A Learning Algorithm for Predicting Mental Health Systems and Substance Use: *Technical Supplement*

Anthony T. Fojo, Catherine R. Lesko, Kelly Benke, Geetanjali Chander, Bryan Lau, Richard D. Moore, Peter P. Zandi, Scott L. Zeger

October 14, 2020

Contents

1 The Model

1.1 Model Terminology

- Assume we are interested in M classes of observations (symptoms scales or substance use indicators), indexed $m = 1, ..., M$. These classes of observations may either be binary, or continuous and bounded on $[L(m), U(m)]$
- Assume we have n participants, indexed $i = 1, ..., n$, such that the i^{th} participant has $N(i)$ observations, indexed $j = 1, ..., N(i)$.
- Let y_{ij} denote the j^{th} observation on participant i. The class of y_{ij} is denoted by m_{ij}, and the calendar time at which it was taken is denoted t_{ij}. Let $\boldsymbol{Y_i}$ denote the vector of all observations y_{ij} taken on participant i.
- For each observation j on participant i, let $\boldsymbol{X_{ij}}$ denote a row vector of K_{fixed} covariates for fixed effects at the time of that observation. Let $\boldsymbol{X_i}$ denote the $N(i) \times K_{fixed}$ matrix of covariates for participant i where the j^{th} row is $\boldsymbol{X_{ij}}$
- For each observation j on participant i, let $\boldsymbol{Z_{ij}}$ denote a row vector of K_{random} covariates for random effects at the time of that observation. Let $\boldsymbol{Z_i}$ denote the $N(i) \times K_{random}$ matrix of covariates for participant i where the j^{th} row is $\boldsymbol{Z_{ij}}$. We assume that the covariates for random effects are a subset of the covariates for fixed effects.
- We define $\boldsymbol{X_i^M}$ to be a transformation of $\boldsymbol{X_i}$ as follow: for $m = 1, ..., M$, let $\boldsymbol{X_i^m}$ be a $N(i) \times K_{fixed}$ matrix where the j^{th} row is

$$\begin{cases} \boldsymbol{X_{ij}} & if m_{ij} = m \\ \boldsymbol{0} & if m_{ij} \neq m \end{cases}$$

  Then $\boldsymbol{X_i^M}$ is the column union of each of the $\boldsymbol{X_i^m}$.
  This implies that $\boldsymbol{X_i^M}$ is a $N(i) \times \left(K_{fixed} \times M\right)$ matrix such that, for $j = 1, ..., N(i)$ and $k = 1, ..., K_{fixed}$:

$$X^M_{i,j,k+(m_{i,j}-1)\times K_{fixed}} = X_{i,j,k}$$

  While all other elements of $\boldsymbol{X_i^M}$ are zero.
  In constructing $\boldsymbol{X_i^M}$, we have defined a matrix that has each value in $\boldsymbol{X^i}$ exactly once, in a column corresponding to the class of the j^{th} observation and the covariate.
- Similarly, we define $\boldsymbol{Z_i^M}$ to be a transformation of $\boldsymbol{Z_i}$ - the column union of matrices $\boldsymbol{Z_i^m}$ whose j^{th} row is;

$$\begin{cases} \boldsymbol{Z_{ij}} & if m_{ij} = m \\ \boldsymbol{0} & if m_{ij} \neq m \end{cases}$$

  This implies that $\boldsymbol{Z_i^M}$ is a $N(i) \times \left(K_{random} \times M\right)$ matrix such that, for $j = 1, ..., N(i)$ and $k = 1, ..., K_{random}$:

$$Z^M_{i,j,k+(m_{i,j}-1)\times K_{random}} = Z_{i,j,k}$$

  While all other elements of $\boldsymbol{Z_i^M}$ are zero.

1.2 Definition of Latent Variables

1.2.1 Continuous Symptom Scales - a Tobit Model

We represent observations for continuous symptom scales using a Tobit model, which postulates that there is an unbounded latent variable, y'_{ij} such that:

- If the latent variable is less than the lower bound for the symptom scale - $y'_{ij} \leq L(m_{ij})$ - then we observe $y_{ij} = L(m_{ij})$.
- If the latent variable is greater than the upper bound for the symptom scale - $y'_{ij} \geq U(m_{ij})$ - then we observe $y_{ij} = U(m_{ij})$.
- If the latent variable falls within the bounds of the symptoms scale - $L(m_{ij}) < y'_{ij} < U(m_{ij})$ - then we observe $y_{ij} = y'_{ij}$.

1.2.2 Binary Indicators - a Probit Model

We represent observations for binary indicators using a Probit model, which postulates that there is an unbounded latent variable, y'_{ij} such that:

- If $y'_{ij} < 0$ then we observe the indicator to be negative: $y_{ij} = 0$.
- If $y'_{ij} \geq 0$ then we observe the indicator to be positive: $y_{ij} = 1$.

1.3 The Distribution for Latent Variables

The vector of all latent variables for participant i (for both continuous and binary observations) is denoted $\boldsymbol{Y_i'}$. We specify that $\boldsymbol{Y_i'}$ follows a Multivariate Normal distribution:

$$\boldsymbol{Y_i'} \sim Multivariate\text{-}Normal\Big(\boldsymbol{X_i^M\beta},\ \boldsymbol{Z_i\Gamma(Z_i^M)^\mathsf{T} + \Phi_i + \mathcal{E}_i}\Big) \tag{1}$$

Where $\boldsymbol{\beta}$ is a vector of fixed effects, and $\boldsymbol{X_i^M}$ denotes the matrix of covariates for those fixed effects for participant i at all times observations are taken.

The variance-covariance matrix has three components:

1. $\boldsymbol{Z_i^M\Gamma(Z_i^M)^\mathsf{T}}$ incorporates the random effects, where $\boldsymbol{\Gamma}$ is the variance-covariance matrix for the random effects, and $\boldsymbol{Z_i^M}$ are the covariates for random effects
2. Φ_i encodes a continuous autoregressive process that governs how observations for participant i are correlated over time.
3. $\boldsymbol{\mathcal{E}_i}$ is a matrix of measurement errors.

These elements are described in greater detail below.

1.3.1 Fixed Effects

We define $\beta_{k,m}$ to be the fixed effect for the k^{th} covariate ($k \in 1, ..., K_{fixed}$) for the m^{th} class of observations, where there are K_{fixed} covariates for fixed effects. We define $\boldsymbol{\beta}$ to be the vector of all $K_{fixed} \times M$ fixed effects such that element $k + (m-1) \times K_{fixed}$ of $\boldsymbol{\beta}$ is $\beta_{k,m}$ for $k = 1, ..., K_{random}$ and $m = 1, ..., M$

Each $\beta_{k,m}$ gets a weakly-informative Normal prior with mean 0:

- When the class m of the observation is continuous, the standard deviation of the prior is half the span from the lower to upper bound of the scale:

$$\beta_{k,m} \sim Normal\left(0, \left[\frac{U(m) - L(m)}{2}\right]^2\right)$$

- When the class m of the observation is binary, the standard deviation of the prior is 1:

$$\beta_{k,m} \sim Normal(0, 1)$$

1.3.2 Random Effects

Conceptually, we postulate that $\alpha_{i,k,m}$ denotes a random effect for participant i for the k^{th} covariate (where $k \in 1, ..., K_{random}$) and the m^{th} class of observations, where there are K_{random} covariates for random effects. In actuality, we marginalize over the α's so that instead of having to estimate each $\alpha_{i,k,m}$, we only have to estimate their covariance matrix.

