Received: 18 December 2021 | Revised: 14 November 2022 | Accepted: 17 January 2023

DOI: 10.1002/sim.9676

RESEARCH ARTICLE

Statistics in Medicine WILEY

Analysis of N-of-1 trials using Bayesian distributed lag model with autocorrelated errors

Ziwei Liao[1] | **Min Qian**[1] | **Ian M. Kronish**[2] | **Ying Kuen Cheung**[1]

[1]Department of Biostatistics, Columbia University, New York, USA

[2]Center for Behavioral Cardiovascular Health, Columbia University, New York, USA

Correspondence
Ying Kuen Cheung, Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY, USA
Email: yc632@columbia.edu

Funding information
National Cancer Institute, Grant/Award Number: HHSN261200800001E; National Institutes of Health, Grant/Award Number: R01MH109496; National Science Foundation, Grant/Award Number: DMS-2112938

An N-of-1 trial is a multi-period crossover trial performed in a single individual, with a primary goal to estimate treatment effect on the individual instead of population-level mean responses. As in a conventional crossover trial, it is critical to understand carryover effects of the treatment in an N-of-1 trial, especially when no washout periods between treatment periods are instituted to reduce trial duration. To deal with this issue in situations where a high volume of measurements are made during the study, we introduce a novel Bayesian distributed lag model that facilitates the estimation of carryover effects, while accounting for temporal correlations using an autoregressive model. Specifically, we propose a prior variance-covariance structure on the lag coefficients to address collinearity caused by the fact that treatment exposures are typically identical on successive days. A connection between the proposed Bayesian model and penalized regression is noted. Simulation results demonstrate that the proposed model substantially reduces the root mean squared error in the estimation of carryover effects and immediate effects when compared to other existing methods, while being comparable in the estimation of the total effects. We also apply the proposed method to assess the extent of carryover effects of light therapies in relieving depressive symptoms in cancer survivors.

KEYWORDS

Bayesian distributed lag model, carryover effects, N-of-1 trials, personalized treatment, regression with autocorrelated errors, time series

1 | INTRODUCTION

N-of-1 trials are multi-period crossover studies that compare two or more interventions in single individuals, and are suitable for evaluating personalized treatment effects in those with chronic conditions where the outcome is relatively stable.[1] Advances in mobile and sensor technology[2] and better understanding of patient preferences[3] have improved the implementation of N-of-1 trials. However, their uptake remains very small in clinical practice. In particular, the duration of N-of-1 trials remains a key barrier. To reduce the duration needed to conduct an N-of-1 trial and to reduce the burden of participation, it is often necessary to preclude scheduling washout periods between treatments. When physical washout periods are not feasible, it is critical to have provisions for dealing with carryover effects analytically. To motivate our work, consider an N-of-1 trial series that compare bright white light (10 000 lux) and dim red light (50 lux) in cancer patients with depressive symptoms, where light therapy was delivered by portable light boxes with instructions.[4] Briefly, each individual would use one of two light boxes for 30 minutes each morning over a 12 weeks. Along with the light boxes, a smartphone application would be used to give treatment reminders and to assess daily depressive symptoms and fatigue

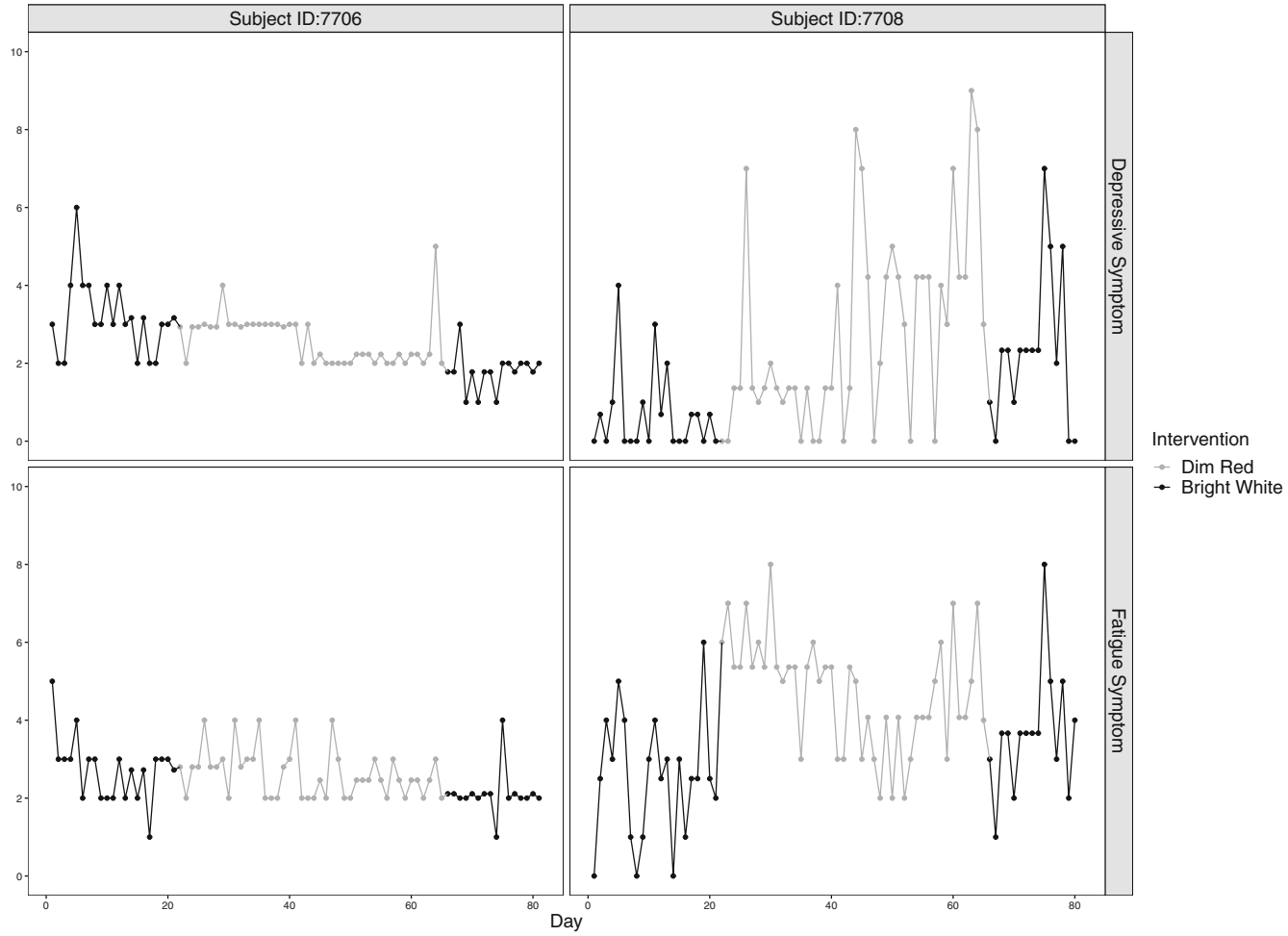

FIGURE 1 Daily assessments of two patients id 7706 and 7708. Black line represents bright white light intervention, and grey line represents dim red light.

level over the entire 12-week period. While theory suggests bright white light may reduce cancer-related depression and fatigue, its effects may vary from individual to individual.[5] Thus, the primary analytical goal in light therapy study is to identify for each individual whether bright white light is superior in terms of symptom control and make light therapy suggestion for their further clinical treatment. Figure 1 shows the daily assessments of two patients during the study course.

