

## RESEARCH ARTICLE

# A fully Bayesian mixture model approach for identifying noncompliance in a regulatory tobacco clinical trial

Alexander M. Kaizer<sup>1</sup>  | Joseph S. Koopmeiners<sup>2</sup> 

<sup>1</sup>Department of Biostatistics and Informatics, University of Colorado-Anschutz Medical Campus, Aurora, Colorado

<sup>2</sup>Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota

## Correspondence

Alex M. Kaizer, Department of Biostatistics and Informatics, University of Colorado-Anschutz Medical Campus, Aurora, CO 80045.  
Email: alex.kaizer@ucdenver.edu

## Funding information

National Cancer Institute, Grant/Award Number: P30-CA077598; National Institute on Drug Abuse, Grant/Award Numbers: R01-DA046320, R03-DA041870, U54-DA031659

Identifying noncompliance in a randomized trial is challenging, but could be improved by leveraging biomarker data to identify participants that did not comply with their assigned treatment. For randomized trials of very low nicotine content (VLNC) cigarettes, the biomarker of total nicotine equivalents (TNE) could be used to identify noncompliance. Compliant participants should have lower levels of TNEs than participants that did not comply and smoked normal nicotine content cigarettes, resulting in a mixture of compliant and noncompliant participants at each dose level. Thresholds of TNE could then be identified from the compliant groups at each dose level and used to determine which study participants were compliant. Furthermore, proposed biological relationships of TNE with nicotine dose could be incorporated into improve the efficiency of estimation, but may introduce bias if misspecified. To account for multiple modeling assumptions across dose levels, we explore model averaging via reversible jump markov chain monte carlo (MCMC) within each dose level to take advantage of improvements in efficiency when the proposed relationship is true and to downweight the biological model when it is misspecified. In simulation studies, we demonstrate that model averaging in the presence of a correct biological relationship results in a decrease in the mean square error (MSE) of up to 85%, but downweights the model in dose levels where the relationship is not appropriate. We apply our approach to data from a randomized trial of VLNC cigarettes to estimate TNE thresholds and probability of compliance curves as a function of TNEs for each nicotine dose used in the trial.

## KEYWORDS

Bayesian model averaging, clinical trial, noncompliance, regulatory science, reversible jump MCMC

## 1 | INTRODUCTION

The Center for the Evaluation of Nicotine in Cigarettes, Project 1, Study 1 (CENIC-p1s1), was a 6-week randomized multicenter trial designed to evaluate the effect of nicotine reduction on tobacco use behavior.<sup>1</sup> Eight hundred and thirty nine current smokers underwent randomization, with 780 completing the 6-week study after being equally randomized to one of the seven groups, including a usual brand normal nicotine condition and one of the six experimental cigarettes with nicotine content ranging from 15.8 mg per gram of tobacco (normal nicotine controls) to 0.4 mg per gram of tobacco

(very low nicotine content [VLNC] cigarettes). At the end of six weeks, participants randomly assigned to the lowest nicotine condition had significantly reduced tobacco use, dependence, and nicotine exposure relative to the normal nicotine conditions.

In 2009, Congress passed the Family Smoking Prevention and Tobacco Control Act (FSPTCA), which gave the Food and Drug Administration (FDA) the authority to regulate the content, marketing, and sale of tobacco products. In particular, the FSPTCA gives the FDA the authority to reduce the nicotine content of cigarettes to non-addictive levels (but not zero) if it can be shown that this will improve public health. CENIC-p1s1 provides key scientific evidence in support of a new nicotine standard for cigarettes, but understanding the impact of mandated nicotine reduction on tobacco use behavior is difficult due to the presence of substantial noncompliance to randomized treatment assignment. That is, in a future regulatory environment where nicotine reduction is mandated by regulation, only VLNC cigarettes would be legally available and compliance would be much higher than what was observed in CENIC-p1s1.

In CENIC-p1s1, noncompliance is defined as smoking any commercially available, nonstudy cigarettes in place of, or in addition to, the study cigarettes provided by the trial. Identifying noncompliance in CENIC-p1s1 is difficult because the most direct measure of noncompliance, self-reported smoking of nonstudy cigarettes, is known to be unreliable.<sup>2</sup> An alternative strategy is to use biomarkers of nicotine exposure (ie, cotinine or total nicotine equivalents [TNEs]) to identify noncompliance among self-reported compliers.<sup>3</sup> For example, a recent study characterized the distribution of biomarkers of nicotine exposure in participants who were sequestered in a hotel for five days to ensure that they only smoked cigarettes with 0.4 mg of nicotine per gram of tobacco.<sup>4</sup> In this study, the 95th percentile for TNEs in fully compliant subjects was estimated at 6.41 nmol/ml, which has been used as a threshold for identifying noncompliance in secondary analyses of CENIC-p1s1.<sup>5,6</sup> Alternatively, Boatman et al<sup>7</sup> used a mixture model approach, which incorporated the data from Denlinger et al<sup>4</sup> to estimate the probability that a subject randomized to the VLNC group was compliant conditional on their biomarker values, which facilitated estimation of the causal effect of nicotine reduction on cigarettes smoked per day.

While the results of Denlinger et al<sup>4</sup> provide thresholds for identifying noncompliance in the VLNC group, they can not be used to identify noncompliance at the intermediate dose levels (ie, nicotine dose levels between 0.4 and 15.8 mg/g). The mixture model approach proposed by Boatman et al<sup>7</sup> could be used to identify noncompliance at the intermediate dose levels but fitting mixture models is known to be challenging due to label switching and overlap between the mixture components.<sup>8-10</sup> Alternatively, we could leverage our understanding of the biological relationship between the nicotine content of cigarettes and biomarkers of nicotine exposure to better estimate the mixture components.<sup>3</sup> We expect the nicotine content of the cigarettes to have an additive effect on biomarkers of nicotine exposure and we could specify a model that incorporates this relationship, in which case the mean of the compliant mixture component is estimated relative to the normal nicotine conditions.<sup>3</sup> However, it is possible that smoking lower nicotine content cigarettes may lead to compensatory smoking, such as puffing more intensely, which may result in observed data not following the hypothesized relationship. This could improve estimation of the mixture components and result in narrower credible intervals due to increased precision but is also susceptible to model misspecification.

The presence of two distinct approaches to modeling (assuming no relationship versus assuming an additive effect) represents a challenge given the uncertainty as to which model is most appropriate for the data. The Bayesian framework is naturally able to account for the uncertainty around selecting the “best” approach by considering multiple models simultaneously and averaging the results, thus avoiding the need to select one model. In fact, we need not limit ourselves to the two, extreme modeling approaches listed above, and could instead also consider models where some dose levels follow the biological relationship and others do not. This use of the Bayesian model averaging allows us to “mix and match” which approach is most appropriate for each dose level so that any gains in efficiency from assuming a biological relationship can be used for the appropriate dose levels while avoiding the introduction of bias when it is inappropriate, either due to misspecification of the true relationship or due to compensatory behaviors of participants.

The remainder of this article proceeds as follows. In Section 2, we define a finite Gaussian mixture model with two components (compliant and noncompliant) for the distribution of biomarkers of nicotine exposure and discuss how our model specification can be adapted to account for the hypothesized relationship between the nicotine content of cigarettes and biomarkers of nicotine exposure. An reversible jump markov chain monte carlo (RJMCMC) algorithm for averaging over various models is described in Section 2.2. Simulation results evaluating our modeling approach when the hypothesized relationship does and does not hold are presented in Section 3 and we apply our proposed method to CENIC-p1s1 in Section 4. Finally, we conclude with a brief discussion in Section 5.

## 2 | PROBABILITY MODELS AND ESTIMATION

We first introduce the notation that will be used throughout the remainder of this manuscript. Let  $j = 1, \dots, J$  index the different randomized groups, and  $i = 1, \dots, n_j$  be an index for each observation within group  $j$ . Furthermore, let  $\mathbf{y}$  represent the observed outcome such that  $y_{ij}$  represents the value of the biomarker in participant  $i$  belonging to randomization group  $j$ . For CENIC-p1s1, the biomarker of interest is TNE on the log-scale, transformed to achieve a more Gaussian distribution.

Within any randomized treatment assignment for group  $j$ , there are individuals who are compliant and noncompliant, resulting in a two-component mixture distribution for biomarkers used to estimate compliance which we represent as  $k = 1, 2$  such that  $k = 1$  represents compliance and  $k = 2$  represents noncompliance. Our goal is to identify biomarker thresholds for identifying noncompliance to randomized treatment assignment. For example, as the nicotine content of the study cigarettes increase in CENIC-p1s1, so should the biomarker threshold for identifying non-compliance, but the exact relationship is unknown. When no dose-response relationship is assumed across dose levels, we will label this the “IND model.”