Let $\boldsymbol{\alpha_i}$ be the vector of all $K_{random} \times M$ random effects for participant i such that element $\big[k + (m-1) \times K_{random}\big]$ of $\boldsymbol{\alpha_i}$ is $\alpha_{i,k,m}$ for $k = 1, ..., K_{random}$ and $m = 1, ..., M$.

We give $\boldsymbol{\alpha_i}$ a Multivariate-Normal prior, which allows random effects across covariates and observation classes to covary. We take the mean for each random effect to be zero, since all covariates for random effects are also used as covariates for fixed effects.

$$\boldsymbol{\alpha_i} \sim Multivariate\text{-}Normal(\boldsymbol{0}, \boldsymbol{\Gamma})$$

We fold the random effects into the distribution for $\boldsymbol{Y*_i}$ $\boldsymbol{\alpha_i}$s as follows:

$$\boldsymbol{Y_i'} = \boldsymbol{X_i^M\beta} + \boldsymbol{Z_i^M\alpha_i} + error$$

Using the simplified term *error* to encompass both the measurement error and autoregressive error components, where we assume $error \sim Multivariate\text{-}Normal(\boldsymbol{0}, \boldsymbol{\Sigma})$.

Since $\boldsymbol{\alpha_i} \sim Multivariate\text{-}Normal(\mathbf{0}, \boldsymbol{\Gamma})$,

$$\boldsymbol{Y_i'} \sim Multivariate\text{-}Normal\big(\boldsymbol{X_i^M}\boldsymbol{\beta}, \boldsymbol{Z_i^M}\boldsymbol{\Gamma}(\boldsymbol{Z_i^M})^{\mathsf{T}} + \boldsymbol{\Sigma}\big)$$

This allows us to avoid estimating each participants $\boldsymbol{\alpha_i}$ by marginalizing them out and estimating the variance-covariance matrix $\boldsymbol{\Gamma}$ only.

In formulating the prior for $\boldsymbol{\Gamma}$, we decompose it into a correlation matrix, $\boldsymbol{\Theta}$, and a vector of $K_{random} \times M$ standard deviations, $\boldsymbol{\lambda}$:

$$\boldsymbol{\Gamma} = \boldsymbol{\lambda}\boldsymbol{\Theta}\boldsymbol{\lambda}^{\mathsf{T}} \tag{2}$$

where element $\big[k + (m-1) \times K_{random}\big]$ of $\boldsymbol{\lambda}$ is $\lambda_{k,m}$, representing the standard deviation $\alpha_{i,k,m}$.

- For observations of class m where m is binary, we give $\lambda_{k,m}$ $(k = 1, ..., K_{random})$ a weakly-informative Half-Normal prior with scale parameter equal to 1:
$$\lambda_{k,m} \sim Half\text{-}Normal(1)$$
- For observations of class m where m is continuous, we give $\lambda_{k,m}$ weakly-informative Half-Normal prior with scale parameter equal to half the span from the lower to upper bound of the scale:
$$\lambda_{k,m} \sim Half\text{-}Normal\left(\Big[\frac{U(m) - L(m)}{2}\Big]^2\right)$$
- We give $\boldsymbol{\Theta}$ a Lewandowski-Kurowicka-Joe (LKJ) prior:
$$\boldsymbol{\Theta} \sim LKJ(0.25)$$

1.3.3 Continuous Autoregressive Error

We define $\boldsymbol{\mathcal{P}}$ to be a $M \times M$ variance-covariance matrix where $\mathcal{P}_{m_1,m_2}$ is the covariance between observations of two classes, m_1 and m_2, when they are taken at the same time. We also define a continous autoregressive coefficient, ρ such that the covariance between observations of two classes taken at different times, t_1 and t_2, is $\mathcal{P}_{m_1,m_2} \times \rho^{|t_2 - t_1|}$

From $\boldsymbol{\mathcal{P}}$ and ρ we construct $\boldsymbol{\Phi_i}$, a $N(i) \times N(i)$ variance-covariance matrix which represents the covariance of the $N(i)$ observations on participant i as a function of their separation in time. For $j1 = 1, ..., N(i)$ and $j2 = 1, ..., N(i)$:

$$\Phi_{i,j_1,j_2} = \mathcal{P}_{m_1,m_2} \times \rho^{|t_{i,j_1} - t_{i,j_2}|}$$

In formulating the prior for $\boldsymbol{\mathcal{P}}$, we decompose it into a correlation matrix, $\boldsymbol{\Psi}$, and a vector of M standard deviations, $\boldsymbol{\pi}$:

$$\boldsymbol{\mathcal{P}} = \boldsymbol{\pi}\boldsymbol{\Psi}\boldsymbol{\pi}^{\mathsf{T}} \tag{3}$$

- For observations of class m where m is binary, we give π_m a weakly-informative Half-Normal prior with scale parameter equal to 1:
$$\pi_m \sim Half\text{-}Normal(1)$$
- For observations of class m where m is continuous, we give π_m weakly-informative Half-Normal prior with scale parameter equal to half the span from the lower to upper bound of the scale:
$$\pi_m \sim Half\text{-}Normal\left(\Big[\frac{U(m) - L(m)}{2}\Big]^2\right)$$
- We give $\boldsymbol{\Psi}$ a Lewandowski-Kurowicka-Joe (LKJ) prior:
$$\boldsymbol{\Psi} \sim LKJ(0.25)$$

1.3.4 Measurement Error

We allow measurement errors for different classes of observations taken at the same time to be correlated by defining an $M \times M$ covariance matrix, $\boldsymbol{E}$ of measurement errors, such that E_{m_1,m_2} represents the covariance between contemporaneous measurement errors for scales m_1 and m_2.

We create the $N(i) \times N(i)$ participant-specific measurement error matrix, $\boldsymbol{\mathcal{E}_i}$ from $\boldsymbol{E}$ such that for $j_1 = 1, ..., N(i)$ and $j_2 = 1, ..., N(i)$:

$$\mathcal{E}_{i,j_1,j_2} = \begin{cases} E_{m_{ij_1},m_{ij_2}} & \text{if} = t_{i,j_1} = t_{i,j_2} \\ 0 & \text{if} = t_{i,j_1} \neq t_{i,j_2} \end{cases}$$

In formulating the prior for $\boldsymbol{E}$, we decompose it into an $M \times M$ correlation matrix, $\boldsymbol{\Sigma}$, and a vector of M standard deviations, $\boldsymbol{\nu}$:

$$\boldsymbol{E} = \boldsymbol{\nu}\boldsymbol{\Sigma}\boldsymbol{\nu}^{\mathsf{T}} \tag{4}$$

- For observations of class m where m is binary, we set $\nu_m = 1$. (Allowing the standard deviations to vary for all three components of error - random effects, autoregressive error, and measurement error - would introduce an identifiability problem, so we fix the measurement error at 1).
- For observations of class m where m is continuous, we give ν_m weakly-informative Half-Normal prior with scale parameter equal to half the span from the lower to upper bound of the scale:

$$\nu_m \sim Half\text{-}Normal\left(\left[\frac{U(m) - L(m)}{2}\right]^2\right)$$

- We give $\boldsymbol{\Sigma}$ a Lewandowski-Kurowicka-Joe (LKJ) prior:

$$\boldsymbol{\Sigma} \sim LKJ(0.25)$$

2 Making Predictions from a Fitted Model

Because fitting the model is computationally expensive, we developed a computationally efficient method to predict new symptom scores and binary indicators from a fitted MCMC chain.