In a systematic review of 108 N-of-1 trial series published between 1985 and 2010, Gabler et al reported on the analytic methods used to compare the effectiveness of two or more treatments being studied in an N-of-1 trial, including graphical comparison, hypothesis tests (eg, *t*-test, nonparametric tests), and regression models.[6] While there is no single agreed-upon analysis method, these methods ignore two key features of experimental N-of-1 data. First, most methods do not account for temporal dependence (ie, autocorrelation) between assessments. Second, the methods do not capture the carryover effects of an intervention. The second data feature, which motivates this article, can be partly addressed by using a distributed lagged model (DLM), which is widely used in economics,[7,8] advertising,[9] and environmental health studies.[10,11] A DLM postulates that the current value of the outcome variable depends on the previous values (lags) of an exposure as well as the current exposure value, thus allowing the total exposure effect to be distributed over a time period and facilitating explicit modeling of carryover effects. A potential challenge in fitting a DLM is collinearity of the exposure lags. The N-of-1 trial design will further aggravate the problem: as illustrated in Figure 1, the exposure (light box) often remains the same as in the previous day in order to avoid switching intervention too frequently during a trial. A strategy to handle collinearity in DLM is by putting parametric constraints on the lag coefficients such as geometric lags,[7] or polynomial lags.[8] Alternatively, one may consider putting informative prior on the coefficients in a Bayesian framework.[10]

In this article, we adopt the Bayesian framework and propose a Bayesian distributed lag model with autocorrelated errors (BDLM-AR) as an extension of DLMs for N-of-1 trial data. The model is novel in several ways. First, we propose a prior distribution that constrains the lag coefficients with shrinkage factors increasing over time. Second, we impose a fused ridge-type penalty to address collinearity, which may be viewed as a variant of the fused lasso method.[12,13] Third, while current DLM methods assume independent error terms, we incorporate temporal correlations using an autoregressive error model. We will introduce the proposed BDLM-AR with details in Section 2, and describe the posterior computations in Section 3. The performance of BDLM-AR will be evaluated and compared with other methods by simulation studies presented in Section 4. We will apply the proposed method to the light therapy data in Section 5, and will conclude this article with a discussion in Section 6. Technical details and additional numerical results are given in the online Supporting Information.

2 | BAYESIAN DISTRIBUTED LAG MODEL WITH AUTOCORRELATED ERROR

2.1 | Proposed model

Suppose we observe data from a patient on n consecutive days. On day $t = 1, \ldots, n$, let X_t and Y_t denote the binary treatment indicator and the continuous outcome of interest, respectively. We consider a distributed lag autoregressive model for Y, described as follows:

$$Y_t = \mu + Z_t'\boldsymbol{b} + \sum_{l=0}^{L} \beta_l X_{t-l} + \epsilon_t, \tag{1}$$

for $t = p+1, \ldots, n$, where Z_t is a q-dimension vector, representing time varying covariates and the error term ϵ_t follows an autoregressive process,

$$\epsilon_t = \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + \cdots + \phi_p\epsilon_{t-p} + w_t, \tag{2}$$

w_t is a white Gaussian noise with mean zero and unknown variance $\sigma^2 > 0$, and L and p are prespecified. Note that for $t \le L$, the maximum lag effect is of order $t-1$, and terms involving X with nonpositive subscript are not included in the model.

Model (1) is composed of two parts. First, for the structural component, the mean model is specified by lag coefficients $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_L)'$ and control mean μ. Specifically, the effect of a given treatment sequence is represented by the sum of the corresponding lag coefficients. For example, if a patient receives the treatment for two successive days with no prior treatment, the effect due to this sequence is indicated by $\beta_0 + \beta_1$. Likewise, if another patient receives treatment today and the day before yesterday, but not yesterday, the effect of this treatment sequence will be $\beta_0 + \beta_2$. While the proposed method will allow us to estimate individual coefficients (hence the effect of any given treatment sequence), we will focus on a few specific parameters in this article. We define the total treatment effect by

$$\sum_{l=0}^{L} \beta_l = E(Y_t|X_t = 1, \ldots, X_{t-L} = 1) - E(Y_t|X_t = 0, \ldots, X_{t-L} = 0). \tag{3}$$

The total effect (3) is clinically meaningful because one of the goals of an N-of-1 trial is to determine whether the treatment should be given to the patient indefinitely if deemed effective.

The immediate treatment effect is measured by β_0, and the carryover effect due to treatment on l days ago is measured by β_l for $l > 0$. In the model, we assume the carryover effect beyond day L is zero. As such, the total carryover treatment effect due to treatment assuming the patient receives the treatment at all-time points before the current time but not at the current time is captured as

$$\delta \triangleq \sum_{l=1}^{L} \beta_l = E(Y_t|X_{t-1} = 1, \ldots, X_{t-L} = 1, X_t) - E(Y_t|X_{t-1} = 0, \ldots, X_{t-L} = 0, X_t). \tag{4}$$

Hence, the model naturally breaks down total treatment effect (3) into β_0 and δ. Apparently, if a patient has received the treatment on fewer timepoints than L from the current time, the carryover effect due to that treatment sequence will be different. However, the total carryover effect (4) indicates the maximal impact of the treatment over time.

Second, for the stochastic component, temporal dependency between errors is specified using an order-p autoregressive error model with autoregression coefficient $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$. Let B denote the backshift operator and $\Phi(B)$ be a polynomial in the backshift operator, that is, having $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ so that the autoregression model for the error terms can be written as $\Phi(B)\epsilon_t = w_t$. It is often convenient to work with the transformed data $Y_t^* = \Phi(B)Y_t$ and $X_t^* = \Phi(B)X_t$ in the estimation steps. Thus, applying $\Phi(B)$ to both sides of model (1), we will rewrite the model

$$Y_t^* = \mu^* + Z_t^{*'}\boldsymbol{b} + \sum_{l=0}^{L} \beta_l X_{t-l}^* + w_t, \tag{5}$$

for $t = p+1, \ldots, n$, where $\mu^* = \Phi(B)\mu$ and $Z_t^* = \Phi(B)Z_t$. To stack the data in vector form, we have

$$(\boldsymbol{Y}^* | \boldsymbol{X}^*, \mu^*, \boldsymbol{b}, \boldsymbol{\beta}) \sim N(\mu^* \mathbf{1}_{n-p} + \boldsymbol{Z}^*\boldsymbol{b} + \boldsymbol{X}^*\boldsymbol{\beta}, \sigma^2 \mathbf{I}_{n-p}) \tag{6}$$

where $\boldsymbol{Y}^* = (Y_{p+1}^*, \ldots, Y_n^*)'$, $\boldsymbol{X}^*$ is a $(n-p)\times(L+1)$ matrix with $X_{k-l+p+1}^*$ being the (k,l)th element of $\boldsymbol{X}^*$, $\boldsymbol{Z}^* = (Z_{p+1}^*, \ldots, Z_n^*)'$, $\mathbf{1}_{n-p}$ is a 1-vector of length $n-p$, and $\mathbf{I}_{n-p}$ is the identity matrix of dimension $n-p$. We denote $\tilde{\boldsymbol{\beta}} = (\mu, \boldsymbol{b}', \boldsymbol{\beta}')'$, $\tilde{\boldsymbol{X}} = (\mathbf{1}_n, \boldsymbol{Z}, \boldsymbol{X})$ and $\tilde{\boldsymbol{X}}^* = (\Phi(B)\mathbf{1}_{n-p}, \boldsymbol{Z}^*, \boldsymbol{X}^*)$, so that $\tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}} = \mu^*\mathbf{1}_{n-p} + \boldsymbol{Z}^*\boldsymbol{b} + \boldsymbol{X}^*\boldsymbol{\beta}$.