To avoid label switching (ie, the peaks of the posterior distribution for the two components switch such that the non-compliant component actually estimates the compliant component and vice versa) and because we make the assumption that the biomarker mean for noncompliant subjects will be higher, we propose a likelihood parameterization that addresses these issues. For group  $j$ , let  $\mu_j$  be the mean of the biomarker in the compliant component ( $k = 1$ );  $\mu_j + \theta_j$ , where  $\theta_j \geq 0$ , be the mean of the noncompliant component ( $k = 2$ );  $\tau_j$  be the common precision of the Gaussian distributions for compliant and noncompliant subjects; and  $p_j$  be the mixture weight, which represents the proportion of non-compliant subjects in group  $j$ . The specification on the components' means ensures that the noncompliant mean will be greater than the compliant component mean. Assuming a common precision for compliant and noncompliant components that can differ by group on the log-scale results in a common coefficient of variation on the original scale for each group whenever the standard deviation is small compared to the mean, which has been shown to be a reasonable assumption for these data.<sup>11</sup> In addition, to more easily derive the conditional posteriors, we introduce a latent indicator variable  $\mathbf{z}$  through data augmentation.<sup>12</sup> Define  $z_{ijk} = 1$  when  $y_{ij}$  belongs to component  $k$ , and 0 otherwise. These changes induce a complete data likelihood for group  $j$  consisting of both our known  $\mathbf{y}$  and unknown  $\mathbf{z}$  of

$$p(\mathbf{y}, \mathbf{z} | \mu_j, \theta_j, \tau_j, p_j) = \prod_{i=1}^{n_j} [(1 - p_j) \times \mathcal{N}(y_i | \mu_j, \tau_j)]^{z_{ij1}} [p_j \times \mathcal{N}(y_i | \mu_j + \theta_j, \tau_j)]^{z_{ij2}}. \quad (1)$$

Extending Equation (1) to multiple  $j$ , let  $\Theta = (\mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\tau})$ , where the bold notation represents vectors, including the corresponding parameter for each of the  $j$  groups. The complete data likelihood for all subjects can be written as:

$$p(\mathbf{y}, \mathbf{z} | \Theta) = \prod_{j=1}^J \left\{ \prod_{i=1}^{n_j} [(1 - p_j) \mathcal{N}(y_i | \mu_j, \tau_j)]^{z_{ij1}} [p_j \mathcal{N}(y_i | \mu_j + \theta_j, \tau_j)]^{z_{ij2}} \right\}, \quad (2)$$

where it can be noted that noncompliance is not possible in group  $J$  since all participants receive normal nicotine content cigarettes, which implies that  $p_J = 0$  and the likelihood for this group does not correspond to a mixture distribution.

The complete data likelihood can be modified to incorporate the hypothesized biological relationship between the nicotine content of the cigarette and the biomarkers, which we refer to as the “REL model” for the remainder of this article. First, within CENIC-p1s1, assume that group  $J$  represents the combined normal nicotine conditions (ie, the usual brand and 15.8-mg/g groups). Participants in these groups received cigarettes with nicotine content that was comparable to commercial cigarettes and, as a result, smoking nonstudy cigarettes was not considered noncompliance. We then assume that a relationship exists between  $\mu_j$  and  $\mu_J$  that can be defined by a function  $h$ :  $\mu_j^* = h_j(\mu_J)$ . In this case, we no longer estimate each  $\mu_j, j \neq J$ , but only  $\mu_J$ . Estimates for each  $\mu_j$  are based on the proposed relationship,  $h_j(\mu_J)$ . Note that data are now utilized across all levels to estimate  $\mu_J$ , but estimates of  $\theta_j$  are derived within each level to account for differences in the mean for the noncompliers across dose levels due to differences in the prevalence of partial compliance or smoking behavior across dose levels.

Specifically in the context of CENIC-p1s1 for the log-normally distributed TNE biomarker, we will consider the following proposed relationship, noting that participants were current smokers not actively attempting to quit smoking, with minimal environmental exposure to nicotine or use of other tobacco products, we expect that cigarettes will be their

primary source of nicotine exposure.<sup>2</sup> Letting  $w_j$  represent the nicotine content for dose level  $j$ , we propose the following function,  $h_j$ , that maps the ratio of the nicotine contents to the mean of the biomarker distribution on the log scale:

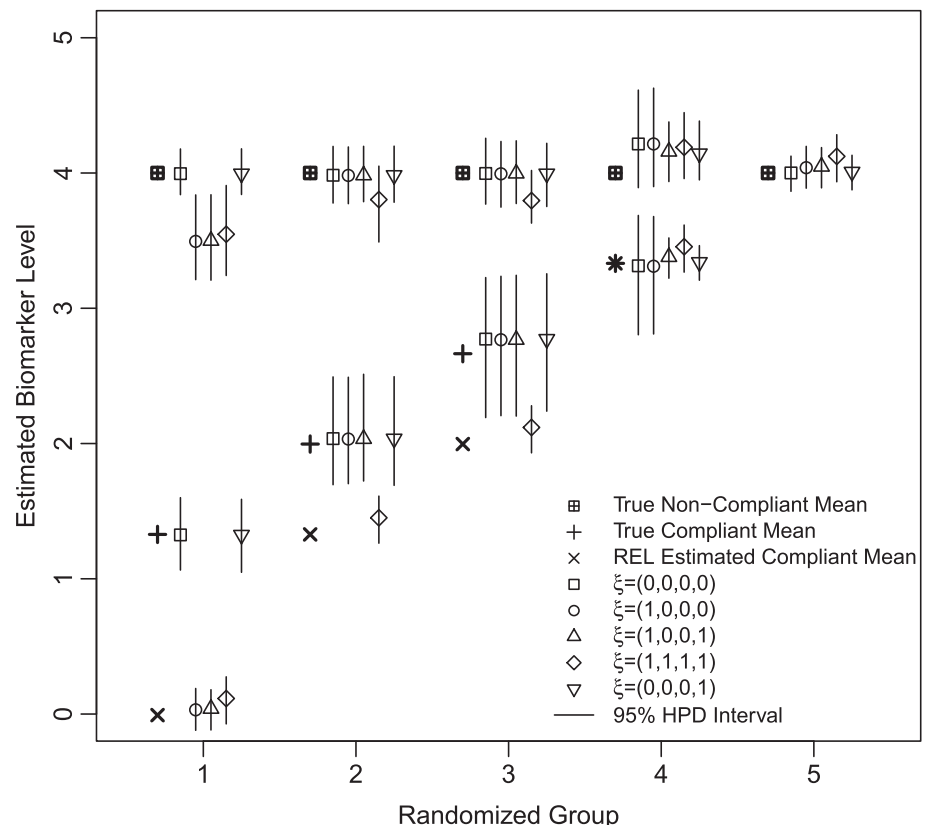
$$\mu_j^* = h_j(\mu_J) = \log\left(\frac{w_j}{w_J}\right) + \mu_J. \quad (3)$$

Biomarkers of nicotine exposure are thought to be linearly related to the nicotine content of the cigarettes and the proposed relationship imposes a linear relationship between the nicotine content of the cigarettes and the median on the original scale.<sup>13</sup> Given this relationship, the complete data likelihood for the REL model is

$$p(\mathbf{y}, \mathbf{z} | \Theta) = \prod_{j=1}^J \left\{ \prod_{i=1}^{n_j} [(1 - p_j) \mathcal{N}(y_i | \mu_j^*, \tau_j)]^{z_{ij1}} [p_j \mathcal{N}(y_i | \mu_j^* + \theta_j, \tau_j)]^{z_{ij2}} \right\}. \quad (4)$$

When considering all  $j$  levels, it may be the case that some levels are better estimated with the IND approach and others the REL approach. In this case, we can also average over intermediate models where some dose levels are estimated assuming the IND model, while others follow the REL model. For example, within CENIC-p1s1, there are a total of 16 potential models,  $m = 1, \dots, 16$ , after combining two groups with 15.8-mg/g nicotine (commercial versus study cigarettes) and two groups with 0.4-mg/g nicotine (normal versus high tar study cigarettes) to reduce the overall number of groups from 7 to 5, based upon no observed differences in the original CENIC-p1s1 analyses.<sup>1</sup> These 16 models consider all possible combinations of modeling frameworks assumed for the four conditions with reduced nicotine levels where the state of the model can be denoted by a vector,  $\xi_m$ , where 0 indicates the IND model and 1 indicates the REL model. For example,  $\xi_m = (1, 0, 0, 0)$  indicates that group 1 is estimated from the REL model, while groups 2 to 4 are estimated from the IND model. To illustrate the tradeoffs, Figure 1 presents the estimates from the IND and REL models for 500 simulated data sets based on correctly and incorrectly specified relationships where it can be noted that REL helps to reduce the 95% highest posterior density (HPD) interval, but can introduce bias when incorrectly specified. In Section 2.2, we will propose an RJMCMC algorithm that can traverse the model space of 16 models with different parameter dimensionalities to average over all possible combinations.