At each of the R iterations of the chain, we use the sampled model parameters to formulate a multivariate normal distribution for each participant that links latent variables corresponding to prior observations to the latent variables corresponding to future observations. For each iteration, we simulate a number (B) of values of prior latent variables consistent with observed prior values; we condition on each of the B simulated sets of values to derive B predictive, multivariate normal distributions for the future.

We aggregate the $R \times B$ multivariate normal distributions for each set of simulated values at each iteration of the chain into a mixture of multivariate normal distributions, which constitutes our prediction distribution for a new participant.

2.1 Terminology for Predictions

- As before, we are dealing with M classes of observations (symptoms scales or substance use indicators), indexed $m = 1, ..., M$. These classes of observations may either be binary, or continuous and bounded on $[L(m), U(m)]$
- We are given n participants for whom we have $N(i)$ prior observations, indexed $j = 1, ..., N(i)$. Each observation, y_{ij} is of class m_{ij}, taken at calendar time t_{ij}. We define $\boldsymbol{Y_i}$ to be the vector of all $N(i)$ prior observations on participant i.
- For each participant, i, we are given matrices of covariates for K_{fixed} fixed effects, $\boldsymbol{X_i}$ and covariates for K_{random} random effects, $\boldsymbol{Z_i}$, which we transform into matrices $\boldsymbol{X_i^M}$ and $\boldsymbol{Z_i^M}$ as described in Section 1.1: "Model Terminology"
- For each participant i, we desire to make $N^*(i)$ predictions, indexed $j^* = 1, ..., N^*(i)$. Each observation is to be of class m_{ij^*} at time t_{ij^*}. The $N^*(i)$ desired predictions correspond to matrices of covariates for fixed and random effects $\boldsymbol{X_i^*}$ and $\boldsymbol{Z_i^*}$, which we transform into matrices $\boldsymbol{X_i^{*M}}$ and $\boldsymbol{Z_i^{*M}}$.
- We also presume that we are given an MCMC chain consisting of R iterations. For each iteration $r = 1, ..., R$, we have the following parameters, corresponding to the parameters defined in 1.3: "The Distribution for Latent Variables"
  - $\boldsymbol{\beta^{[r]}}$ - A vector of $K_{fixed} \times M$ fixed effects
  - $\boldsymbol{\Gamma^{(r)}}$ - A $(K_{random} \times M) \times (K_{random} \times M)$ variance-covariance matrix for the M random effects for each covariate
  - $\boldsymbol{\mathcal{P}^{(r)}}$ and $\rho^{(r)}$ - An $M \times M$ variance-covarience matrix and scalar continuous autoregressive coefficients, respectively, that govern the covariance of observations as a function of separation in time
  - $\boldsymbol{E^{(r)}}$ - An $M \times M$ variance-covariance matrix for measurement errors

2.2 Making Predictions

2.2.1 Recap of Latent Variables

Recall that our model handles observations of class m using a Tobit model when class m is continuous and a Probit model when class m is binary.

This postulates that for each y_{ij} there is a corresponding unbounded latent variable, y'_{ij}. When m_{ij} denotes a continuous class of observations:

$$y_{ij} = \begin{cases} L(m_{ij}) & if y'_{ij} \leq L(m_{ij}) \\ U(m_{ij}) & if y'_{ij} \geq U(m_{ij}) \\ y'_{ij} & if L(m_{ij}) < y'_{ij} < U(m_{ij}) \end{cases}$$

When m_{ij} denotes a binary class of observations:

$$y_{ij} = \begin{cases} 0 & if y'_{ij} < 0 \\ 1 & if y'_{ij} \geq 0 \end{cases}$$

2.2.2 The Multivariate Distribution for Latent Variables

For each participant i and iteration r of the MCMC, we construct:

$$\boldsymbol{Y_i'} \sim Multivariate\text{-}Normal\Big(\boldsymbol{X_i^M \beta^{(r)}},\ \boldsymbol{Z_i^M \Gamma^{(r)} (Z_i^M)^\intercal + \Phi_i^{(r)} + \mathcal{E}_i^{(r)}}\Big) \tag{5}$$

Where $\boldsymbol{\Phi_i^{(r)}}$ is constructed from $\boldsymbol{\mathcal{P}^{(r)}}$ and $\rho^{(r)}$ as described in Section 1.3.3: "Continuous Autoregressive Error" and $\boldsymbol{\mathcal{E}_i^{(r)}}$ is constructed from $\boldsymbol{E^{(r)}}$, as described in Section 1.3.4: "Measurement Error"

2.2.3 Simulating Unobserved Latent Variables from Observations

We observe some latent variables directly - the observations of continuous scales where the observed values fall between the lower and upper bounds for the scale. Based on this, for each participant i, we partition $\boldsymbol{Y_i}$ into two subsets:

- $\boldsymbol{Y_i^o}$ is the vector of all observations y_{ij} where the class m_{ij} of the observation is continuous and $L(m_{ij}) < y_{ij} < U(m_{ij})$. This represents the subset of observations on participant i where we directly observe the latent variable as $y'_{ij} = y_{ij}$
- $\boldsymbol{Y_i^u}$ is the vector of all observations y_{ij} where the class m_{ij} of the observation is binary or $y_{ij} = L(m_{ij})$ or $y_{ij} = U(m_{ij})$. This represents the subset of observations on participant i where the latent variable is unobserved $\left(y'_{ij} \neq y_{ij}\right)$

$\boldsymbol{Y_i^{o\prime}}$ and $\boldsymbol{Y_i^{u\prime}}$ represent the corresponding partition of $\boldsymbol{Y_i'}$. We note that by definition, $\boldsymbol{Y_i^{o\prime}} = \boldsymbol{Y_i^0}$. Based on equation (5), and using the properties of the Multivariate Normal distribution, we can derive the conditional distribution of $\boldsymbol{Y_i^{u\prime}}$ given $\boldsymbol{Y_i^{o\prime}}$. Further, we know that the values of each element of $\boldsymbol{Y_i^{u\prime}}$ are bounded as follows:

- When the class of the observation m_{ij} is continuous and $y_{ij} = L(m_{ij})$, $y'_{ij} \leq L(m_{ij})$
- When the class of the observation m_{ij} is continuous and $y_{ij} = U(m_{ij})$, $y'_{ij} \geq U(m_{ij})$
- When the class of the observation m_{ij} is binary, and the observed indicator is negative, $y'_{ij} < 0$
- When the class of the observation m_{ij} is binary, and the observed indicator is positive, $y'_{ij} \geq 0$

Using the distribution $\boldsymbol{Y_i^{u\prime}}|\boldsymbol{Y_i^{o\prime}}$, we simulate values for each element of $\boldsymbol{Y_i^{u\prime}}$ from a truncated multivariate normal distribution according to these bounds.

We repeat this procedure some number (B) of times; each time, this procedure gives us a complete set of the prior latent variables $\boldsymbol{Y_i'}$.