2.2 | Prior Distribution on the mean model

We consider normal prior distribution for $\tilde{\boldsymbol{\beta}}$, that is, having

$$\tilde{\boldsymbol{\beta}} \sim N(\mathbf{0}, \sigma^2 \tilde{\boldsymbol{\Omega}}^{-1}), \tag{7}$$

where $\tilde{\boldsymbol{\Omega}} = diag(c_0\mathbf{I}_{q+1}, \boldsymbol{\Omega})$ so that the prior variance of μ is $\sigma^2 c_0^{-1}$, the prior variance-covariance matrix of $\boldsymbol{b}$ is $\sigma^2 c_0^{-1}\mathbf{I}_q$ and the prior variance-covariance matrix of $\boldsymbol{\beta}$ is $\sigma^2\boldsymbol{\Omega}^{-1}$. We note that the prior variance depends on the variance σ^2 of the observations: such dependence renders a fused ridge penalized estimation procedure that is free of the variance parameters, resulting in computational stability; see Equation (9) below. We will postulate a noninformative prior on μ and $\boldsymbol{b}$ by setting c_0 to be a small number, and we will consider $(L+1)\times(L+1)$ matrix $\boldsymbol{\Omega}$ of the following form:

$$\begin{pmatrix} \lambda_0 + \lambda_0^* & -\lambda_0^* & 0 & \ldots & \ldots & 0 \\ -\lambda_0^* & \lambda_1 + \lambda_0^* + \lambda_1^* & -\lambda_1^* & \ldots & \ldots & 0 \\ 0 & -\lambda_1^* & \lambda_2 + \lambda_1^* + \lambda_2^* & \ldots & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_{L-1} + \lambda_{L-2}^* + \lambda_{L-1}^* & -\lambda_{L-1}^* \\ 0 & 0 & 0 & \ldots & -\lambda_{L-1}^* & \lambda_L + \lambda_{L-1}^* + \lambda_L^* \end{pmatrix}, \tag{8}$$

where the hyperparameters $\lambda_l, \lambda_l^* > 0$, for $l = 0, \ldots, L$, are constrained to increase over l. As a result of the monotonicity constraint, a lag coefficient β_l at a greater lag l is associated with a larger diagonal element in $\boldsymbol{\Omega}$, thus shrinking β_l toward the prior mean (zero) to a greater extent. This effectively addresses collinearity of the lag coefficients without imposing strong parametric structure to $\boldsymbol{\beta}$. In addition, using the normal prior (7) with precision matrix $\sigma^{-2}\tilde{\boldsymbol{\Omega}}$, we can show that the maximum *a posteriori* probability estimate of $\tilde{\boldsymbol{\beta}}$ minimizes a fused ridge-type penalty:

$$(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}}) + c_0\mu^2 + c_0\sum_{i=1}^{q} b_i^2 + \sum_{l=0}^{L} \lambda_l \beta_l^2 + \sum_{l=0}^{L} \lambda_l^*(\beta_l - \beta_{l+1})^2, \tag{9}$$

where $\beta_{L+1} \triangleq 0$, thus giving insights on how the proposed prior constrains the lag coefficients: it regularizes not only the ℓ_2-norm of the coefficients but also their successive differences, thereby enhancing local smoothness. The equivalence between the Bayesian inference and the fused ridge regularization (9) is proved in the online Supporting Information Appendix A.

There are many ways to specify λ_l and λ_l^* to meet the monotonicity constraints. In this article, we consider $\lambda_l = \exp\{\gamma_1(l+1)\} - 1$ and $\lambda_l^* = \exp\{\gamma_2(l+1)\} - 1$ for $\gamma_1, \gamma_2 > 0$, so that γ_1 controls the rate at which the ridge penalty in (9) increases, and γ_2 controls the increasing rate of smoothness of the coefficient curve $\boldsymbol{\beta}$. Instead of treating these hyperparameters as fixed, we postulate a standard exponential hyperprior on (γ_1, γ_2), that is, having probability density function

$$\pi(\gamma_1, \gamma_2) \propto \exp(-\gamma_1 - \gamma_2). \tag{10}$$

As such, the degree of ridge and smooth penalization can be determined according to the posterior distribution of the pair.

2.3 | Prior distribution on the error model

We put the Jeffreys prior for the error variance σ^2, that is, having density function

$$\pi(\sigma^2) \propto 1/\sigma^2. \tag{11}$$

Note that any inverse-gamma prior for σ^2 would maintain conjugacy, and the Jeffreys prior can be regarded as an improper limit of inverse-gamma prior distribution.

For the autoregressive process, we consider a truncated normal prior for $\boldsymbol{\phi}$ subject to the constraint that the error process is stationary. Specifically, we postulate

$$\boldsymbol{\phi} \sim N_p\left(0_p, \sigma_{\boldsymbol{\phi}}^2 \mathbf{I}_p\right) \mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi}), \tag{12}$$

where $S_{\boldsymbol{\phi}}(\boldsymbol{\phi})$ denotes the support where all roots of the polynomial $\Phi(B) = 1 - \sum_{l=1}^{p} \phi_l B^l$ are outside the unit circle. The process $\{\epsilon_t : t = 1, 2, \dots\}$ is stationary when $\boldsymbol{\phi} \in S_{\boldsymbol{\phi}}(\boldsymbol{\phi})$.[14,15] Note that the range of each ϕ_l is $(-1, 1)$ and Figure A1 in the online Supporting Information shows the probability density function of truncated normal prior with different values of $\sigma_{\boldsymbol{\phi}}^2 = 50$, 200, and 400, as well as a uniform prior on $(-1, 1)$. A prior variance of $\sigma_{\boldsymbol{\phi}}^2 = 200$ is used in (12), which essentially amounts to a flat prior.[15]

3 | CONDITIONAL POSTERIOR DISTRIBUTIONS

The proposed Bayesian model includes several conditionally conjugate priors, which facilitate posterior computations via a hybrid Metropolis-Hastings/Gibbs algorithm. We describe the conditional posterior distributions in this section; the details of derivation can be found in the online Supporting Information Appendix B.

Working with the likelihood (6) based on the transformed data Y_t^*, we obtain that $\tilde{\boldsymbol{\beta}}$ is conditionally normally distributed *a posteriori*:

$$\tilde{\boldsymbol{\beta}} | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \sigma^2, \boldsymbol{\phi}, \boldsymbol{\gamma} \sim N_{L+1}\left\{[\tilde{\boldsymbol{X}}^{*'}\tilde{\boldsymbol{X}}^* + \tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})]^{-1}\tilde{\boldsymbol{X}}^{*'}\boldsymbol{Y}^*, \sigma^2[\tilde{\boldsymbol{X}}^{*'}\tilde{\boldsymbol{X}}^* + \tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})]^{-1}\right\} \tag{13}$$

and that σ^2 has an inverse-gamma conditional posterior:

$$\sigma^2 | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\phi}, \boldsymbol{\gamma} \sim \mathrm{IG}\left[\frac{n - p + L + q + 2}{2}, \frac{(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y}^* - \tilde{\boldsymbol{X}}^*\tilde{\boldsymbol{\beta}}) + \tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\tilde{\boldsymbol{\beta}}}{2}\right]. \tag{14}$$

Note that the dependence of (13) and (14) on $\boldsymbol{\phi}$ is via the transformed data $\boldsymbol{Y}^*$.

Working with model (2) and (12), we obtain the conditional posterior distribution of $\boldsymbol{\phi}$ is truncated multivariate normal:

$$\boldsymbol{\phi} | \boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \sigma^2, \boldsymbol{\gamma} \sim N_p\left[\left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^* + \sigma_{\boldsymbol{\phi}}^{-2}\mathbf{I}\right)^{-1}\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{\epsilon}^*, \left(\sigma^{-2}\boldsymbol{E}^{*'}\boldsymbol{E}^* + \sigma_{\boldsymbol{\phi}}^{-2}\mathbf{I}\right)^{-1}\right] \mathbf{1}_{S_{\boldsymbol{\phi}}}(\boldsymbol{\phi}), \tag{15}$$

where $\boldsymbol{\epsilon}^* = (\epsilon^*_{p+1}, \ldots, \epsilon^*_n)'$, $\epsilon^*_t = Y_t - \mu - Z_t'\boldsymbol{b} - \sum_{l=0}^{L}\beta_l X_{t-l}$, and $\boldsymbol{E}^*$ is a $(n-p)\times p$ matrix with ϵ^*_{p+k-j} being the (k,j)-th element. Because of conjugacy, the parameters $\tilde{\boldsymbol{\beta}}, \sigma^2$, and $\boldsymbol{\phi}$ can be easily updated in a Gibbs sampling fashion.

Using the likelihood (6) and prior of $\boldsymbol{\gamma}$ and $\tilde{\boldsymbol{\beta}}$, the conditional posterior distribution can be expressed as

$$\pi(\boldsymbol{\gamma}|\boldsymbol{Y}, \tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\phi}, \sigma^2) \propto |\sigma^{-2}\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})|^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\tilde{\boldsymbol{\beta}}\right] \exp(-\gamma_1 - \gamma_2). \tag{16}$$

We propose to sample $\boldsymbol{\gamma}$ using a Metropolis-Hastings (MH) step with a uniform $U(-a, a)$ proposal distribution, that is, having an updating step $\gamma_{i,new} = \gamma_i + U(-a, a)$, where the tuning parameter a is chosen such that the acceptance rate of proposed sample is around 50%.[16] Note that updating the hyperparameter $\boldsymbol{\gamma}$ involves the calculation of the matrix $\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})$, which needs to be positive definite. The $(L+q+2)\times(L+q+2)$ matrix $\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})$ is a special case of tridiagonal matrix and it can be shown that $\tilde{\boldsymbol{\Omega}}(\boldsymbol{\gamma})$ is positive definite (see online Supporting Information Appendix C). The complete algorithm is summarized in the online Supporting Information Appendix D.