**FIGURE 1** Illustration of REL versus IND model for estimation of biomarker levels when the proposed biomarker levels for the compliant subgroup estimated from the REL model are misspecified for randomized groups 1, 2, and 3, but correctly specified for group 4. The proposed compliant subgroup REL biomarker level is represented by “x” and the true biomarker level is represented by “+” for the compliant subgroups. The true biomarker level for the noncompliant subgroup is represented by “⊞.” Different points correspond to the estimated biomarker level under different scenarios of  $\xi$  and its 95% highest posterior density (HPD) interval



## 2.1 | Prior distributions and posterior inference

In this section, we discuss prior distributions and posterior inference for the IND and REL approaches. In general, the posterior will not be available in closed form and must be approximated using markov chain monte carlo (MCMC).

Let  $\mathcal{HN}(0, \tau)$  denote a half-normal distribution such that if  $A \sim \mathcal{N}(0, \tau)$ , then  $B = |A|$  follows a half-normal distribution. For treatment group  $j$ , we assume the following prior distributions for various model parameters:

$$\begin{aligned}\mu_j &\sim \mathcal{N}(0, s_{\mu j}), \\ \theta_j &\sim \mathcal{HN}(0, s_{\theta j}), \\ \tau_j &\sim \Gamma(a_j, b_j), \\ p_j &\sim \text{Beta}(\alpha_j, \beta_j).\end{aligned}\quad (5)$$

Letting  $n_{jk}$  represent the number of observations in component  $k$  of group  $j$  and  $n_j = \sum_{k=1}^2 n_{jk}$ , the joint posterior for the IND model is proportional to the following:

$$\begin{aligned}p(\Theta|\mathbf{y}, \mathbf{z}) &\propto \prod_{j=1}^J \left\{ (1-p_j)^{n_{j1}+\beta_j-1} p_j^{n_{j2}+\alpha_j-1} \times \tau_j^{\frac{n_j}{2}+\alpha_j-1} \exp(-b_j \tau_j) \times \exp\left(-\frac{s_{\mu j}}{2} \mu_j^2\right) \right. \\ &\quad \left. \times \exp\left(-\frac{s_{\theta j}}{2} \theta_j^2\right) \times \exp\left(-\frac{\tau_j}{2} \left[ \sum_{i=1}^{n_j} z_{ij1} (y_i - \mu_j)^2 + z_{ij2} (y_i - [\mu_j + \theta_j])^2 \right] \right) \right\}.\end{aligned}\quad (6)$$

From (6), the conditional posterior for each parameter in the IND model within group  $j$  is:

$$p(\mu_j | \theta_j, \tau_j, \mathbf{y}, \mathbf{z}) \propto \mathcal{N}\left(\frac{\tau_j(n_{j1}\bar{y}_{j1} + n_{j2}(\bar{y}_{j2} - \theta_j))}{n_j \tau_j + s_{\mu j}}, n_j \tau_j + s_{\mu j}\right), \quad (7)$$

$$p(\theta_j | \mu_j, \tau_j, \mathbf{y}, \mathbf{z}) \propto \mathcal{N}\left(\frac{n_{j2} \tau_j (\bar{y}_{j2} - \mu_j)}{n_{j2} \tau_j + s_{\theta j}}, n_{j2} \tau_j + s_{\theta j}\right), \quad (8)$$

$$\begin{aligned}p(\tau_j | \mu_j, \theta_j, \mathbf{y}, \mathbf{z}) &\propto \Gamma\left(\frac{n_j}{2} + a_j, \frac{1}{2} \left[ (n_{j1} - 1)s_{j1}^2 + (n_{j2} - 1)s_{j2}^2 \right. \right. \\ &\quad \left. \left. + n_{j1}(\bar{y}_{j1} - \mu_j)^2 + n_{j2}(\bar{y}_{j2} - (\mu_j + \theta_j))^2 + 2b_j \right] \right),\end{aligned}\quad (9)$$

$$p(p_j | \mathbf{y}, \mathbf{z}) \propto \text{Beta}(n_{j2} + \alpha_j, n_{j1} + \beta_j), \quad (10)$$

$$p(z_{ij2} = 1 | y_{ij}, \Theta) \propto \text{Ber}\left(\frac{p_j \mathcal{N}(y_{ij} | \mu_j + \theta_j, \tau_j)}{(1-p_j) \mathcal{N}(y_{ij} | \mu_j, \tau_j) + p_j \mathcal{N}(y_{ij} | \mu_j + \theta_j, \tau_j)}\right), \quad (11)$$

where  $\bar{y}_{jk}$  is the mean of  $y_{ijk}$  for group  $j$  and component  $k$ .

The previous derivation can be easily extended to the REL model, where  $h_j(\mu_j)$  is specified as in Equation (3). The conditional posteriors for the REL model, with some additional notation, where  $\mu_j^* = h_j(\mu_j)$ , are:

$$d_j = \log\left(\frac{w_j}{w_J}\right), \quad (12)$$

$$g_j = \tau_j(n_j d_j + n_{j2} \theta_j - n_{j1} \bar{y}_{j1} - n_{j2} \bar{y}_{j2}), \quad (13)$$

$$p(\mu_J | \{\Theta \setminus \mu_J\}, \mathbf{y}, \mathbf{z}) \propto \mathcal{N}\left(\frac{n_J \tau_J \bar{y}_J - \sum_{j=1}^{J-1} g_j}{\sum_{j=1}^J (n_j \tau_j) + s_{\mu J}}, \sum_{j=1}^J (n_j \tau_j) + s_{\mu J}\right). \quad (14)$$



Note that the previously derived conditional posteriors for  $\theta$ ,  $\tau$ ,  $\mathbf{p}$ , and  $\mathbf{z}$  are unchanged from the IND model, except that estimates for  $\mu_j$  are derived from the proposed relationship,  $h_j(\mu_J)$ .

## 2.2 | Model averaging with RJMCMC

We now discuss how we will average over the IND, REL, and intermediate models using RJMCMC.<sup>14</sup> RJMCMC accounts for the varying dimensionality of the parameter space during model averaging. For example, the IND model estimates  $\mu_j$  from the group  $j$  observed data only, but in the REL model,  $\mu_j$  is derived from the proposed relationship with  $\mu_J$ , representing a change in the dimensionality of the parameter space due to the assumed relationship between  $\mu_j$  and  $\mu_J$ .

As noted above, we will consider intermediate models, in addition to the IND and REL models, to account for variability in the appropriateness of the hypothesized relationship between nicotine content and biomarkers of nicotine exposure across dose levels. That is, the hypothesized relationship may be appropriate for some dose levels but not others. For example, if  $\xi_m = (1, 0, 0, 1)$ ,  $\mu_J$  is updated using *only* groups  $j = 1, 4$  and the normal nicotine condition  $j = 5$ , while groups  $j = 2$  and  $3$  are not included because only data from their own group are used to derive parameter estimates.

While described in greater detail below, the general steps of the RJMCMC algorithm are:

1. Update the parameters conditional on the current model using the Gibbs sampler discussed in Section 2.1.
2. Update the model,  $m$ , conditional on the current parameter values using the RJMCMC algorithm that includes the following steps:
  - 2a. Randomly select a group  $j'$  from  $j = 1, \dots, J - 1$  with equal probability as a candidate to change states from REL to IND or vice versa.
  - 2b. Accept this proposed move with some probability.