2.3 Deriving the Predictive Distribution from Prior Latent Variables

If we define :

- $\boldsymbol{Y_i^*}$ to represent the set of observations for which we wish to make predictions for participant i
- $\boldsymbol{Y_i^{*\prime}}$ to represent the corresponding latent variables
- $\hat{\boldsymbol{Y}}_{\boldsymbol{i}}' = \begin{bmatrix} \boldsymbol{Y_i^{*\prime}} \\ \boldsymbol{Y_i'} \end{bmatrix}$ to be the joint vector of $\boldsymbol{Y_i^{*\prime}}$ and $\boldsymbol{Y_i'}$
- $\hat{\boldsymbol{X}}_{\boldsymbol{i}}^{\boldsymbol{M}} = \begin{bmatrix} \boldsymbol{X_i^{*M}} \\ \boldsymbol{X_i^M} \end{bmatrix}$ - to be the row union of covariates for fixed effects for future and prior observations for participant i
- $\hat{\boldsymbol{Z}}_{\boldsymbol{i}}^{\boldsymbol{M}} = \begin{bmatrix} \boldsymbol{Z_i^{*M}} \\ \boldsymbol{Z_i^M} \end{bmatrix}$ - to be the row union of covariates for random effects for future and prior observations for participant i

We can define a joint multivariate distribution for the latent variables corresponding to both future and prior observations:

$$\hat{\boldsymbol{Y}}_{\boldsymbol{i}}' \sim \textit{Multivariate-Normal}\Big(\hat{\boldsymbol{X}}_{\boldsymbol{i}}^{\boldsymbol{M}}\boldsymbol{\beta^{(r)}},\ \hat{\boldsymbol{Z}}_{\boldsymbol{i}}^{\boldsymbol{M}}\boldsymbol{\Gamma^{(r)}}(\hat{\boldsymbol{Z}}_{\boldsymbol{i}}^{\boldsymbol{M}})^{\mathsf{T}} + \hat{\boldsymbol{\Phi}}_{\boldsymbol{i}}^{\boldsymbol{(r)}} + \hat{\boldsymbol{\mathcal{E}}}_{\boldsymbol{i}}^{\boldsymbol{(r)}}\Big) \tag{6}$$

Where $\hat{\boldsymbol{\Phi}}_{\boldsymbol{i}}^{\boldsymbol{(r)}}$ is constructed from $\boldsymbol{\mathcal{P}^{(r)}}$ and $\rho^{(r)}$ for the set of both future and prior observation times, analagous to the construction of $\boldsymbol{\Phi}_{\boldsymbol{i}}^{\boldsymbol{(r)}}$ as described in Section 1.3.3: "Continuous Autoregressive Error" and $\hat{\boldsymbol{\mathcal{E}}}_{\boldsymbol{i}}^{\boldsymbol{(r)}}$ is constructed from $\boldsymbol{E^{(r)}}$, analagous to the construction of $\boldsymbol{\mathcal{E}}_{\boldsymbol{i}}^{\boldsymbol{(r)}}$ described in Section 1.3.4: "Measurement Error"

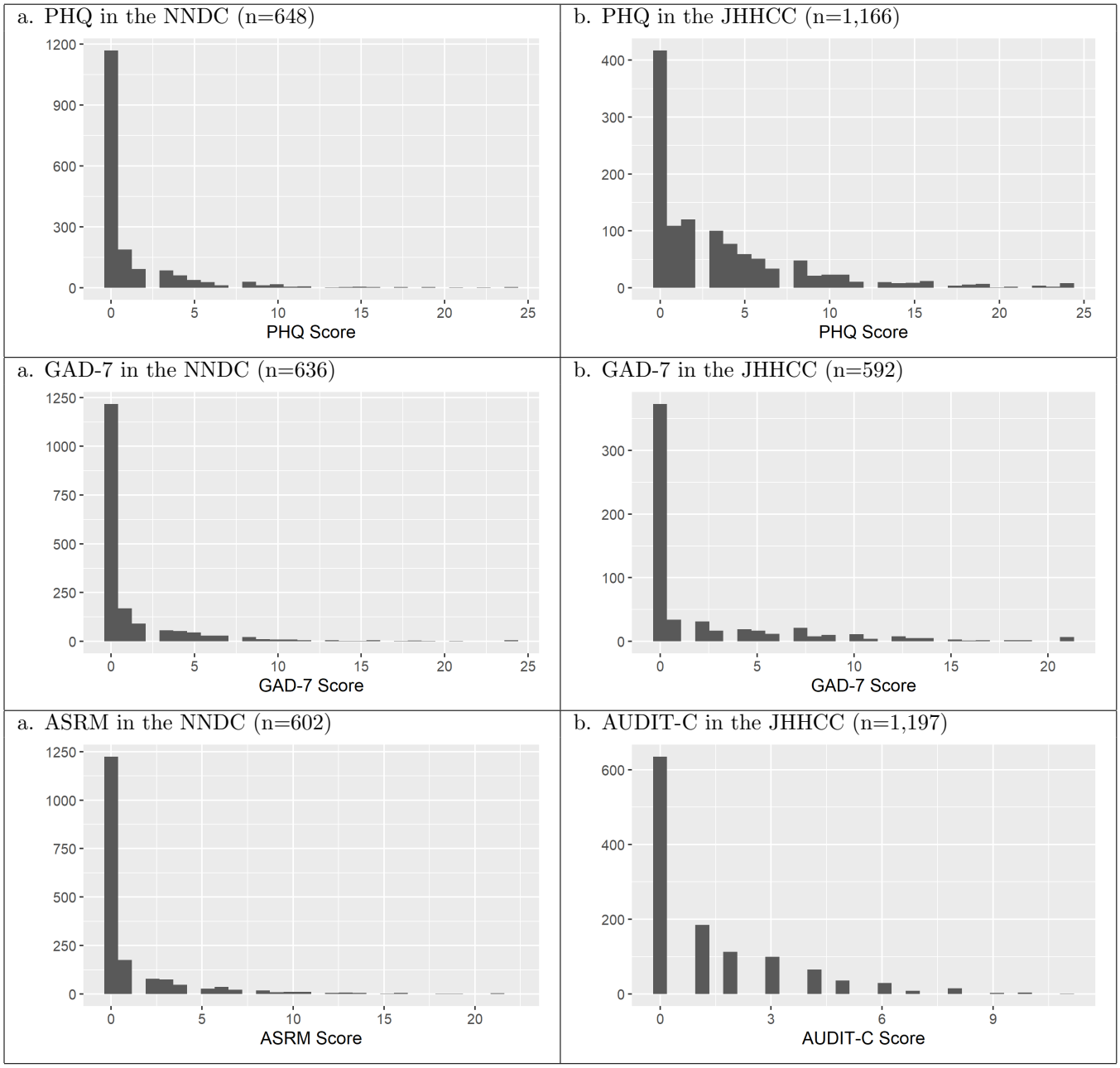
For each iteration of the MCMC chain and each of the B sets of simulated latent variables for that iteration, we can use equation (6) to derive the conditional distribution of $\boldsymbol{Y}_i^{*\prime}$ given $\boldsymbol{Y_i'}$. This yields a Multivariate Normal distribution which represents the predictive distribution for the latent variables corresponding to future observations for the given iteration of the chain and simulated prior latent variables.