4 | SIMULATION STUDY

4.1 | Comparison methods

In this section, we evaluate the performance of the proposed BDLM-AR using simulation studies. At the end of each simulated trial, we fitted BDLM-AR with lag $L = 7$ and AR(1), that is, having

$$Y_t = \mu + \sum_{l=0}^{7}\beta_l X_{t-l} + \epsilon_t, \tag{17}$$

where $\epsilon_t = \phi\epsilon_{t-1} + w_t$ and $w_t \sim N(0, \sigma^2)$. Posterior distributions were derived using the hybrid Metropolis Hastings/Gibbs algorithm described in the previous section with 50 000 iterations, a burn-in period of 25 000, and $a = 0.2$ for sampling γ in the MH step.

We compared BDLM-AR with some existing methods including the Bayesian distributed lag model (BDLagM), which incorporates prior knowledge about the shape of the DL function through a normal prior with a specified covariance matrix,[10] Bayesian ridge DLM (BR-DLM) with a mean zero normal prior for $\tilde{\boldsymbol{\beta}}$, and a noninformative prior Bayesian DLM (NB-DLM) with a flat improper priors on each parameter in $\tilde{\boldsymbol{\beta}}$. These existing methods would use the same mean model (17) but assume independent errors without accounting for autocorrelation.

In addition, as a benchmark, we include the parametric Koyck's DLM[7] which assumes the knowledge of the true autoregressive coefficients is known. Details of the model specifications of the competing methods are given in the online Supporting Information Appendix E. The difference of four Bayesian distributed lag models can be found in Table A1 in the online Supporting Information. For Bayesian models, we estimate the parameters using the posterior means and for Koyck model, we use the maximum likelihood estimates.

4.2 | Simulation scenarios and data generation

In each simulated N-of-1 trial, measurements were collected daily for 120 days, under one of two treatment sequences. In the first sequence, a participant would receive $x_t = 1$ on the first 30 days and the last 30 days, and receive $x_t = 0$ between days 31 and 90; that is,

$$x_t^{(1)} = \begin{cases} 1 & t = 30s+1, \ldots, 30s+30 \text{ for } s = 0, 3, \\ 0 & t = 30s+1, \ldots, 30s+30 \text{ for } s = 1, 2. \end{cases}$$

In the second treatment sequence, a participant would switch treatments more frequently; specifically,

$$x_t^{(2)} = \begin{cases} 1 & t = 15s+1, \ldots, 15s+15 \text{ for } s = 0, 3, 5, 6, \\ 0 & t = 15s+1, \ldots, 15s+15 \text{ for } s = 1, 2, 4, 7. \end{cases}$$

For each treatment sequence, the data were generated according to model (17) under five sets of lag coefficients (lag curves, LC):

LC1. Exponential decay curve: $\boldsymbol{\beta} = (5, 2.5, 1.25, 0.625, 0.3125, 0, 0, 0)'$;
LC2. Exponential decay curve with oscillation: $\boldsymbol{\beta} = (5, 2.5, -1.25, -0.625, 0.3125, 0, 0, 0)'$;
LC3. Slow absorption curve: $\boldsymbol{\beta} = (1.51, 2.75, 3.36, 2.03, 0.34, 0, 0, 0)'$;
LC4. Slow absorption curve with oscillation: $\boldsymbol{\beta} = (1.51, 2.75, -3.36, -2.03, 0.34, 0, 0, 0)'$;
LC5. No carryover effect: $\boldsymbol{\beta} = (10, 0, 0, 0, 0, 0, 0, 0)'$.

The exponential decay curves (LC1 and LC2) specify coefficients that diminish in magnitude as lag lengthens. Specifically, the coefficients under LC1 decrease geometrically, which is aligned with the assumption of Koyck DLM. The slow absorption curves (LC3 and LC4) reflect scenarios where the carryover effect peaks at day 2 after treatment. LC5 is the null scenario where there is no carryover effect. The magnitudes of the coefficients were chosen in these scenarios so that $\|\boldsymbol{\beta}\|_1 \approx 10$; in addition, the total carryover effects (δ) are 4.69, 0.94, 8.48, −2.30, and 0, respectively for LC1–LC5. We consider $\sigma = 10, 20$ and $\phi = 0.5, 0.2$ for the stochastic component in data generation. In addition to the two main treatment sequences $x_t^{(1)}$ and $x_t^{(2)}$, other treatment switching frequency sequences are investigated to understand the effect of treatment design. $x_t^{(1)}$ and $x_t^{(2)}$ can be regarded as repeated treatment for a consecutive 30 and 15 days, respectively. Additional treatment sequences with repeated treatment for a consecutive 1, 2, 6, 7, and 10 days are considered. For time-varying covariates, we considered two scenarios: (1) linear time trend, that is, $Z_t = t$ for $t = 1, \ldots, n$, and (2) effect of weekend, that is, setting $Z_1 = Z_2 = \ldots = Z_5 = 0$ and $Z_6 = Z_7 = 1$ and so on, where Z_t is a weekend indicator. Time effect b is set to be 0.3 and 3 respectively for linear time trend and effect of weekend. For each of these scenarios, the methods were evaluated using 100 simulated trials.

4.3 | Simulation results

The convergence of all MCMC simulations are checked by the Gelman-Rubin diagnostics, which use stable and consistent estimators of Monte Carlo variance.[17,18] To be specific, the multivariate potential scale reduction factors (PSRF) are estimated to check the convergence of multiple parameters simultaneously. The point estimates of the multivariate PSRF range from 1.000021 to 1.000073 (mean 1.000047, median 1.000048), indicating the convergence of posterior samples. Figure 2 shows the bias and root mean squared error (RMSE) of the posterior means of individual lag coefficients. As expected, the biases of NB-DLM are relatively small; however, the method also has the largest RMSE uniformly because of the use of noninformative prior. The biases of the other methods are comparable. While the ℓ_2 penalty in BR-DLM on the lag coefficients reduces variability when compared to NB-DLM, the additional constraints on diminishing coefficients imposed by BDLagM and the proposed BDLM-AR further reduce RMSE for large lag l. Additionally, since the proposed BDLM-AR explicitly incorporates ridge-type regularization on the lag coefficients, it results in smaller RMSE for β_0 and the earlier lag coefficients (eg, β_1). However, as a trade-off, the bias of BDLM-AR for early lag coefficients will be slightly inflated as compared to BDLagM, BR-DLM and NB-DLM, especially when true lag coefficient curve has frequent fluctuation. The benchmark method, Koyck DLM, performs best in the exponential decay case, where the true coefficient of autoregressive error (ϕ) is assumed to be known. The proposed BDLM-AR has very similar performance as Koyck DLM. Note that the coefficient of autoregressive error is estimated directly from the proposed BDLM-AR model, which is more practical in real application. In summary, the proposed BDLM-AR generally results in the smallest RMSE for all lag coefficients.

To further examine the performance of each method in estimating the lag curve in aggregate, Figure 3 plots the Euclidean distance between the vector of estimated lag coefficients and the vector of true lag coefficients. Under LC1 (exponential decay), the Koyck DLM has the best performance overall. This is not surprising because (i) the Koyck model mimics the coefficients under LC1 closely, and (ii) Koyck DLM assumes knowledge of the true autoregressive coefficients used in the simulation and hence it is not a method that can be implemented in practice. Thus, this comparison serves as a benchmark about the efficiency of the proposed BDLM-AR and other Bayesian methods. The figure demonstrates that the proposed BDLM-AR produces smaller distance from the true lag curve than the other Bayesian methods.