We introduce additional notation for RJMCMC. Let  $q(\cdot)$  be an (arbitrary) proposal density for parameters that must be generated as the result of proposed moves that increase the dimensionality of the parameter space,  $f(\mathbf{y}|\Theta, m)$  represents the likelihood of our data given  $\Theta$  in model  $m$ , and  $p(m)$  represent the prior probability of model  $m$ . Let  $I_j = 1$  be an indicator that group  $j$  is estimated using the IND model, with  $I_j = 0$  indicating that group  $j$  is estimated using the REL model and define  $p(I_j = 1)$  to be the prior probability that  $I_j = 1$  with  $p(I_j = 0) = 1 - p(I_j = 1)$ , the prior probability that  $I_j = 0$ . We define  $p(m)$  as the product of these priors:  $\prod_{j=1}^{J-1} [p(I_j = 1)]^{I_j} [1 - p(I_j = 1)]^{1-I_j}$ . If  $p(I_j = 1) = 0.5$ , all models receive equal prior weight, whereas  $p(I_j = 1) > 0.5$  favors IND for group  $j$  and  $p(I_j = 1) < 0.5$  favors REL for group  $j$ .

In step 2a, the intermediate approaches can be conceptualized as a set of nested models of the REL approach where all groups are estimated assuming the relationship, so the proposal scheme suggested by Green and Hastie<sup>15</sup> for scenarios with nested models is implemented. In the scheme for nested models, at each iteration, we propose a move from  $(\Theta, m)$  to  $(\Theta', m')$  by randomly selecting  $j' \in 1, \dots, J - 1$  with equal probability as the candidate dose level to flip from IND to REL or vice versa. The change in the state of  $j'$  will increase or decrease the dimension of the parameter space by 1. For example, if  $\xi_m = (1, 0, 0, 0)$  and  $j' = 3$ , then  $\xi_{m'} = (1, 0, 1, 0)$  with group 3 switching from IND to REL and the parameter dimensionality decreases by 1. We note that, while we are considering an RJMCMC algorithm that only considers state-changes one-dose-level-at-a-time, an alternative proposal framework would be to consider jumping to any of the 15 other permutations of the model states for  $m'$  rather than flipping one group at each iteration.

In greater detail, if the proposed move changes group  $j'$  from the REL model to the IND model (increasing dimensionality), we define the bijective functions (ie, one-to-one and onto functions that map to only a single value in the range of possible values) for our proposed values to be:

$$\begin{aligned} \mathbf{p}' &= \mathbf{p}, \\ \tau' &= \tau, \\ \theta' &= \theta, \\ \mu'_{j \neq j'} &= \mu_{j \neq j'}, \\ \mu'_{j'} &= u, \end{aligned} \tag{15}$$

where  $u \sim q(\cdot)$  such that  $q(\cdot)$  is some (arbitrary) proposal distribution for  $\mu'_{j'}$ . We define  $q(\cdot)$  as the corresponding conditional posterior distribution from Equation (7).

Conversely, the reverse move for group  $j'$  from IND to REL (decreasing dimensionality) is determined completely by the move from REL to IND described above:  $\mathbf{p} = \mathbf{p}'$ ,  $\boldsymbol{\tau} = \boldsymbol{\tau}'$ ,  $\boldsymbol{\theta} = \boldsymbol{\theta}'$ ,  $\mu_{j \neq j'} = \mu'_{j \neq j'}$ , and  $u = \mu'_{j'}$ . We note that when decreasing the dimensionality, we do not have to simulate from a proposal distribution, but set  $\mu_{j'}$  at the value determined by the proposed relationship to  $\mu_j$ .

In step 2b, the probability of accepting a proposed move is equal to  $\min(1, A)$ , where:

$$A = \frac{p(\boldsymbol{\theta}', m' | \mathbf{y}) P(m | m') q'(\mathbf{u}')}{p(\boldsymbol{\theta}, m | \mathbf{y}) P(m' | m) q(\mathbf{u})} \left| \frac{\partial(\boldsymbol{\theta}', \mathbf{u}')}{\partial(\boldsymbol{\theta}, \mathbf{u})} \right|, \quad (16)$$

where  $P(m | m')$  denotes the probability of proposing to move from  $m$  to  $m'$ ,  $\left| \frac{\partial(\boldsymbol{\theta}', \mathbf{u}')}{\partial(\boldsymbol{\theta}, \mathbf{u})} \right|$  is the Jacobian, and  $p(\boldsymbol{\theta}, m | \mathbf{y})$  is the joint distribution of  $\boldsymbol{\theta}$  and the model  $m$ :  $p(\boldsymbol{\theta}_m, m | \mathbf{y}, \mathbf{z}) \propto p(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}_m, m) p(\boldsymbol{\theta}_m | m) p(m)$ .

In our implementation, each group  $j$  has an equal chance of being selected and flipped within each iteration,  $P(m' | m) = P(m | m')$ . Therefore, the  $A$  term simplifies to

$$\begin{aligned} A_{j' : \text{REL} \rightarrow \text{IND}} &= \frac{p(\boldsymbol{\theta}', m' | \mathbf{y}, \mathbf{z}) P(m | m') q'(\mathbf{u}')}{p(\boldsymbol{\theta}, m | \mathbf{y}, \mathbf{z}) P(m' | m) q(\mathbf{u})} \left| \frac{\partial(\boldsymbol{\theta}', \mathbf{u}')}{\partial(\boldsymbol{\theta}, \mathbf{u})} \right| \\ &= \frac{p(\boldsymbol{\theta}', m' | \mathbf{y}, \mathbf{z})}{p(\boldsymbol{\theta}, m | \mathbf{y}, \mathbf{z}) q(\mathbf{u})} \\ &= \frac{f(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}', m') p(\boldsymbol{\theta}' | m') p(m')}{f(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}, m) p(\boldsymbol{\theta} | m) p(m) q(\mathbf{u})}. \end{aligned} \quad (17)$$

The reverse move from the model where the estimation of group  $j'$  is changed from IND to REL can be described by the inverse of the stated acceptance probability such that  $\min(1, A_{j' : \text{IND} \rightarrow \text{REL}}) = \min(1, A_{j' : \text{REL} \rightarrow \text{IND}}^{-1})$ .

The final step in 2b is to simulate a single value  $v \sim \text{Unif}(0, 1)$  and accept the move to the proposed state if  $v \leq \min(1, A)$ . If the proposed move is not accepted, the original model  $m$  continues to the next iteration with its original parameter values.

### 3 | SIMULATION STUDIES TO ESTABLISH SMALL SAMPLE PROPERTIES

In this section, we present the results of a simulation study to evaluate the small sample properties of the model averaging approach discussed in Section 2.2. Our simulation study will mimic the data collected in CENIC-p1s1; we consider five dose levels, with dose level 5 treated as the normal nicotine condition, dose levels 1 to 4 treated as varying levels of reduced nicotine content conditions and  $n_j = 100$  for all dose levels. Data for the normal nicotine condition are drawn from a single component,  $\text{LogNorm}(\mu = 4, \sigma = 0.668)$ , where  $\mu$  and  $\sigma$  are the mean and standard deviation on the natural log scale. The standard deviation,  $\sigma = 0.668$ , was chosen such that the coefficient of variation was approximately 0.75, which is similar to what was observed in CENIC-p1s1. Data for the four reduced nicotine conditions are simulated from a two-component mixture distribution with true probability of noncompliance,  $p_j$ , equal to 0.70, which is approximately equal to the estimated probability of noncompliance for the 0.4-mg/g group reported by Nardone et al<sup>2</sup> and Boatman et al.<sup>7</sup> Data for the noncompliant component were drawn from a  $\text{LogNorm}(\mu = 4, \sigma = 0.668)$  distribution for all dose levels. We note that this is equal to the distribution of the biomarkers in the normal nicotine condition but this is not an assumption of our model, nor do we assume a common distribution for noncompliers across conditions. Data for the compliant components were drawn from a normal distribution with  $\sigma = 0.668$  and a mean that varied as a function of dose level. For the remainder of this section, we specify the mean for simulating the data for the reduced nicotine content groups by the effect size, defined as  $ES = \frac{4 - \mu_j}{0.668}$ . That is, we specify the effect size, which in turn specifies  $\mu_j$ . For purposes of estimation, the hypothesized dose-response relationship for the REL model assumes reductions of 93.1%, 86.5%, 73.7%, and 48.7% in the mean of the biomarker for groups 1 through 4 relative to the normal nicotine condition, corresponding to ES of 4, 3, 2, and 1 for groups 1 through 4, respectively. For all  $j$  dose levels, we assume “noninformative” prior specifications of  $\alpha_j = \beta_j = 1$  for the beta prior on  $p_j$ ,  $a_j = b_j = 0.001$  for the gamma prior on  $\tau_j$ , and  $s_{\mu_j} = s_{\theta_j} = 0.00001$  for the normal priors on  $\mu_j$  and  $\theta_j$ .