We aggregate the $R \times B$ multivariate normal distributions as a mixture, which we use as our predictive distribution. To make predictions for a particular outcome at a particular time, we consider the mixture of marginal normal distributions for the one latent variable corresponding to that outcome:

- To predict a continuous outcome, we derive the mean (or median) and credible interval for the corresponding latent variable from the mixture, and truncate if any of those values exceeds the upper limit or falls below the lower limit of the scale
- To predict the probability of a binary outcome happening, we calculate the probability that the corresponding latent variable is ≥ 0 across the mixture
- To predict the probability that a continuous outcome is above some threshold, we calculate the probability that the corresponding latent variable is $\geq$ that threshold across the mixture

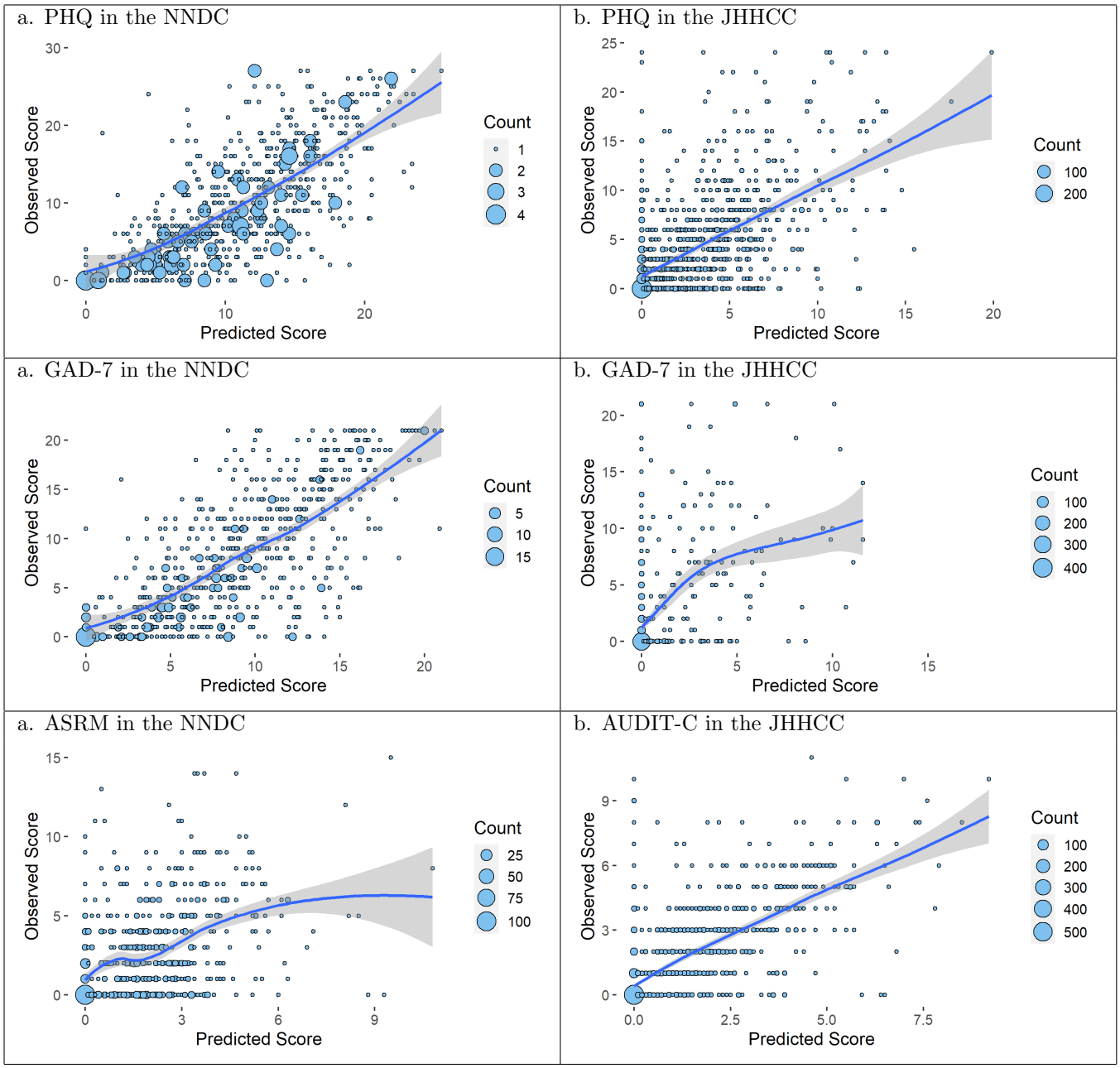
3 Additional Figures for Evaluating Predictive Performance

Figure S1: Distribution of Outcomes Used for Validation*



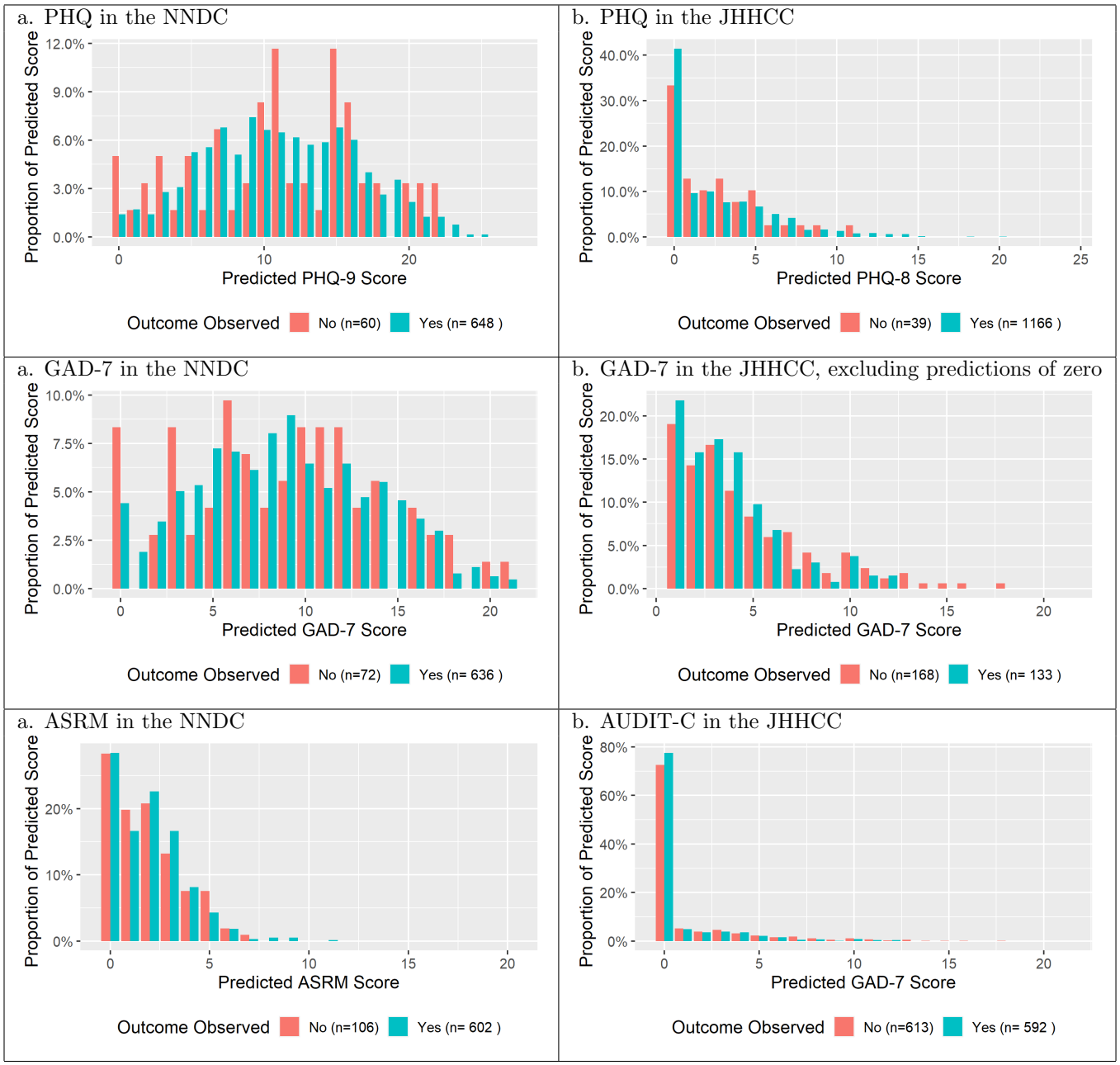
*Includes only the outcomes used in validating the prediction algorithm.