Table 1 gives the bias and RMSE in the estimation of the total effect ($\sum_{l=0}^{7}\beta_l$), the total carryover effect ($\delta = \sum_{l=i}^{7}\beta_l$) and the immediate effect (β_0) under different lag curves with $\sigma = 10$ and $\phi = 0.5$ under treatment sequence $x_t^{(1)}$. Results

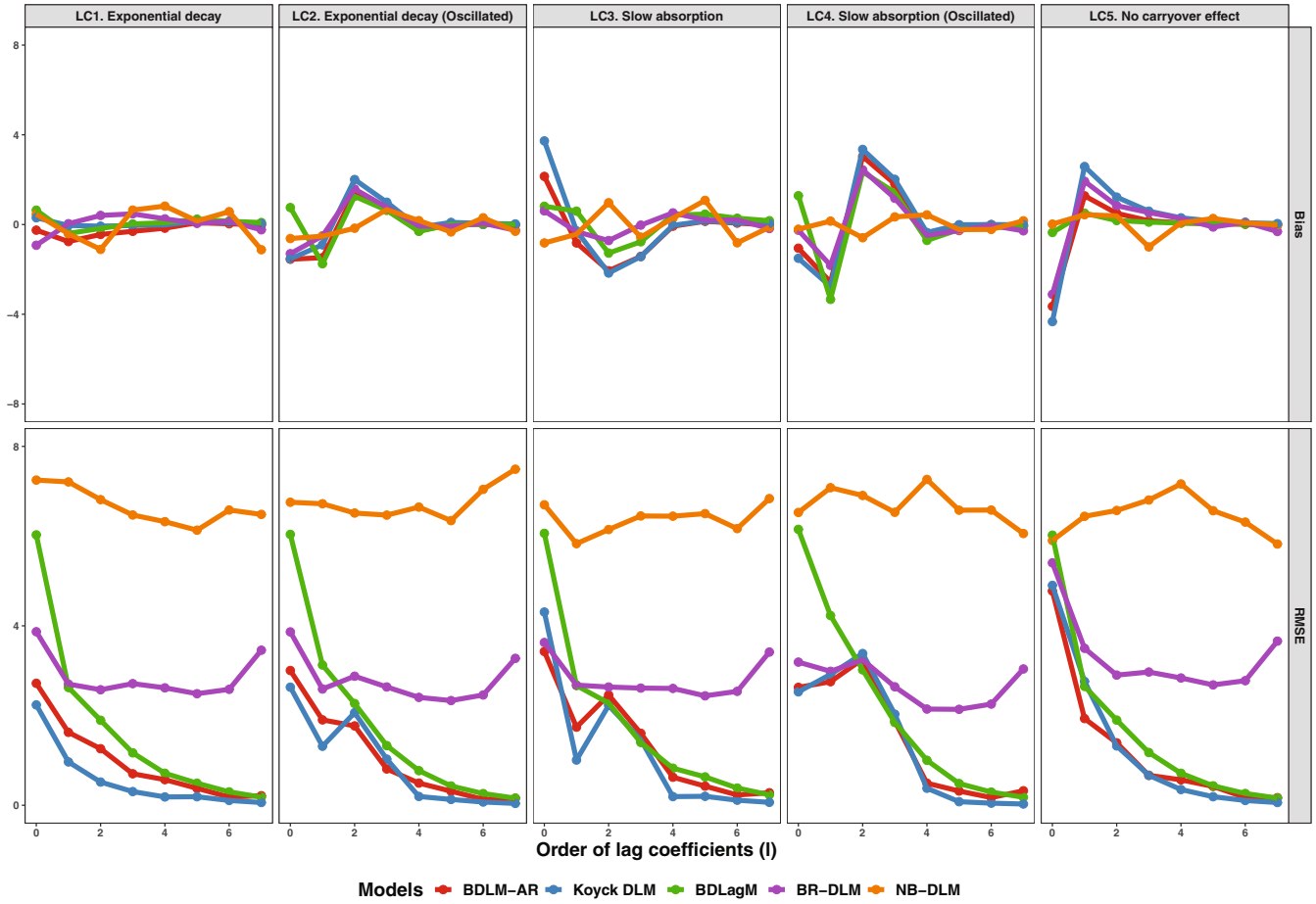

FIGURE 2 Bias and RMSE of lag coefficients estimates under five lag curves, treatment sequence $x_t^{(1)}$, $\sigma = 10$ and $\phi = 0.5$.

for other values of ϕ, σ and treatment sequence are similar and are provided in Figure A2 to A4 in the online Supporting Information. Overall, the proposed BDLM-AR yields consistently lower RMSE in estimating total effect, carryover effects, and immediate effects than other comparison methods, except for Koyck DLM. This is consistent with what we observe in Figures 2 and 3. And the bias of the proposed BDLM-AR is larger than the other models as a trade-off. We note that the advantages of BDLM-AR in terms of RMSE for the total carryover effect (δ) and immediate effect (β_0) are more pronounced than that for total effect ($\delta + \beta_0$). This is indeed the motivation that we set out to accomplish: to decompose the treatment effects and separate carryover effect from the immediate effect. The simulation results for time-varying covariates are summarized in Table 2. The proposed BDLM-AR with time varying covariates can yield unbiased estimates for time effect and has potentially better performance in estimating total effect, total carryover effect and immediate effect compared to model without time varying covariates.

4.4 | Effects of design

Figure 3 shows that the Euclidean distance between the vector of estimated lag coefficients and the truth under treatment sequence $x_t^{(1)}$ and $x_t^{(2)}$. The relative performance among methods is the same regardless of the treatment sequence, that is, the proposed BDLM-AR yields the shortest distance from the true lag coefficients $\boldsymbol{\beta}$. The simulation results for other treatment switching frequency sequences are summarized in Figure 4. As we observe, frequently switching treatments will generally help improve the performance of the proposed BDLM-AR. This is in line with what we expect because collinearity of exposure lags will be lessened as treatments change frequently, while the total duration is held fixed. However, when the true lag curve deviates from the monotone decreasing trend as LC3 and LC4, switching the treatment sequence in the most extreme case will yield worse performance. From a practical perspective, switching treatments frequently