We consider five scenarios to evaluate the effect of varying the relationship between the mean of the compliant component and the mean of the normal nicotine on our ability to estimate quantiles of the compliant component. In four of

the scenarios, we vary the ES for one dose level in increments of 0.5, while assuming that the hypothesized relationship holds true for the other three dose levels. For example, when the proposed relationship for dose level 1 varies and all other levels are fixed, we have  $ES = (X, 3, 2, 1)$ , where the model is correctly specified when  $X = 4$ , and misspecified otherwise. A fifth scenario misspecifies the relationship for all dose levels while varying dose level 3 as a sensitivity analysis where  $ES = (6, 2, X, 0.1)$ .

Five hundred simulated studies were completed for each scenario and each ES increment. Simulation results are presented for the IND model, the REL model with our hypothesized relationship, and for our RJMCMC approach that averages over the IND, REL, and intermediate approaches. Four different model priors are chosen for the RJMCMC simulations: assuming each model specification is equally likely (RJ<sub>50</sub>) with  $p(I_j = 1) = 0.5$  for each group  $j$ , versus two priors favoring models assuming IND for each level (RJ<sub>95</sub> and RJ<sub>99</sub>) with  $p(I_j = 1) = 0.95$  and  $p(I_j = 1) = 0.99$  for each group  $j$ , versus favoring models assuming REL for each level (RJ<sub>05</sub>) with  $p(I_j = 1) = 0.05$  for each group  $j$ . For each simulated data set, model parameters were estimated using MCMC with a single chain with 10 000 total iterations, with the first 1000 iterations excluded for the burn-in period. The performance of each model is summarized as a ratio relative to the IND model in order to illustrate the gain in efficiency due to borrowing on the mean square error (MSE) for the 95th percentile on the original scale, the primary quantity of interest for the analysis of the data from CENIC-p1s1. Additional scenarios with performance summarized in greater detail by bias, standard deviation (SD), and MSE within each dose level  $j$  for the 80th, 90th, and 95th percentile of the distribution of the biomarker on the original scale are presented in the Supplementary Materials.

Simulation results for the scenario where dose level 3 varies and all other dose levels are fixed at their correctly specified ES are presented in Figure 2, with the other “fixed” scenario figures included in the Supplementary Materials since the resulting behavior is similar. We first note that varying the ES in dose level 3 does not impact estimation of the other, correctly specified, dose levels for all models. That is, the MSE for dose levels 1, 2, and 4 are constant as dose level 3 varies, with the REL model having the lowest MSE and the MSE increasing as the prior probability of independence increases. The RJ models perform in-between the IND and REL models by retaining improved MSE when the relationship is correct, but some inflation to the MSE when it is incorrect. For dose level 3, the RJ<sub>05</sub> and RJ<sub>50</sub> borrow too aggressively, but RJ<sub>95</sub> and RJ<sub>99</sub> provide a 50% and 25% reduction, respectively, in the MSE versus the IND model when the model is correctly specified (compared to 94% for REL) but have a maximum inflation of 43% and 17%, respectively, compared to 1032% and 288% for RJ<sub>05</sub> and RJ<sub>50</sub>, respectively. These results suggest that priors that are more conservative in assuming the REL model may be more appropriate.

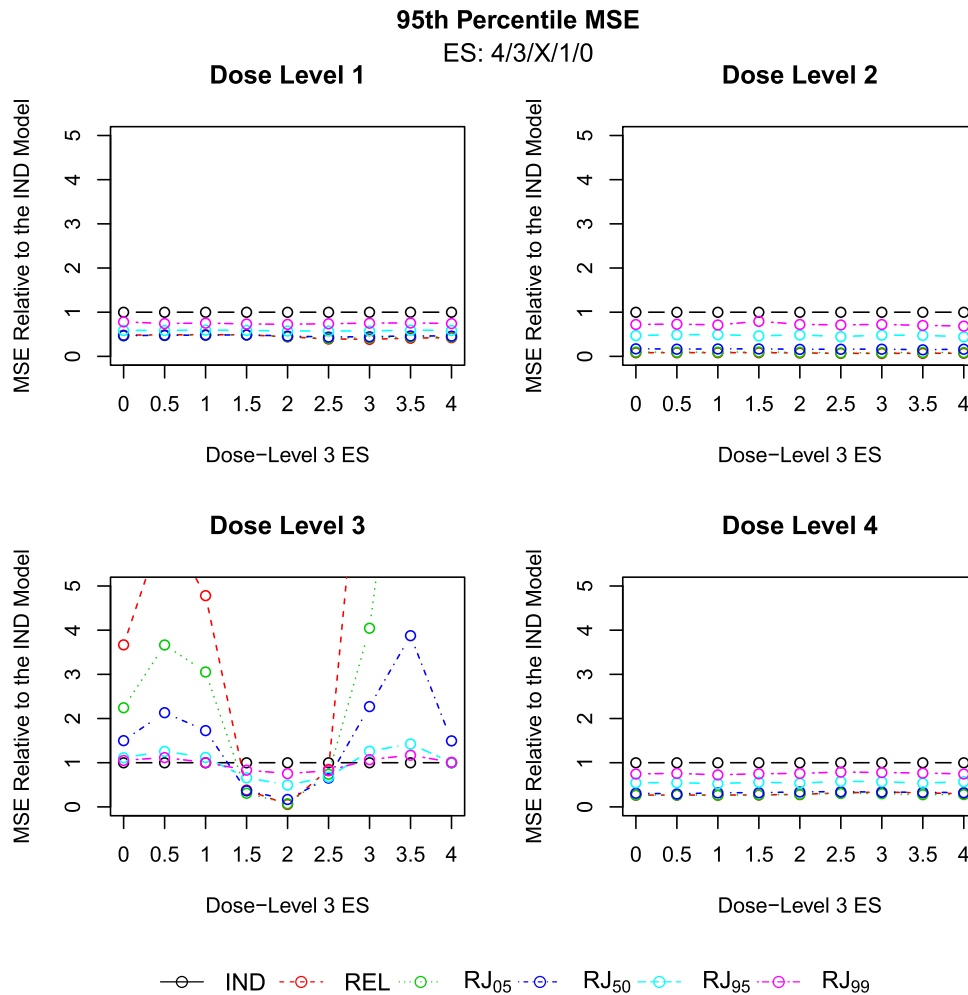
Results for the scenario where the relationship is misspecified for all dose levels while varying dose level 3 can be found in Figure 3. This scenario allows us to evaluate if misspecification of the other dose levels impacts the MSE of the dose level with varying ES. We observe that the trajectories of dose level 3 are similar to those in Figure 2, suggesting that, for a given dose level, the correct or incorrect specification of other dose levels has minimal impact on estimation.

## 4 | APPLICATION TO CENIC-P1S1

This section provides a case study of our model averaging approach applied to the data from CENIC-p1s1. As mentioned in Section 1, previous research has identified TNE cutoffs that can be used to identify noncompliance among subjects randomized to the 0.4-mg/g groups. However, investigators are also interested in understanding the effect of nicotine reduction among smokers that complied to the intervention in the intermediate nicotine content groups where biomarker thresholds for identifying noncompliance are not currently available.

Our analysis will focus on TNE (a biomarker of nicotine exposure that measures most nicotine metabolites) measured at the end of the 6-week study as our biomarker for identifying noncompliance among those who self-report compliance. For the purposes of this analysis, we will analyze TNE on the log scale and assume that all mixture components for the log-transformed biomarker values are normally distributed. CENIC-p1s1 included two normal nicotine conditions: a usual brand (UB) condition, who received their preferred brand of commercial cigarettes, and a 15.8-mg/g study cigarette (roughly equivalent to the nicotine content of commercial cigarettes). In addition, CENIC-p1s1 included five experimental conditions: a 5.2-mg/g group, a 2.4-mg/g group, a 1.3-mg/g group, a 0.4-mg/g group, and a 0.4-mg/g condition with elevated tar yield (0.4-mg/g [HT] group) to understand the impact of tar yield on the effect of nicotine reduction. For the purposes of this analysis, we combine the two normal nicotine conditions and the two 0.4-mg/g groups because these groups have the same nicotine content, respectively, which is the primary factor influencing the distribution of the





