian hierarchical model are voxel-specific and hence are estimated by performing EM on a voxel-by-voxel basis, whereas earlier work [Friston et al., 2002] assumes global hyperparameters which could be estimated more efficiently with a single EM using sample covariance calculated across voxels [Friston et al., 2002]. The reason for us to spend considerable computational effort in estimating voxel-specific hyperparameters is that variance hyperparameters play a crucial role in predicting both the magnitude and variability of post-treatment brain activity (see equations (7) and (9)). Often, variance hyperparameters may change considerably across voxels due to different brain activity levels and functions. For the working memory study, the average coefficient of variation (CV) of the estimated variance hyperparameters is about 1.1, representing considerable amount of variations in the hyperparameters across voxels. Therefore, specifying voxel-specific hyperparameters is crucial to improve our prediction accuracy. It is important to point out that the heavy computations for estimation of the prediction algorithm only need to be performed during the algorithm development stage using the training data. After the prediction algorithm is established, applying the algorithm to calculate a new patient's predicted post-treatment brain maps is quite rapid. For example, for our working memory study which involves 16 subjects and PET images with $91 \times 109 \times 91$ voxels, the computation time for the training stage was 48.7 hours on a cluster of 4 processors with AMD Operton 850 CPUs (2.4Ghz) and 8GB RAM. The prediction for a new subject only took 2.18 minutes on 1 processor. For fMRI, one needs to first estimate the subject-specific pre-treatment effects using OLS estimation, which involves fast computations and can be implemented using existing neuroimaging analysis software such as SPM.

We are aware that the current approach to using pre-treatment neuroimaging data to predict clinical treatment response focuses on their relationship to symptom reduction, rather than the focus here on the response defined by task-related neural processing. We believe the predicted information for each
approach to be complementary to the ultimate goal of forecasting response and non-response. The use of clinical instruments to quantify symptom severity is not exact and patients with disorders such as schizophrenia generally have difficulty verbally articulating the nature or extent of their functional impairments. The focus on neural processing corollaries of psychiatric illness as a measure of treatment response that is more proximal to the molecular mechanisms of action of pharmacotherapies rather than behavioral symptoms is similar to the argument for similar intermediate measures in assessing genotype-phenotype correlations [Meyer-Lindenberg and Weinberger, 2006]. Finally, the ability to predict outcomes based on task-related neural activations has the advantage of focusing on neural processing corollaries of domains of function (e.g., cognitive, affective, social) targeted by illness and affected by treatments.

An important extension of our proposed framework is to develop a formalized algorithm for predicting the symptom response to treatment using baseline scans, predicted post-treatment activity, and patient characteristics. To predict treatment response to therapies, our prediction framework may combine with classification procedures such as discriminant analysis, support vector classifiers [Hastie et al., 2001] or Bayes classifiers [Ripley, 1996] to identity features in the pre- and post-treatment brain activity that best differentiate treatment responders from non-responders. Such a prediction model for treatment response may potentially enable physicians to make more informed treatment decisions either prior to initiating treatment or at an early phase of therapy. Our prediction framework for post-treatment brain activity is an important first step at aiding treatment decisions using information from in vivo functional neuroimaging scans. Furthermore, developing an extended algorithm to identify treatment responders could make use of the predicted post-treatment brain activity maps, produced by our prediction algorithm in this paper and shown to have high accuracy, so our current approach would play a critical role in this extended framework.
ACKNOWLEDGMENTS

Support for this research was provided by the National Institute of Mental Health, grant number K25-MH65473. Funding for the PET study of working memory in patients with schizophrenia was provided by a grant from Janssen Pharmaceutica. The authors thank Mr. Timothy Ely from the Department of Psychiatry and Behavioral Sciences at Emory University for his contributions toward data acquisition and preprocessing.

REFERENCES


Appendix A

This appendix describes how to make inferences for the proposed Bayesian hierarchical model (equation (1) and (2) in the paper) using the EM algorithm.