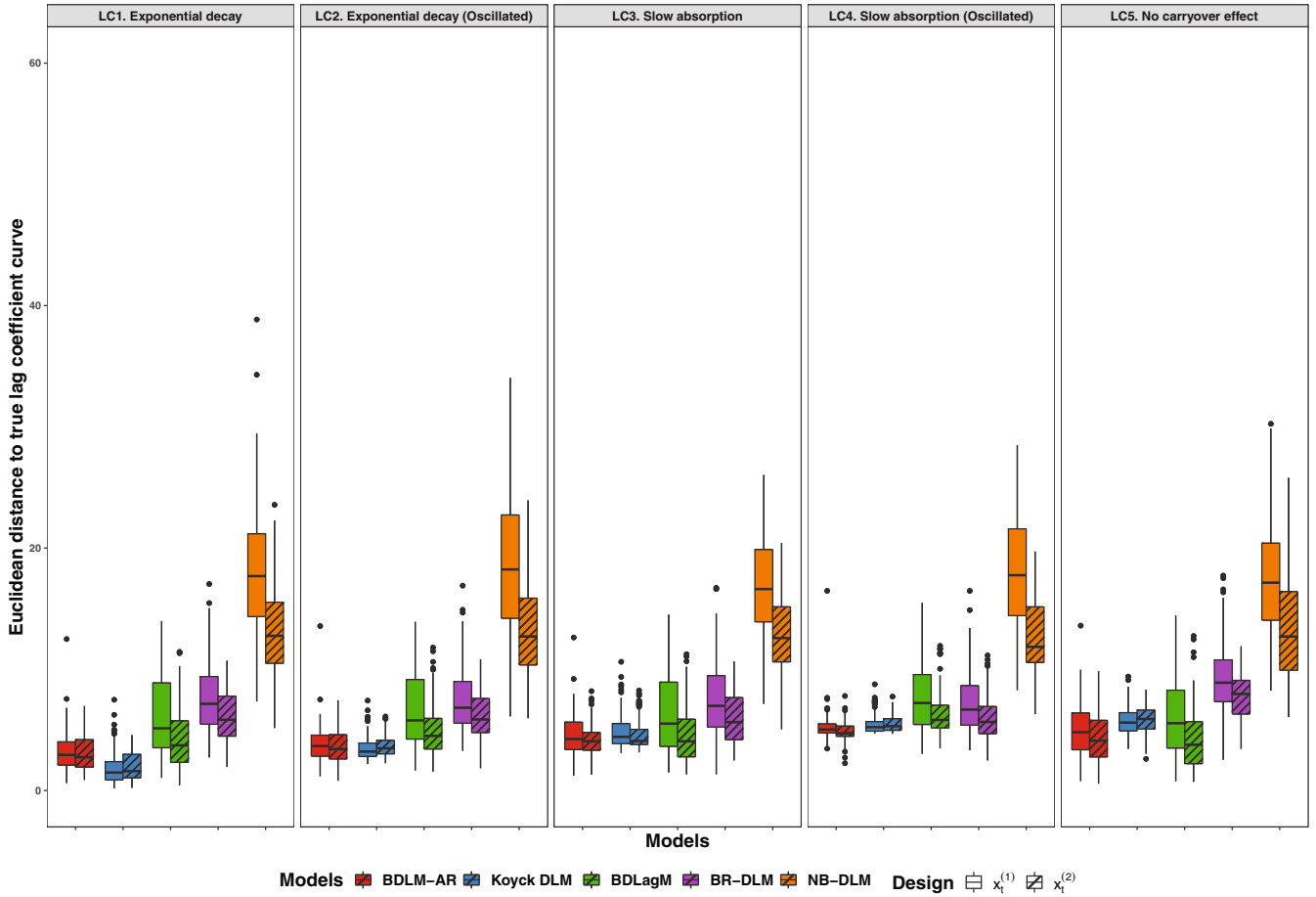

FIGURE 3 Euclidean distance to true lag curves under: Treatment sequence $x_t^{(1)}$ vs Treatment sequence $x_t^{(2)}$. Other simulation parameters are fixed as: $\sigma = 10$ and $\phi = 0.5$.

(eg, everyday) is infeasible as it will impose excessive burdens on patients and may cause treatment nonadherence. In practice, the number of crossovers in an N-of-1 trial is expected to be determined based on the trade-off between statistical accuracy and practical considerations.

4.5 | Effects of model misspecification

In the previous subsections, BDLM-AR and other methods use a working mean model with $L = 7$ and an AR(1) model for autoregressive errors. These working models correctly specify (or include) the data generation model in the previous simulation study. In this subsection, we investigate the robustness of BDLM-AR under model mis-specifications. Specifically, we will consider (1) the working mean model with $L = 0, 1, \ldots, 6, 7, 15$; (2) the stochastic components that assume autoregressive error order of $p = 0, 1, 7$. That is, we consider a total of 27 BDLM-AR models.

In data generation, we use LC1 as the true mean model, where $\beta_l > 0$ for $l = 0, 1, 2, 3, 4$, and we consider true scenarios for the errors:

E1. AR(1) with $\phi = 0.5$;
E2. Autoregressive model with $\phi_1 = 0.5, \phi_2 = 0, \phi_3 = 0, \phi_4 = 0.3, \phi_5 = 0, \phi_6 = 0.2$.

Note that, under the scenario LC1/E1, a working model with $L < 4$ or $p = 0$ under-specifies the true model. Likewise, under LC1/E2, a working model with $L < 4$ or $p = 0, 1$ under-specifies the true model.

Table 3 summarizes the RMSE of these 27 models under the two scenarios (LC1/E1 and LC1/E2) with $\sigma = 10$ under $x_t^{(1)}$. It can be seen that misspecified lag length has little influence on estimating total effect, total carryover

TABLE 1 Summary of evaluation metrics (best values in bold) of total effect, total carryover effect and immediate effect (β_0) under five lag curves, treatment sequence $x_t^{(1)}$, $\sigma = 10$ and $\phi = 0.5$. Monte Carlo standard errors are in brackets.

		Truth	BDLM-AR	Koyck DLM	BDLagM	BR-DLM	NB-DLM
Bias	**Total effect**						
	LC1. Exponential decay	10	−1.82 (0.31)	0.41 (0.39)	0.65 (0.40)	0.21 (0.42)	**0.02 (0.40)**
	LC2. Exponential decay (Oscillated)	5.94	−1.24 (0.27)	0.61 (0.39)	0.61 (0.40)	**0.20 (0.41)**	−0.79 (0.39)
	LC3. Slow absorption	10	−2.04 (0.32)	**0.12 (0.39)**	0.71 (0.40)	0.30 (0.42)	−0.42 (0.44)
	LC4. Slow absorption (Oscillated)	−0.79	0.73 (0.22)	0.73 (0.39)	0.59 (0.40)	0.46 (0.40)	**−0.15 (0.40)**
	LC5. No carryover	10	−1.60 (0.32)	0.65 (0.39)	0.58 (0.40)	**0.14 (0.42)**	0.25 (0.39)
	Total carryover effect						
	LC1. Exponential decay	4.69	−1.57 (0.26)	0.10 (0.20)	**0.01 (0.66)**	1.13 (0.47)	−0.44 (0.76)
	LC2. Exponential decay (Oscillated)	0.94	0.31 (0.22)	2.14 (0.19)	**−0.15 (0.66)**	1.50 (0.46)	−0.17 (0.63)
	LC3. Slow absorption	8.48	−4.19 (0.30)	−3.61 (0.21)	**−0.11 (0.66)**	−0.31 (0.47)	0.40 (0.70)
	LC4. Slow absorption (Oscillated)	−2.30	1.78 (0.24)	2.23 (0.19)	−0.70 (0.66)	0.71 (0.44)	**0.06 (0.64)**
	LC5. No carryover	0	2.05 (0.27)	4.98 (0.20)	0.94 (0.66)	3.25 (0.51)	**0.23 (0.66)**
	Immediate effect						
	LC1. Exponential decay	5	**−0.25 (0.27)**	0.30 (0.22)	0.64 (0.60)	−0.92 (0.38)	0.46 (0.72)
	LC2. Exponential decay (Oscillated)	5	−1.55 (0.26)	−1.53 (0.21)	0.76 (0.60)	−1.30 (0.36)	**−0.62 (0.67)**
	LC3. Slow absorption	1.51	2.15 (0.27)	3.73 (0.22)	0.81 (0.60)	**0.61 (0.36)**	−0.82 (0.66)
	LC4. Slow absorption (Oscillated)	1.51	−1.05 (0.24)	−1.51 (0.20)	1.29 (0.60)	−0.25 (0.32)	**−0.21 (0.65)**
	LC5. No carryover	10	−3.65 (0.31)	−4.33 (0.23)	−0.35 (0.60)	−3.11 (0.44)	**0.02 (0.59)**
RMSE	**Total effect**						
	LC1. Exponential decay	10	**3.61 (0.21)**	3.87 (0.24)	4.05 (0.25)	4.17 (0.26)	3.96 (0.24)
	LC2. Exponential decay (Oscillated)	5.94	**2.95 (0.17)**	3.90 (0.25)	4.05 (0.25)	4.11 (0.26)	3.93 (0.25)
	LC3. Slow absorption	10	**3.81 (0.22)**	3.85 (0.24)	4.06 (0.26)	4.16 (0.26)	4.40 (0.27)
	LC4. Slow absorption (Oscillated)	−0.79	**2.28 (0.14)**	3.93 (0.25)	4.04 (0.25)	4.01 (0.26)	3.99 (0.24)
	LC5. No carryover	10	**3.54 (0.21)**	3.91 (0.25)	4.04 (0.25)	4.18 (0.26)	3.93 (0.24)
	Total Carryover Effect						
	LC1. Exponential decay	4.69	3.03 (0.19)	**2.01 (0.13)**	6.57 (0.36)	4.83 (0.28)	7.61 (0.44)
	LC2. Exponential decay (Oscillated)	0.94	**2.25 (0.19)**	2.87 (0.17)	6.56 (0.36)	4.83 (0.28)	6.31 (0.42)
	LC3. Slow absorption	8.48	5.13 (0.25)	**4.15 (0.19)**	6.58 (0.36)	4.64 (0.27)	6.94 (0.38)
	LC4. Slow absorption (Oscillated)	−2.30	2.96 (0.19)	**2.94 (0.16)**	6.62 (0.38)	4.44 (0.27)	6.39 (0.39)
	LC5. No carryover	0	**3.39 (0.23)**	5.36 (0.20)	6.65 (0.36)	6.06 (0.35)	6.56 (0.39)
	Immediate effect						
	LC1. Exponential decay	5	2.71 (0.16)	**2.23 (0.14)**	6.00 (0.36)	3.85 (0.21)	7.21 (0.44)
	LC2. Exponential decay (Oscillated)	5	2.99 (0.17)	**2.62 (0.16)**	6.01 (0.36)	3.85 (0.20)	6.72 (0.43)
	LC3. Slow absorption	1.51	**3.42 (0.22)**	4.30 (0.21)	6.03 (0.36)	3.61 (0.24)	6.67 (0.38)
	LC4. Slow absorption (Oscillated)	1.51	2.62 (0.18)	**2.52 (0.15)**	6.12 (0.37)	3.18 (0.21)	6.49 (0.39)
	LC5. No carryover	10	**4.77 (0.26)**	4.90 (0.22)	5.99 (0.35)	5.39 (0.29)	5.87 (0.37)