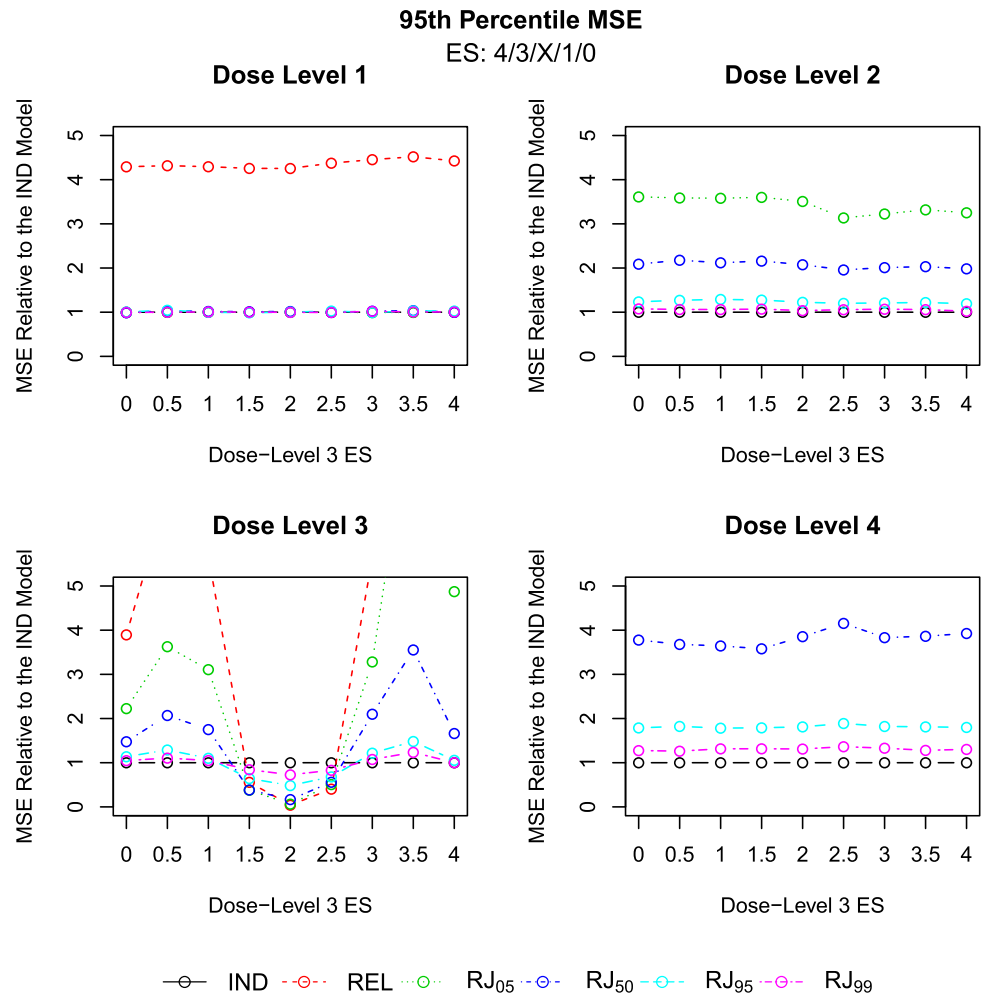
**FIGURE 2** MSE for each modeling approach relative to IND model for a true ES = (4, 3, 2, 1) where the proposed REL assumes ES = (4, 3, X, 1), such that dose level 3 varies over a grid of values from 0 to 4 and dose levels 1, 2, and 4 are correctly specified [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

biomarkers. The proposed relationship from Section 2.1 is used with  $w_j = (0.4, 1.3, 2.4, 5.2, 15.8)$ , corresponding to a 97.5%, 91.8%, 85.0%, and 66.7% reduction (effect size of  $-3.7, -2.5, -1.9$ , and  $-1.1$ ) in the average biomarker value for the VLNC group, 1.3-mg/g group, 2.4-mg/g group, and 5.2-mg/g group, respectively, relative to the combined usual brand/15.8-mg/g normal nicotine group. Table 7 and Figure 5 in the Supplementary Materials provide additional descriptive summaries for the treatment groups in CENIC-p1s1. There is a clear bimodal pattern in the VLNC and 1.3-mg/g groups, but not in the 2.4- and 5.2-mg/g groups, which could be due the extreme context of there being no truly compliant subjects or, more likely, the distributions of  $\log(\text{TNE})$  for compliant and noncompliant subjects have sufficient overlap to make distinct bimodal patterns difficult to identify.

For all dose levels, we assume “noninformative” prior specifications of  $\alpha_j = \beta_j = 1$  for the beta prior on  $p_j$ ,  $a_j = b_j = 0.001$  for the gamma prior on  $\tau_j$ , and  $s_{\mu_j} = s_{\theta_j} = 0.00001$  for the normal priors on  $\mu_j$  and  $\theta_j$ . The REL and IND models were fit using three chains with different, overdispersed initial values of length 20 000 with a burn-in of 2000 iterations. Chain length, burn-in, and MCMC convergence of the parameters were established by Gelman-Rubin diagnostic values<sup>16</sup> below 1.005, autocorrelation plots with a lag up to 50 iterations, and trace plots. The RJMCMC models were fit on three chains with a length of 400 000 with a burn-in of 200 000 iterations. Convergence is challenging for reversible jump algorithms since other methods for MCMC diagnostics are not applicable due to the changing model space,<sup>15</sup> but utilizing longer chains and discarding the first half of the observations provide a greater opportunity to explore the total model space of 16 potential models.

The results of applying both approaches and the RJMCMC algorithm assuming a prior that favors the IND model,  $p(I_j) = 0.95$ , for each group  $j$  are presented in Table 1. The RJMCMC prior of  $p(I_j) = 0.95$  was selected based on the simulation results in Section 3 identifying more appropriate tradeoffs with MSE for priors favoring the IND model. Presented are the posterior means and 95% HPD intervals for the proportion compliant ( $\hat{p}_c$ ), the compliant component mean for

**FIGURE 3** MSE for each modeling approach relative to IND model for a true ES = (4, 3, 2, 1) where the proposed REL assumes ES = (6, 2, X, 0.1), such that dose level 3 varies over a grid of values from 0 to 4 and dose levels 1, 2, and 4 are all misspecified [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



TNE ( $\hat{\mu}_1$ ), the noncompliant component mean for TNE ( $\hat{\mu}_2$ ), and the 90th and 95th percentile of the biomarker distribution for compliant subjects. The last quantities have been suggested as thresholds for identifying noncompliance in trials of reduced nicotine content cigarettes.<sup>4</sup>

The resulting point estimates for the IND and REL models are similar for the VLNC, 1.3-mg/g, and 5.2-mg/g groups, suggesting that the hypothesized relationship between nicotine dose and biomarkers of nicotine exposure may be reasonable, with REL showing an improvement in efficiency through narrower 95% HPD intervals. As a result, the RJ<sub>95</sub> model, which heavily favors independence a priori, chooses the relationship over 60% of the time for the VLNC, 1.3-mg/g, and 5.2-mg/g groups, improving efficiency. Compared to the IND approach the width of the HPD intervals were 37% to 55% and 24% to 51% narrower for the 90th and 95th percentile, respectively. Figure 4 presents histograms of log(TNE) by group and the fitted distributions from the RJMCMC algorithm. We see that our approach results in a reasonable fit to the data, with the possible exception of the 2.4-mg/g group, where it is difficult to identify the two components. In additional simulation studies provided in Section 1.2 of the Supplemental Materials, we observed that substantial overlap in the mixture components can lead to bias in the estimated marginal probability of compliance, which may explain why the estimated probability of compliance is higher for the 5.2-mg/g group than the other groups.