Figure S2: Calibration Plots for Continuous Outcomes



The size of each point is proportional to the number of observations which have corresponding predicted (rounded to the nearest tenth) and observed values. A gamma smoother is applied to each plot.

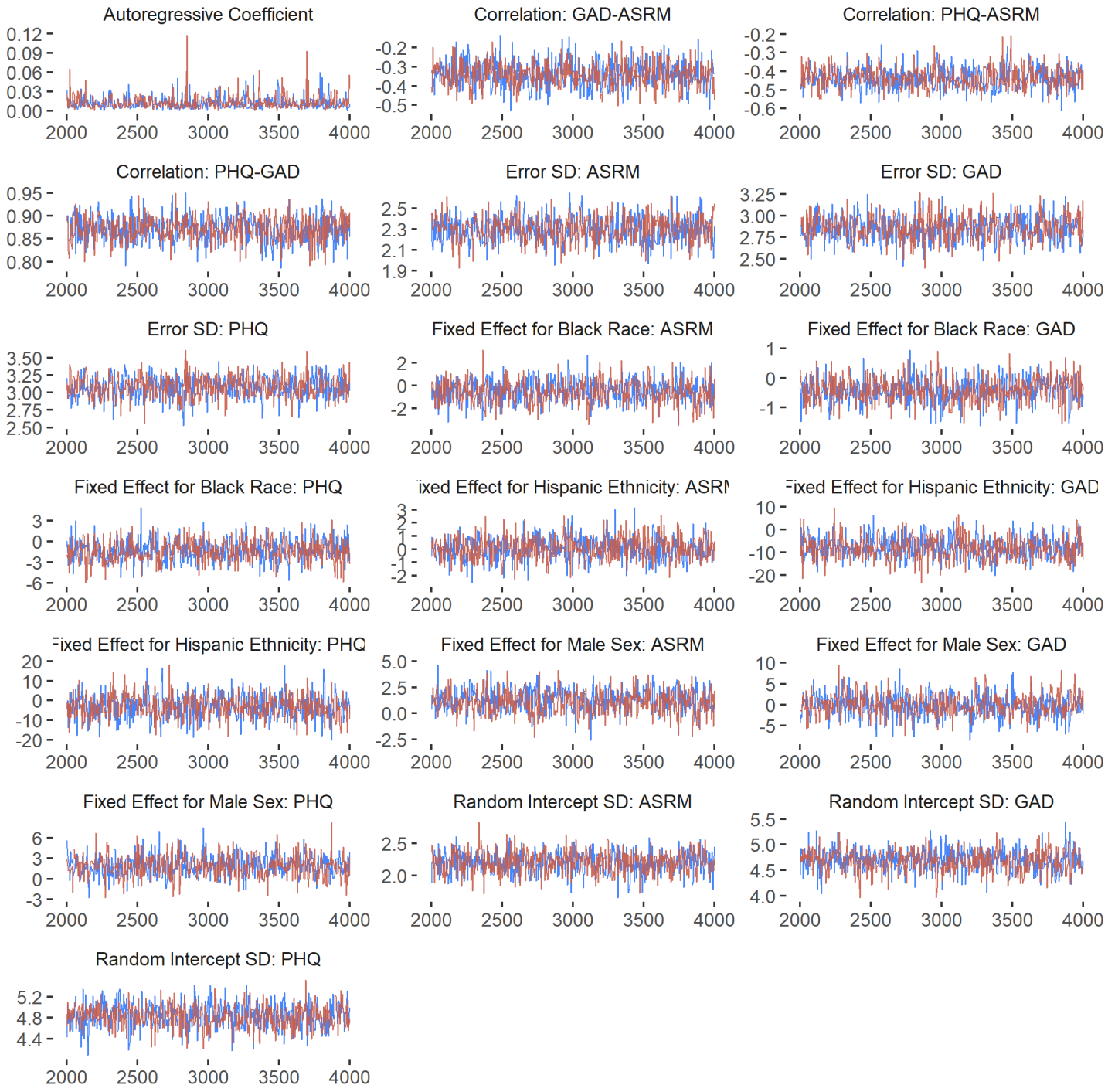
Figure S3: Distribution of Predicted Outcomes When the True Outcome was Observed vs. Missing*



*Only for outcomes where ¿20 observations were missing. This excludes the AUDIT-C score and heroin and cocaine use indicators from the JHHCC.