For simplicity of presentation, we suppress the notation for voxel $v$ in the following. For a given voxel, the two-stage model presented in (1) and (2) for all $N$ subjects can be collapsed into the following single-level representation:

$$Y = W\theta + \epsilon^{(i)} ,$$  \hspace{1cm} (A1)

where $Y = [Y_1^T, ..., Y_N^T]^T$ is the response variables from all $N$ subjects. The covariate matrix

$W = [X^{(i)}, X^{(i)}X^{(2)}, Z]$ where $X^{(j)} = diag(X_1^{(j)}, ..., X_N^{(j)})$ for $j = 1, 2$ and $Z = diag(Z_1, ..., Z_N)$. The parameter $\theta = [\epsilon^{(2)}^T, \beta^T, \alpha^T]^T$ with $\epsilon^{(2)} = [\epsilon_1^{(2)}^T, \epsilon_2^{(2)}]^T$, $\beta^T = [\beta_1^T, \beta_2^T]^T$ and $\alpha^T = [\alpha_1^T, \alpha_2^T]^T$, and $\epsilon^{(1)} = [\epsilon_1^{(1)}^T, \epsilon_2^{(1)}]^T$. The parameter $\theta$ contains the population-level parameters, and the error terms in the second-level of the model which are treated as parameters in Bayesian framework because they act as priors for the subject-specific parameters on the first-level of the hierarchical model [Friston et al., 2002]. Based on our model specification, $\theta$ is assumed to follow Gaussian distributions with mean and covariance of

$$\eta_\theta = E(\theta) = [0^T, \eta_\beta^T, \eta_\alpha^T]^T$$, and $$\Sigma_\theta = Cov(\theta) = \begin{bmatrix} \Sigma^{(2)} & 0 & 0 \\ 0 & \Sigma_\beta & 0 \\ 0 & 0 & \Sigma_\alpha \end{bmatrix} .$$

The likelihood and priors for (A1) under the Gaussian distribution assumptions are

$$p(Y \mid \theta) \propto \exp\{-\frac{1}{2}(Y - W\theta)^T \Sigma^{(1)-1} (Y - W\theta)\} ,$$ and
\[ p(\theta) \propto \exp\left\{ -\frac{1}{2} (\theta - \eta_\theta)^T \Sigma_\theta^{-1} (\theta - \eta_\theta) \right\}, \]

which lead to a Gaussian posterior density for \( \theta \),

\[ p(\theta \mid Y) \propto \exp\left\{ -\frac{1}{2} (\theta - \eta_{\theta Y})^T \Sigma_{\theta Y}^{-1} (\theta - \eta_{\theta Y}) \right\}, \]

where

\[
\Sigma_{\theta Y} = (W^T \Sigma^{(1)}^{-1} W + \Sigma_\theta^{-1})^{-1}, \quad \text{and} \quad 
\eta_{\theta Y} = \Sigma_{\theta Y} (W^T \Sigma^{(1)}^{-1} Y + \Sigma_\theta^{-1} \eta_\theta). \tag{A2}
\]

When we adopt the empirical Bayesian scheme with \( \Sigma_\beta = \infty \) and \( \Sigma_\alpha = \infty \) [Friton et al., 2002],

\( \Sigma_\theta^{-1} \eta_\theta = 0 \) in (A2) which means we do not need to specify the prior mean \( \eta_\beta \) and \( \eta_\alpha \) since they do not influence the posterior. Therefore, the hyperparameters only include the parameters in the covariance matrices \( \Sigma^{(1)} \) and \( \Sigma^{(2)} \), i.e. \( \lambda = (\lambda^{(1)}_k, \lambda^{(2)}_k)^T \).

To estimate the parameters and hyperparameters in (A1), we use the EM algorithm described below.

**E-step:**

At the E-step, we obtain the conditional distribution of the parameter \( \theta \) given the observed data and the current estimate of the hyperparameters \( \lambda^{(m)} \), i.e. \( p(\theta \mid \lambda^{(m)}, Y) \). Based on the model specification, the conditional distribution is Gaussian with mean and covariance of

\[
\Sigma^{(m)}_{\theta Y} = (W^T \hat{\Sigma}^{(1)}^{-1} W + \hat{\Sigma}_\theta^{-1})^{-1}, \quad \eta^{(m)}_{\theta Y} = \Sigma^{(m)}_{\theta Y} (W^T \hat{\Sigma}^{(1)}^{-1} Y + \hat{\Sigma}_\theta^{-1} \eta_\theta),
\]

where \( \hat{\Sigma}^{(1)} \) and \( \hat{\Sigma}_\theta \) are based on \( \lambda^{(m)} \).