Based on an application of the Bayes rule, the probability that a subject was compliant conditional on the observed biomarker values,  $P(K_{ij} = 1|y_{ij})$ , can be written as a function of the components of the biomarker mixture density:

$$P(K_{ij} = 1|y_{ij}) = \frac{(1 - p_j)\mathcal{N}(y_{ij}|\mu_j, \tau_j)}{(1 - p_j)\mathcal{N}(y_{ij}|\mu_j, \tau_j) + p_j\mathcal{N}(y_{ij}|\mu_j + \theta_j, \tau_j)}.$$

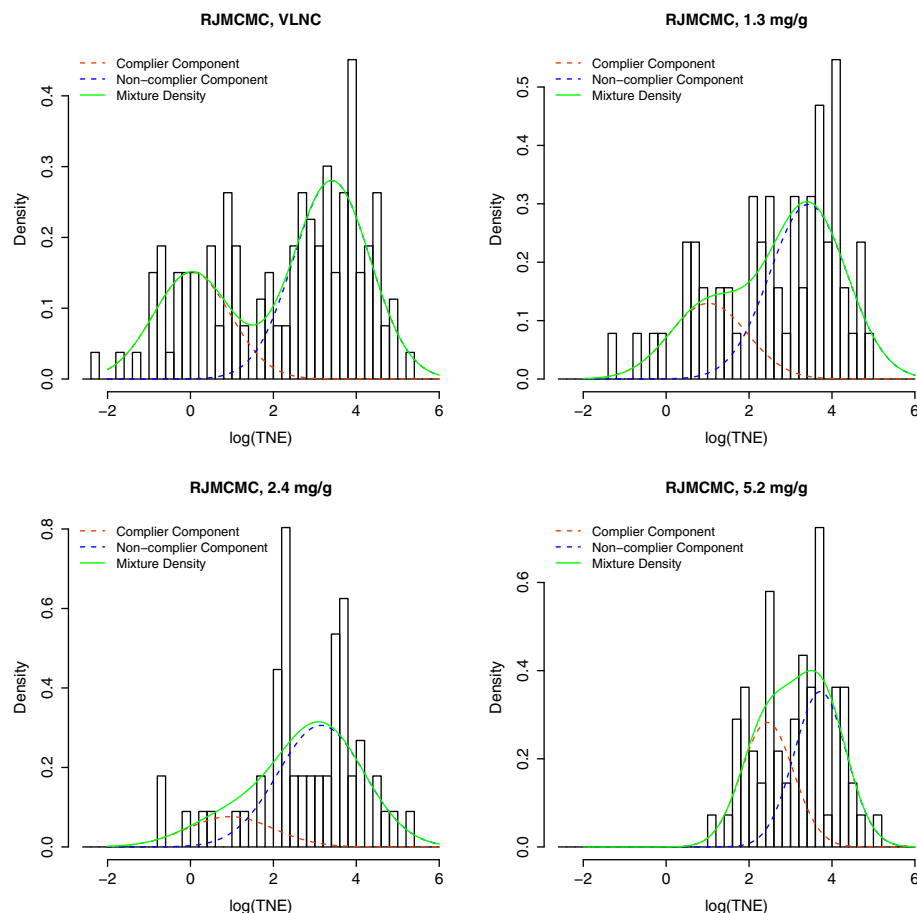
Figure 5 presents the estimated probability of compliance as a function of the biomarkers from the IND and RJ<sub>95</sub> models. With the exception of the 2.4-mg/g group, the curves for RJ<sub>95</sub> are similar to IND, suggesting that any bias due to

**TABLE 1** Mean (95% HPD interval) of proportion in compliant group ( $\hat{p}_c$ ), compliant component mean TNE ( $\hat{\mu}_1$ ), noncompliant component mean TNE ( $\hat{\mu}_2$ ), 90/95th percentile of  $\hat{\mu}_1$ , and proportion of the chain spent in the REL approach for the RJMCMC with  $p(I_j = 1) = 0.95$  (RJ<sub>95</sub>)

Group	Approach	$\hat{p}_c$	$\hat{\mu}_1$	$\hat{\mu}_2$	Percentile for $\mu_1$		Proportion in REL
					90th	95th	
VLNC	IND	0.36 (0.27, 0.45)	1.90 (1.28, 2.57)	47.95 (36.90, 60.17)	4.04 (2.71, 5.60)	5.64 (3.64, 7.99)	
	REL	0.34 (0.26, 0.43)	1.42 (1.17, 1.69)	45.99 (35.45, 57.13)	3.03 (2.40, 3.71)	4.26 (3.24, 5.41)	
	RJ <sub>95</sub>	0.35 (0.26, 0.44)	1.62 (1.15, 2.34)	46.81 (35.96, 58.56)	3.46 (2.39, 5.05)	4.85 (3.20, 7.16)	0.626
1.3 mg/g	IND	0.28 (0.11, 0.45)	4.08 (1.11, 7.69)	47.21 (30.25, 65.94)	8.75 (2.41, 17.11)	12.78 (3.25, 25.71)	
	REL	0.31 (0.17, 0.46)	4.68 (3.41, 6.29)	49.57 (33.49, 68.24)	9.99 (6.63, 14.12)	14.17 (8.65, 21.67)	
	RJ <sub>95</sub>	0.30 (0.15, 0.45)	4.46 (2.10, 6.41)	48.87 (32.74, 67.16)	9.50 (4.21, 14.27)	13.49 (5.41, 21.52)	0.893
2.4 mg/g	IND	0.16 (0.01, 0.43)	4.52 (0.08, 19.09)	39.87 (23.54, 59.32)	9.98 (0.19, 42.95)	15.08 (0.31, 66.91)	
	REL	0.27 (0.05, 0.48)	10.32 (7.04, 14.42)	47.38 (27.30, 70.97)	22.92 (14.65, 32.75)	34.60 (19.58, 53.20)	
	RJ <sub>95</sub>	0.20 (0.02, 0.44)	6.69 (0.28, 14.11)	42.43 (24.38, 65.06)	14.81 (0.65, 31.96)	22.32 (0.80, 51.49)	0.455
5.2 mg/g	IND	0.43 (0.12, 0.75)	13.92 (7.24, 28.56)	50.86 (32.55, 67.25)	26.08 (12.67, 59.22)	33.52 (15.11, 80.96)	
	REL	0.45 (0.24, 0.66)	14.73 (11.51, 18.77)	51.64 (35.44, 67.08)	27.30 (18.72, 38.35)	34.70 (22.07, 52.16)	
	RJ <sub>95</sub>	0.44 (0.22, 0.66)	14.32 (9.91, 19.35)	51.22 (35.54, 66.55)	26.44 (16.10, 39.22)	33.57 (18.75, 52.90)	0.919
UB/	IND	—	74.63 (57.76, 92.85)	—	—	—	
15.8 mg/g	REL	—	75.48 (59.83, 92.71)	—	—	—	
	RJ <sub>95</sub>	—	74.06 (57.91, 92.00)	—	—	—	

Abbreviations: HPD, highest posterior density; TNE, total nicotine equivalents.

**FIGURE 4** Histograms for distribution of self-reported compliers of  $\log(\text{TNE})$  in each CENIC-p1s1 treatment arm at week 6 with mixture densities for the RJ<sub>95</sub> approach [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

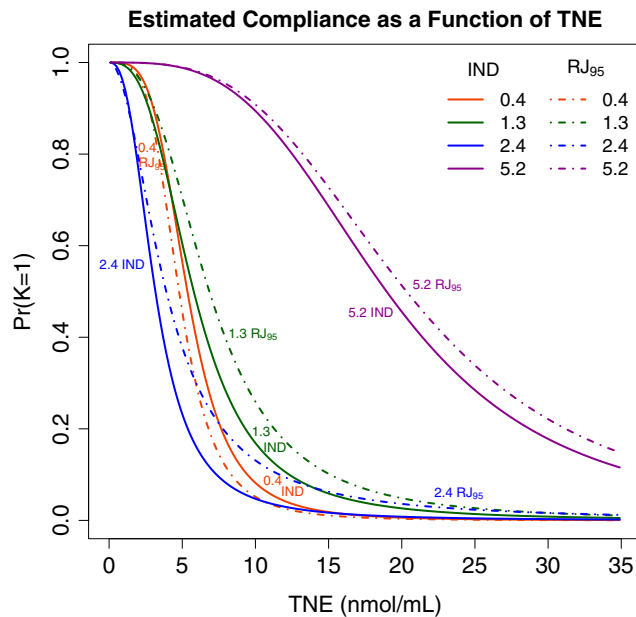


borrowing is minimal and outweighed by the increased precision, as noted in Table 1. We observe a larger difference between the IND and RJ<sub>95</sub> curves for the 2.4-mg/g group, but we note that the curve for the 2.4-mg/g group is to the left of the curve for the 0.4-mg/g group. This suggests that the estimated threshold for declaring noncompliance in the 2.4-mg/g group would be less than the threshold for the 0.4-mg/g group, which is biologically implausible given that the 2.4-mg/g cigarettes have six times the nicotine of the 0.4-mg/g group. The RJ<sub>95</sub> curve for the 2.4-mg/g group is pulled to the right of the IND curve but still results in a biologically implausible curve that is below the curve for the 1.3-mg/g group. In Section 3, we showed via simulation that the estimation of the 2.4-mg/g group should have little impact on estimation for the other three dose levels. To validate this conclusion, we refit our model without the 2.4-mg/g group. The results, which can be found in Section 2.1 of the Supplementary Materials, illustrate that we obtain similar estimates for the VLNC, 1.3-mg/g, and 5.2-mg/g groups as were observed when the 2.4-mg/g group was included in the model.