TABLE 2 Summary of evaluation metrics of total effect, total carryover effect, DL coefficients and autoregressive coefficient under scenarios when different time trends exist: (1) Linear time trend and (2) Effect of weekend. Other simulation parameters are fixed as: treatment sequence $= x_t^{(1)}$, $\sigma = 10$ and $\phi = 0.5$. Monte Carlo standard errors are in brackets.

		Linear time trend			
		Bias		**RMSE**	
	Truth	**Exclude time varying covariate**	**Include time varying covariate**	**Exclude time varying covariate**	**Include time varying covariate**
b_1	0.3	-	0.01 (0.005)	-	0.05 (0.004)
Total effect	9.69	−3.81 (0.34)	−1.85 (0.38)	5.06 (0.25)	4.20 (0.25)
Total carryover effect	4.69	−3.04 (0.19)	−1.61 (0.27)	3.56 (0.16)	3.15 (0.18)
Immediate effect (β_0)	5	−0.77 (0.31)	−0.24 (0.28)	3.18 (0.20)	2.83 (0.18)
β_1	2.5	−1.47 (0.15)	−0.79 (0.15)	2.10 (0.12)	1.69 (0.11)
β_2	1.25	−0.85 (0.08)	−0.48 (0.10)	1.14 (0.06)	1.10 (0.07)
β_3	0.62	−0.47 (0.05)	−0.25 (0.07)	0.69 (0.04)	0.71 (0.05)
β_4	0.31	−0.23 (0.03)	−0.14 (0.04)	0.37 (0.02)	0.43 (0.03)
β_5	0	0.00 (0.02)	0.05 (0.04)	0.20 (0.02)	0.36 (0.03)
β_6	0	0.00 (0.01)	0.01 (0.03)	0.14 (0.01)	0.25 (0.02)
β_7	0	−0.01 (0.01)	−0.01 (0.02)	0.12 (0.01)	0.22 (0.02)
		Effect of weekend			
		Bias		**RMSE**	
	Truth	**Exclude time varying covariate**	**Include time varying covariate**	**Exclude time varying covariate**	**Include time varying covariate**
b_1	3	-	0.09 (0.21)	-	4.56 (0.13)
Total effect	9.69	−1.88 (0.38)	−1.79 (0.38)	4.19 (0.25)	4.17 (0.25)
Total carryover effect	4.69	−1.83 (0.27)	−1.60 (0.28)	3.25 (0.18)	3.18 (0.18)
Immediate effect (β_0)	5	−0.05 (0.28)	−0.20 (0.28)	2.81 (0.18)	2.80 (0.17)
β_1	2.5	−1.01 (0.15)	−0.78 (0.15)	1.78 (0.11)	1.68 (0.10)
β_2	1.25	−0.54 (0.10)	−0.47 (0.10)	1.10 (0.07)	1.11 (0.07)
β_3	0.62	−0.31 (0.06)	−0.25 (0.07)	0.70 (0.05)	0.72 (0.05)
β_4	0.31	−0.11 (0.04)	−0.14 (0.04)	0.41 (0.03)	0.42 (0.03)
β_5	0	0.12 (0.04)	0.04 (0.03)	0.38 (0.03)	0.35 (0.03)
β_6	0	0.04 (0.02)	0.00 (0.03)	0.25 (0.02)	0.26 (0.02)
β_7	0	−0.01 (0.02)	−0.01 (0.02)	0.21 (0.02)	0.22 (0.02)

effect and immediate effect, while under-specified error AR order will increase RMSE of parameters to a higher level than over-specified error AR order. Note that when choosing a small lag length value, we can hardly acquire estimation about the whole DL curve, as well as the information on the duration of carryover effect. Therefore, when the lag length is unknown, we suggest to fit data with a reasonably long lag length. For example, one can use pharmacokinetic parameters such as the half-life of some drugs. For error autoregressive order, when the true orders are unknown, it is also suggested to fit a model with high autoregressive order.

Finally, sensitivity analyses are implemented to understand the model's performance under other error assumptions. ARMA(1,1) and ARMA(1,2) errors are considered as true scenarios. The results are consistent with the previous findings. Detailed comparison results on estimating total effect, total carryover effect and immediate effect can be found in Table A2 in the online Supporting Information.

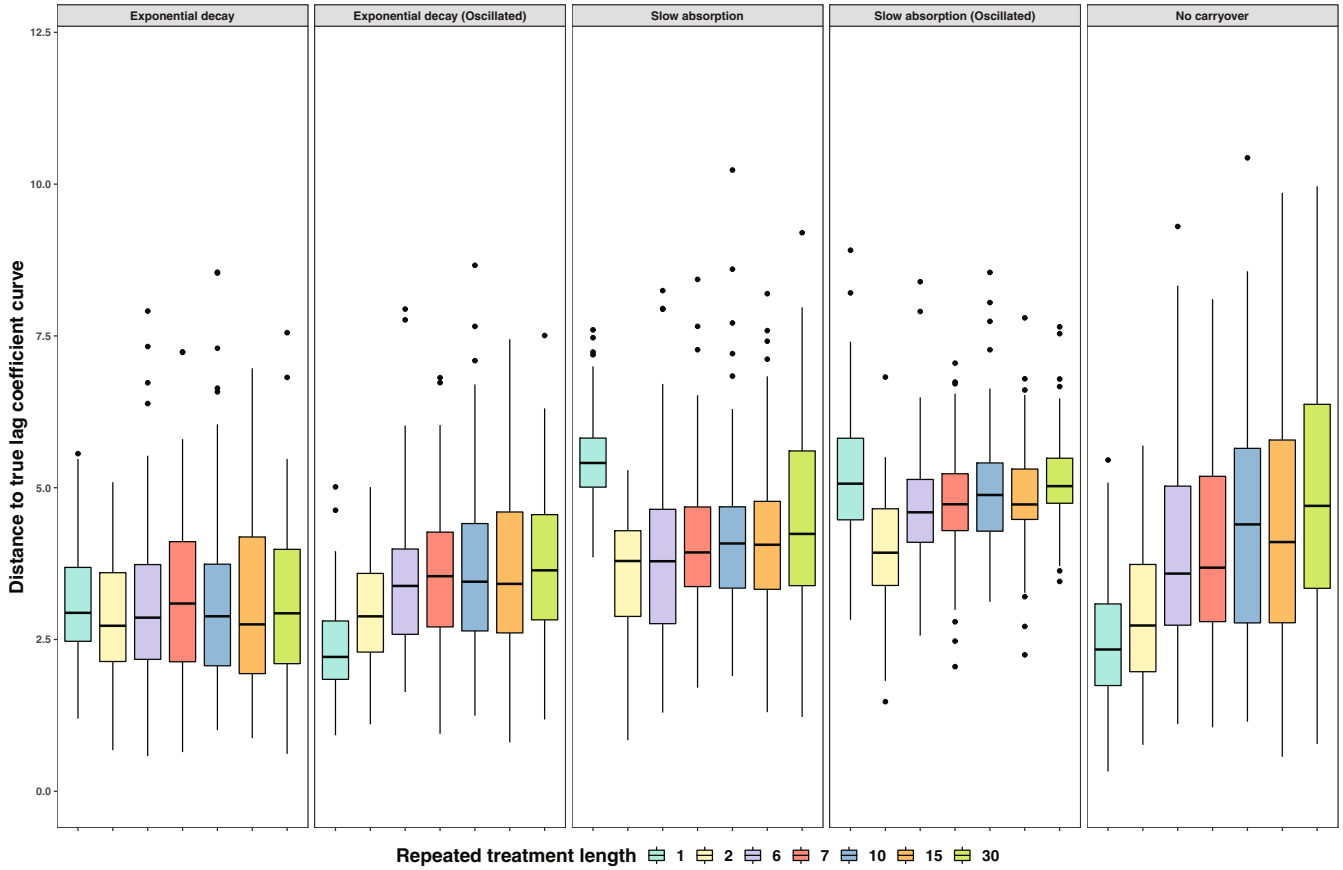

FIGURE 4 Euclidean distance to true lag curves under different repeated treatment lengths.