**M-step:**

At the M-step, we use the estimated conditional distribution of \( \theta \) from the E-step to update the estimates for the hyperparameter \( \lambda \). Specifically, we find \( \lambda^{(m+1)} \) such that maximize the log-likelihood

\[ \ln p(\theta, Y \mid \lambda) \]

integrated over the estimated conditional distribution of \( \theta \) obtained from E-step, i.e.
\[
\lambda^{(m+1)} = \arg \max_{\lambda} \int_\Theta \{ \ln p(\theta, Y \mid \lambda) \} p(\theta \mid \lambda^{(m)}, Y) d\theta . 
\]  
(A3)

We denote the objective function in (A3) as \( F \). One can show that

\[
F = \int_\Theta \{ \ln p(\theta, Y \mid \lambda) \} p(\theta \mid \lambda^{(m)}, Y) d\theta = \\
- \frac{1}{2} \{ (\ln |\Sigma^{(1)}| + \ln |\Sigma_\theta|) + (Y - W \eta_{\theta Y}^{(m)})^T \Sigma^{(1)}^{-1} (Y - W \eta_{\theta Y}^{(m)}) + (\eta_\theta - \eta_{\theta Y}^{(m)})^T \Sigma_\theta^{-1} (\eta_\theta - \eta_{\theta Y}^{(m)}) \\
+ \text{tr}[\Sigma_{\theta Y}^{(m)} (W^T \Sigma^{(1)-1} W + \Sigma_\theta^{-1})]\} + c ,
\]

where \( c \) is a constant. To find the \( \lambda \) that maximizes \( F \), one can use algorithms such as Fisher scoring or Newton-Raphson. For example, using Fisher scoring algorithm,

\[
\lambda = \lambda + H^{-1} g ,
\]

where \( g = \frac{\partial F}{\partial \lambda} \) and \( H = \frac{\partial^2 F}{\partial \lambda^2} \).

We iterate the E-step and M-step until the parameters and the hyperparameters reach convergence. The EM algorithm then returns the posterior mean and covariance for the parameter, i.e. \( \hat{\eta}_{\theta Y} \) and \( \hat{\Sigma}_{\theta Y} \) and the ML estimates of the hyperparameters \( \hat{\lambda} \).
Figure 1: Schematic illustration of the development and application of the prediction algorithm for post-treatment brain activity maps. In the development stage, we first model the relationships between the post-treatment brain activity and factors including pre-treatment brain activity, treatment, and relevant subject characteristics. We then derive the prediction algorithm from the fitted statistical model. In the application step, clinical practitioners obtain scans for a new patient prior to making a treatment decision, input the patient’s data and characteristics into the prediction algorithm, and obtain the predicted post-treatment brain activity maps.
Figure 2

Flow chart for the Bayesian hierarchical model. Symbols represent different roles in the model with diamonds for hyperparameters, ovals for primary parameters, octagons for error terms, rectangles for design matrices, and a rounded rectangle for observations/measurements. Black symbols represent components related to effects of interest for prediction and grey symbols represent components related to effects of no particular interest. The first stage of the hierarchy models brain activity measurements in terms of subject-specific parameters through design matrices containing scan-specific features, and the second stage models the subject-specific effects in terms of population parameters through design matrices containing subject characteristics. The population parameters are modulated by hyperparameters based on prior information.
(a) pre- to post-treatment changes with olanzapine