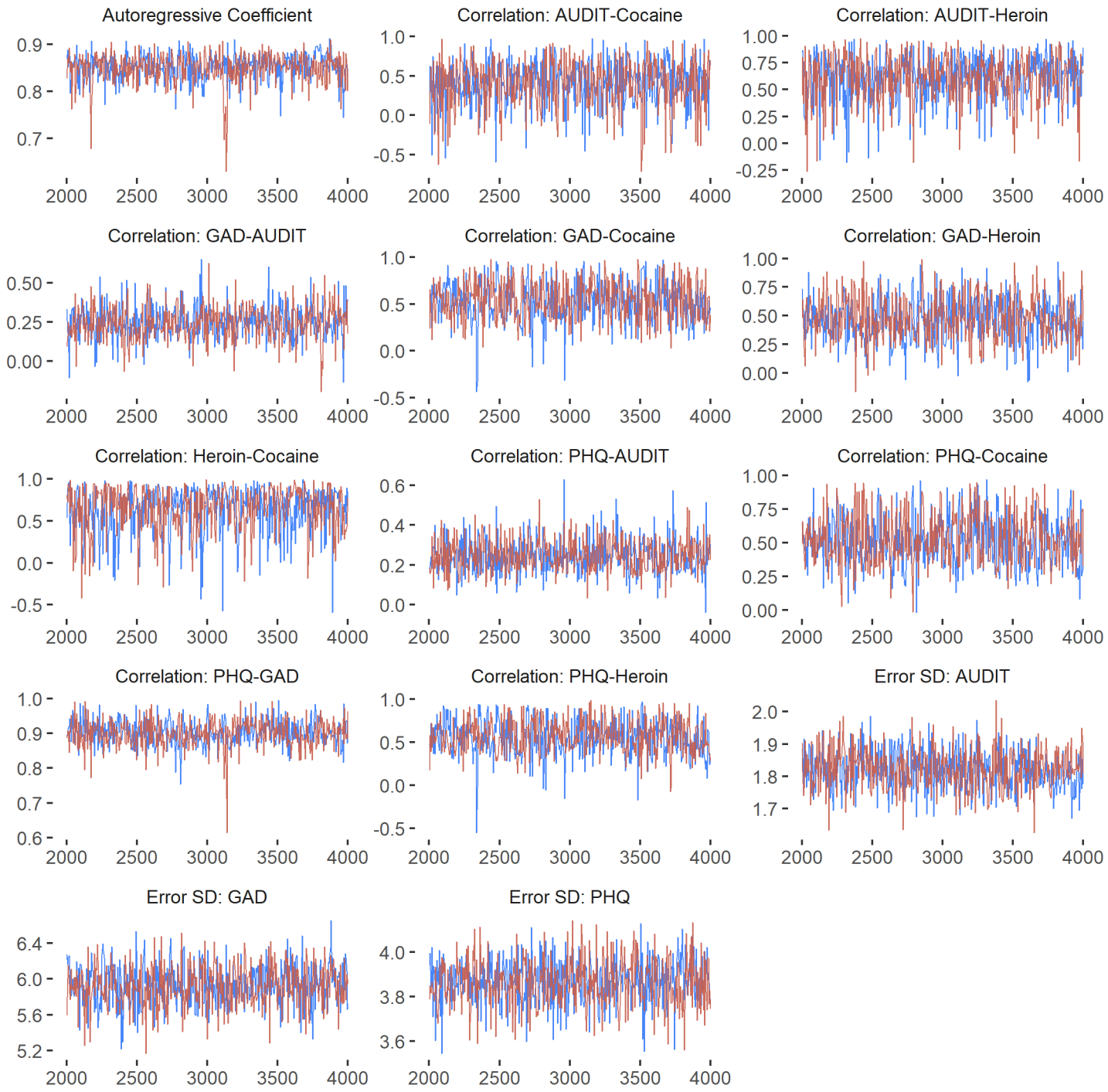
4 Additional Figures regarding MCMC Fitting

Figure S4: Trace Plots for Key Variables in Fitted Model for NNDC*



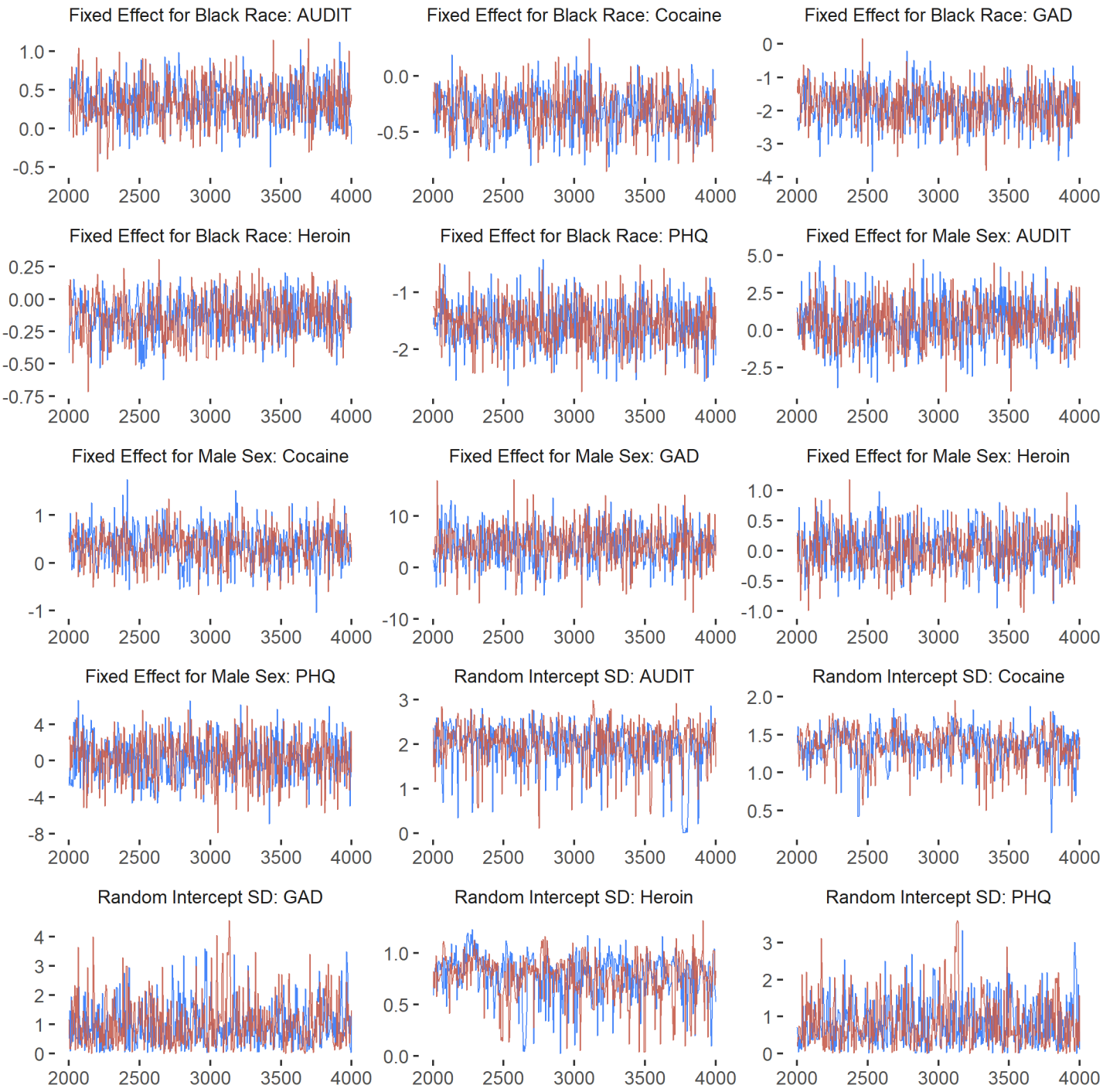
*Shown for the fit to the 1st of 5 validation sets

Figure S5: Trace Plots for Key Variables in Fitted Model for CFAR (Part 1)*



*Shown for the fit to the 1st of 5 validation sets

Figure S6: Trace Plots for Key Variables in Fitted Model for CFAR (Part 2)*



*Shown for the fit to the 1st of 5 validation sets

Figure S7: Posterior Mean Values for Key Variables Across Five Validations for NNDC (Part 1)