While simulation study results demonstrate that any one dose level has minimal impact on estimation of the remaining dose levels, we provide a sensitivity analysis in the Supplementary Materials, which demonstrates that the removal of the 2.4-mg/g group does not greatly change estimation of the other dose levels.

## 5 | DISCUSSION

The ability to identify noncompliance has important implications for interpreting the results of CENIC-p1s1. Previous work to identify noncompliance in CENIC-p1s1 leveraged the availability of auxiliary data to determine cutoffs for identifying noncompliance and to estimate the probability of compliance conditional on the observed biomarker values for the VLNC group, only.<sup>7</sup> However, auxiliary data are not available for the intermediate dose levels. The method proposed by Boatman et al<sup>7</sup> could be used to estimate the probability of noncompliance for the intermediate dose levels, but the estimates will be inefficient in the absence of auxiliary data from known compliers. We



**FIGURE 5** Predicted compliance ( $\Pr(K = 1)$ ) as a function of observed total nicotine equivalent (TNE) by study group for IND approach and model averaging with  $RJ_{95}$  [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

propose a fully Bayesian approach to estimate the mixture components across all groups simultaneously by averaging over the model space, which included different assumptions regarding the association between the nicotine content of the cigarettes and biomarkers of nicotine exposure. Simulation results were encouraging, with model averaging achieving more precise estimates of the mixture components by choosing the state assuming the hypothesized relationship when the relationship was true and downweighting the hypothesized relationship when the relationship was misspecified.

One of the objectives of CENIC-p1s1 was to collect data that would inform the FDA as they consider new product standards for cigarettes, including a potential reduction in the nicotine content of cigarettes. The intention-to-treat analysis provides an estimate of the effect of an intervention, in practice, which may include noncompliance to the intervention, but it does not reflect the potential future reality where policy choices may lead to previous options disappearing from the market due to regulation (ie, the FDA using its power to lower nicotine content of cigarettes). This motivates the desire to estimate the causal effect of nicotine reduction, which requires that investigators are able to accurately identify noncompliance to randomized treatment assignment. A primary concern of nicotine reduction as a regulatory strategy is that it may lead to compensatory behaviors (eg, puffing more intensely to obtain more nicotine), and models to estimate compliance are needed that can flexibly accommodate this behavior. Application to the data from CENIC-p1s1 suggests that the proposed relationship is appropriate for some dose levels, but not for others, potentially due to compensatory behavior or other factors that might lead to a misspecified relationship. Our results indicate that our model is flexible enough to borrow strength when the relationship holds, without leading to bias, when it does not. However, in cases such as the 2.4-mg/g dose where an improbable biological trend is observed relative to other doses, future studies should be conducted before using the estimated cut-offs in practice. Previously, investigators were only able to identify noncompliance in the VLNC group, but the method proposed in this article provides a framework for identifying noncompliance at other dose levels, as well. This is particularly important because a number of randomized trials of reduced nicotine content cigarettes are currently under way and some use dose levels other than the 0.4-mg/g dose.

The selection of a possible biological relationship,  $h_j(\mu_j)$ , is an important consideration when implementing the proposed framework. CENIC-p1s1 benefited from previous work on nicotine metabolism that provided potential biological relationships.<sup>13</sup> In addition, data from the known compliers to VLNC cigarettes from the hotel sequestration study could be leveraged to evaluate the appropriateness of the relationship.<sup>4</sup> In practice, researchers will need to consider prior data or expert opinion to specify biologically reasonable relationships. As one avenue of future work, we will also consider expanding the framework to consider multiple proposed biological relationships. This would facilitate the simultaneous consideration of multiple relationships when there is not a consensus from the prior research.




Mixture models can be challenging to estimate in practice, with problems such as label switching or unstable estimates. In the Bayesian context, these can be addressed through assumptions regarding the relationships between components (ie, forcing one component to have a larger mean than the other) and by monitoring the behavior of the resulting MCMC chains to ensure convergence to a steady state. While our Gaussian mixture model is for a finite, two-component mixture model, our approach could be extended to any finite number of components. For example, our analysis only considered subjects that self-report compliance, in which case there are two groups: subjects that honestly self-report compliance and subjects that self-report compliance but were actually noncompliant. Alternately, we could analyze all subjects, which would result in a third mixture component for subjects that honestly self-report noncompliance, which would be more appropriate than including them with noncompliers who self-reported compliance.<sup>11</sup>


The proposed Bayesian framework represents a flexible approach to identifying noncompliance in regulatory tobacco trials, but could be applied to other settings as well. For instance, the approach may be beneficial in therapeutic clinical trials where multiple doses are considered, but pharmacokinetic data are only available for one dose. In addition, future work will extend our proposed modeling framework to a regression setting where biomarkers can be modeled as a function of other covariates of interest that may impact the observed biomarker values, such as cigarettes per day, with the potential to incorporate longitudinal data. Additionally, the results of this article could be used to estimate graphical compliance networks, which can be combined with causal inference techniques to estimate causal effects for the intermediate dose levels.

## ACKNOWLEDGEMENTS

The authors would like to thank their collaborator, Dr. Eric Donny, for providing the data used in Section 4. This research was partially funded by NIH grants P30-CA077598 from the National Cancer Institute and R01-DA046320, R03-DA041870, and U54-DA031659 from the National Institute on Drug Abuse and FDA Center for Tobacco Products (CTP). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or FDA CTP.

## ORCID

Alexander M. Kaizer  <https://orcid.org/0000-0003-2334-5514>

Joseph S. Koopmeiners  <https://orcid.org/0000-0001-6147-3168>

## REFERENCES

1. Donny EC, Denlinger RL, Tidey JW, Koopmeiners JK, et al. Randomized trial of reduced-nicotine standard for Cigarettes. *N Engl J Med*. 2015;373(14):1340-1349.
2. Nardone N, Donny EC, Hatsukami DK, et al. Estimations and predictors of non-compliance in switchers to reduced nicotine content cigarettes. *Addiction*. 2016;111(12):2208-2216.
3. Benowitz NL, Nardone N, Hatsukami DK, Donny EC. Biochemical estimation of noncompliance with smoking of very low nicotine content cigarettes. *Cancer Epidemiol Prev Biomar*. 2015;24(2):331-335.
4. Denlinger RL, Smith TT, Murphy SE, et al. Nicotine and anatabine exposure from very low nicotine content cigarettes. *Tob Regul Sci*. 2016;2(2):186-203.
5. Tidey JW, Pacek LR, Koopmeiners JS, et al. Effects of 6-week use of reduced-nicotine content cigarettes in smokers with and without elevated depressive symptoms. *Nicotine Tob Res*. 2017;19(1):59-67.
6. Rupprecht LE, Koopmeiners JS, Dermody SS, et al. Reducing nicotine exposure results in weight gain in smokers randomised to very low nicotine content cigarettes. *Tob Control*. 2017;26:e43-e48.
7. Boatman JA, Vock DM, Koopmeiners JS, Donny EC. Estimating causal effects from a randomized clinical trial when noncompliance is measured with error. *Biostatistics*. 2017;19:103-118.
8. Marin J-M, Mengersen K, Robert Christian P. Bayesian modelling and inference on mixtures of distributions. *Handb Stat*. 2005;25: 459-507.
9. Jasra A, Holmes CC, Stephens DA. Markov Chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Stat Sci*. 2005;20:50-67.
10. Stephens M. Dealing with label switching in mixture models. *J Royal Stat Soc: Ser B (Stat Method)*. 2000;62(4):795-809.
11. Boatman JA, Casty K, Vock DM, et al. Classification accuracy of biomarkers of compliance. *Tob Regul Sci*. 2019;5(3):301-319.
12. Chib S. Calculating posterior distributions and modal estimates in Markov mixture models. *J Econ*. 1996;75(1):79-97.
13. Benowitz NL, Hukkanen J, Jacob P. Nicotine chemistry, metabolism, kinetics and biomarkers. *Nicotine psychopharmacology*. Berlin, Heidelberg: Springer; 2009:29-60.

14. Green PJ. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika*. 1995;82(4): 711-732.
15. Green PJ, Hastie DI. Reversible jump MCMC. *Genetics*. 2009;155(3):1391-1403.
16. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7:457-472.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Kaizer AM, Koopmeiners JS. A fully Bayesian mixture model approach for identifying noncompliance in a regulatory tobacco clinical trial. *Statistics in Medicine*. 2020;39:1328–1342.  
<https://doi.org/10.1002/sim.8478>