(b) pre- to post-treatment changes with risperidone

c). comparing pre- to post-treatment changes between olanzapine and risperidone

Figure 3

Thresholded posterior probability maps (PPM) for second-level treatment effects on pre- to post-treatment changes in rCBF based on the Bayesian hierarchical model for the working memory study. The axial slices shown are relative to the anterior commissure. (a) and (b) depict regions that show increase (yellow) or decrease (blue) in rCBF from pre- to post-treatment with posterior probability greater than 0.95 with each of the two treatments. In plot(c), yellow indicates the pre- to post-treatment difference is greater with olanzapine than with risperidone with posterior probability greater than 0.95, i.e. $\Pr[(\text{post-pre})_{\text{olanzapine}} - (\text{post-pre})_{\text{risperidone}} > 0] > 0.95$; blue indicates the pre- to post-treatment difference is greater with risperidone than olanzapine with posterior probability greater than 0.95, i.e. $\Pr[(\text{post-pre})_{\text{risperidone}} - (\text{post-pre})_{\text{olanzapine}} > 0] > 0.95$. Olanzapine results in decreased rCBF responses in more brain locations than risperidone.
Individualized predicted and observed post-treatment rCBF measurements under the *low load* condition for 4 subjects in the working memory study. The axial slice shown is -6mm below the anterior commissure. (a) predicted maps. Notable differences exist between the patients’ predicted brain responses to treatment. (b) observed maps. There is satisfactory agreement between the predicted and observed post-treatment rCBF.
Predicted pre- to post-treatment alterations under the *low load* condition after treatment with olanzapine for subjects 12 and 14 in the working memory study. Subject 12 shows increased task-related activation in the occipital cortex but decreased activations in the medial prefrontal cortex. Subject 14 is predicted to have decreased activations in the occipital and temporal cortex, and inferior frontal cortex but increased activations in the left temporal cortex.
Maps depicting the square root of the prediction mean square error (PMSE) divided by the average brain activity at each voxel for prediction of the post-treatment brain activity for the working memory study. The axial slices range from -18mm to +18 mm from the anterior commissure. (a) Bayesian hierarchical model. The square root of the PMSE generally falls below 10% of the average localized rCBF. The highest prediction error is observed in the areas of the brainstem, hippocampus, inferior frontal cortex, posterior cingulate cortex, and posterior parietal cortex. The prediction in the areas of medial prefrontal cortex and thalamus is fairly accurate where the error is no higher than 5% of the average activity. (b) GLM. Prediction errors based on the proposed Bayesian hierarchical model are 25% lower than those of the GLM on average and the superiority of our prediction model is consistently observed across the brain.
Maps showing coverage probabilities of the 95% prediction intervals for predictions of post-treatment brain activity based on the Bayesian hierarchical model for the working memory study. The axial slices range from -18mm to +18 mm from the anterior comissure. Across all voxels, 91% of the observed post-treatment rCBF measurements fall within the 95% prediction intervals. The areas that show relatively low coverage probabilities include lateral orbitofrontal cortex, left temporal and occipital cortex. These areas also tend to have high ratio of PMSE in Figure 1(a).
Maps showing the prediction error for predicting the post-treatment subject effects for the fMRI study of cocaine dependence. The axial slices range from +12 mm to +48 mm from the anterior commissure. (a) the standardized square root of the prediction mean square error (PMSE). The standardized PMSE falls below 15% for the majority of voxels. The highest prediction error is observed in the areas of medial frontal cortex, superior frontal cortex, occipital cortex and superior parietal cortex. (b) coverage probabilities of the 95% prediction intervals. Across all voxels, 96% of the post-treatment subject effects estimated from the observed post-treatment fMRI data fall within the 95% prediction intervals.
Supplementary Material

Predicting the Brain Response to Treatment using a Bayesian Hierarchical Model with Application to a Study of Schizophrenia

Ying Guo\textsuperscript{1*}, F. DuBois Bowman\textsuperscript{1} and Clinton Kilts\textsuperscript{2}

\textsuperscript{1}Department of Biostatistics, The Rollins School of Public Health, Emory University, Atlanta, GA

\textsuperscript{2}Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA

Supplementary material for this article includes two figures. Figure S1 illustrates the individualized predictions for the rCBF response to the low working memory condition after 12 weeks of treatment for the 16 subjects in the working memory study. Figure S2 presents the observed post-treatment rCBF measurements under the low load condition for the 16 subjects in the working memory study. Comparisons between Figure S2 and S2 reveal satisfactory agreement between the predicted and observed post-treatment brain activity in the subjects.
Individualized predictions of post-treatment rCBF measurements under the low load condition for the 16 subjects in the working memory study. The axial slice shown is -6mm below the anterior commissure.
Figure S2
Observed post-treatment rCBF measurements under the low load condition for the 16 subjects in the working memory study. The axial slice shown is -6mm below the anterior commissure.