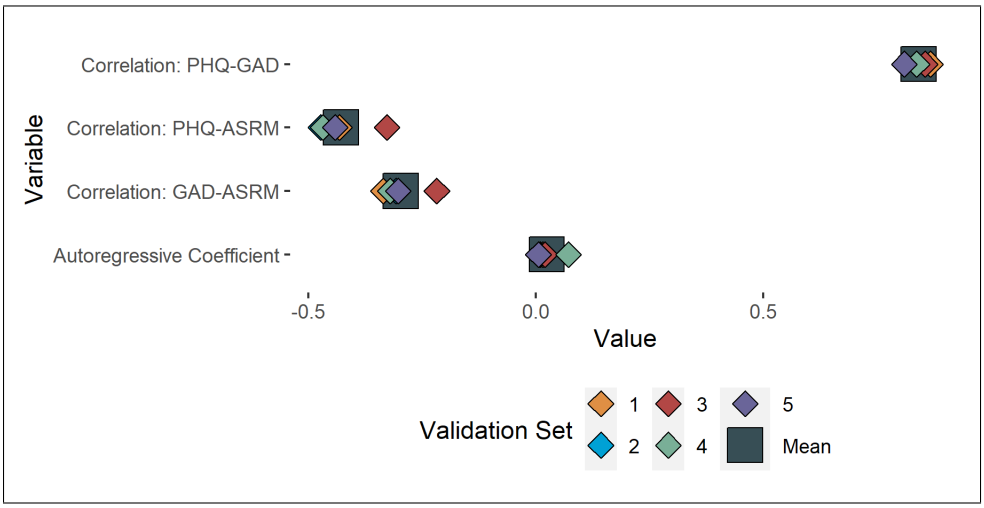


Figure S8: Posterior Mean Values for Key Variables Across Five Validations for NNDC (Part 2)

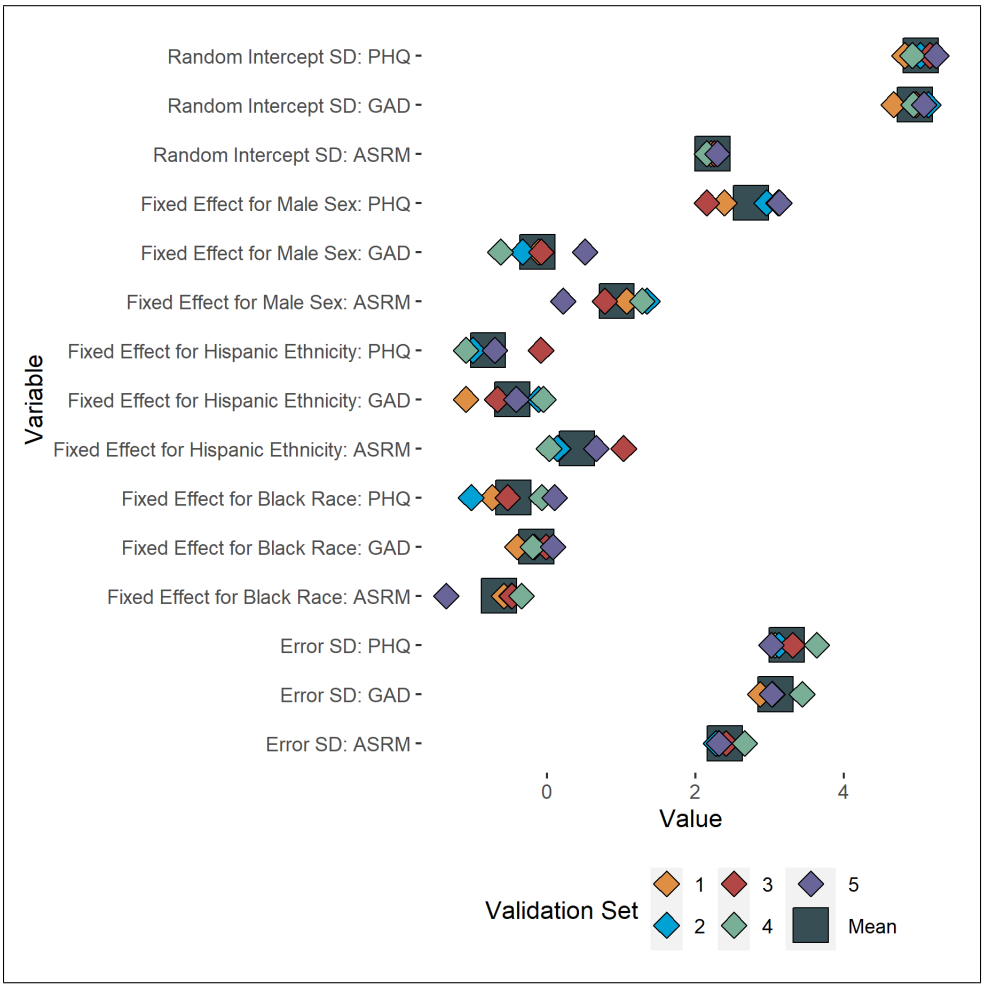


Figure S9: Posterior Mean Values for Key Variables Across Five Validations for CFAR (Part 1)

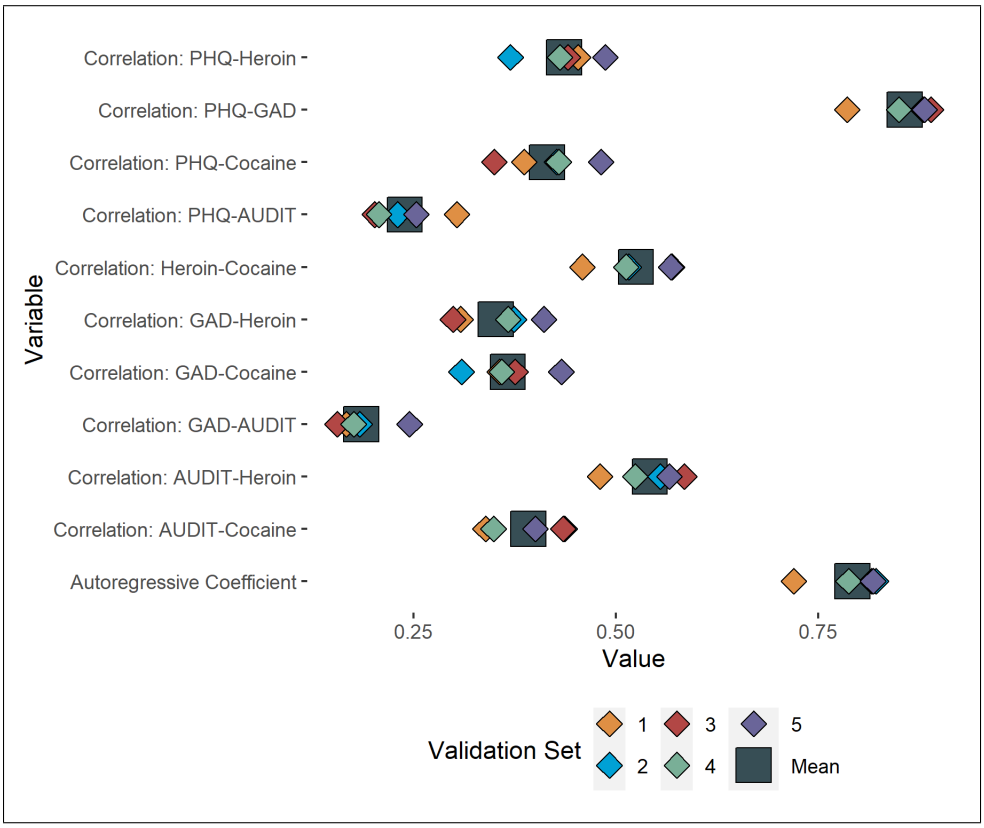


Figure S10: Posterior Mean Values for Key Variables Across Five Validations for CFAR (Part 2)