5 | APPLICATION TO LIGHT THERAPY STUDY

The data set we used is from the light therapy study,[4] which studies the effectiveness of bright white light therapy for depressive symptoms within cancer survivors. Besides bright white intervention (10 000 lux), dim red (50 lux) was used as a control intervention, which lacks sufficient light intensity to affect cells from retina. Patients were instructed to use one of two portable lightboxes each morning for 30 minutes per day. For each patient, the whole study duration was 12 weeks. One intervention was assigned on the first three weeks and last three weeks and the other intervention was assigned between the fourth week and the ninth week. The initial intervention was randomized, either bright white lightbox or dim red lightbox. The collected outcomes were depressive symptom and fatigue symptom, which were tracked using a smartphone application. The outcomes were measured by patient's self-reported standard single-item visual analog scale from 0-not at all depressed/tired to 10-extremely depressed/tired. Some occasional missing outcomes were imputed using Bayesian multiple imputation suggested by Gelman et al[19] To be specific, 20 imputed data sets were generated by drawing missing values from the posterior predictive distribution. Each of the 20 imputed data sets is analyzed separately, and the results are combined together by mixing together the posterior samples from the separate inferences.

We fit the data with the proposed BDLM-AR model with $L = 7$. Two autoregressive orders of BDLM-AR model AR(1) and AR(7) were used. Weekend effect was also added to the model as a time varying covariate. Convergence of all the MCMC were checked using the same method mentioned in the previous section. In addition to the comparison of Bayesian distributed lag models, we also fit the frequentist autoregressive regression models (RegAR) with $p = 1$ and 7. RegAR model with $p = 1$ was originally used as the analysis method in the light therapy study.

Table 4 shows the posterior means of the coefficients by the proposed BDLM-AR model and the maximum likelihood estimates by RegAR using depressive symptom outcome. For patient 7706, the RegAR indicates a weak insignificant total effect of bright white intervention in relieving depressive symptom. However, BDLM-AR(7) model with weekend effect identifies a significant strong effect of bright white intervention as −0.37 (90% CI: −1.04, 0.12). To check the fitness of each

TABLE 3 Summary of RMSE of total effect, total carryover effect and immediate effect (β_0) fitted using BDLM-AR model with different lag length and error autoregressive order. LC1. Exponential decay curve, treatment sequence $x_t^{(1)}$, $\sigma = 10$, E1. AR(1) with $\phi = 0.5$ and E2. Autoregressive model with $\phi_1 = 0.5, \phi_2 = 0, \phi_3 = 0, \phi_4 = 0.3, \phi_5 = 0, \phi_6 = 0.2$ are used to generate simulated data. Monte Carlo standard errors are in brackets.

		Truth for errors: E1			Truth for errors: E2		
	Lag	AR(7)	AR(1)	AR(0)	AR(7)	AR(1)	AR(0)
Total effect	15	4.13 (0.27)	3.69 (0.23)	4.12 (0.25)	6.13 (0.35)	9.01 (0.62)	11.39 (0.76)
	7	4.32 (0.26)	3.85 (0.24)	3.95 (0.25)	5.85 (0.31)	8.32 (0.53)	10.72 (0.72)
	6	4.37 (0.27)	3.88 (0.24)	3.92 (0.25)	5.92 (0.32)	8.25 (0.52)	10.62 (0.71)
	5	4.44 (0.26)	3.91 (0.24)	3.93 (0.24)	5.89 (0.32)	8.21 (0.50)	10.55 (0.70)
	4	4.49 (0.27)	3.98 (0.24)	3.94 (0.24)	6.01 (0.31)	8.06 (0.47)	10.48 (0.69)
	3	4.59 (0.27)	4.02 (0.24)	3.93 (0.24)	6.02 (0.31)	7.81 (0.45)	10.40 (0.68)
	2	4.63 (0.26)	4.07 (0.24)	3.91 (0.24)	6.08 (0.29)	7.68 (0.41)	10.34 (0.67)
	1	4.75 (0.26)	4.15 (0.24)	3.87 (0.24)	6.12 (0.29)	7.43 (0.38)	10.26 (0.66)
	0	4.01 (0.25)	3.71 (0.23)	3.77 (0.23)	5.82 (0.32)	8.58 (0.51)	10.52 (0.67)
Total carryover effect	15	3.97 (0.26)	3.63 (0.24)	4.48 (0.25)	4.12 (0.24)	6.77 (0.52)	8.25 (0.56)
	7	3.56 (0.22)	3.18 (0.20)	3.85 (0.23)	3.55 (0.19)	5.24 (0.35)	6.69 (0.45)
	6	3.51 (0.22)	3.17 (0.19)	3.73 (0.22)	3.60 (0.18)	5.15 (0.34)	6.38 (0.43)
	5	3.42 (0.20)	3.15 (0.19)	3.61 (0.21)	3.60 (0.19)	4.95 (0.31)	6.08 (0.40)
	4	3.41 (0.20)	3.12 (0.18)	3.49 (0.21)	3.73 (0.19)	4.67 (0.28)	5.68 (0.38)
	3	3.41 (0.20)	3.12 (0.19)	3.36 (0.20)	3.70 (0.18)	4.29 (0.23)	5.33 (0.35)
	2	3.38 (0.18)	3.10 (0.18)	3.15 (0.18)	3.73 (0.16)	4.00 (0.21)	4.62 (0.29)
	1	3.58 (0.17)	3.33 (0.17)	2.92 (0.16)	3.86 (0.13)	3.88 (0.19)	3.54 (0.20)
	0	-	-	-	-	-	-
Immediate effect	15	3.53 (0.21)	3.40 (0.21)	4.00 (0.24)	3.96 (0.23)	4.91 (0.28)	8.91 (0.58)
	7	3.09 (0.18)	2.91 (0.18)	3.54 (0.22)	3.61 (0.21)	4.54 (0.27)	8.52 (0.56)
	6	3.02 (0.17)	2.91 (0.18)	3.47 (0.22)	3.56 (0.20)	4.50 (0.27)	8.46 (0.56)
	5	3.03 (0.17)	2.87 (0.18)	3.38 (0.21)	3.58 (0.19)	4.52 (0.27)	8.40 (0.56)
	4	3.04 (0.17)	2.85 (0.18)	3.31 (0.21)	3.57 (0.19)	4.54 (0.27)	8.29 (0.56)
	3	3.03 (0.17)	2.88 (0.18)	3.28 (0.21)	3.54 (0.19)	4.66 (0.28)	8.25 (0.56)
	2	3.09 (0.17)	2.97 (0.19)	3.29 (0.22)	3.64 (0.19)	4.91 (0.30)	8.28 (0.57)
	1	3.15 (0.19)	3.15 (0.21)	3.49 (0.23)	3.71 (0.21)	5.21 (0.33)	8.48 (0.58)
	0	5.39 (0.34)	5.62 (0.34)	6.00 (0.35)	6.35 (0.43)	9.02 (0.60)	11.75 (0.79)

model, we used Ljung–Box test to examine autocorrelation of the residuals,[20] and the corresponding P-values of χ^2-test are also shown in Table 4. No statistically significant autocorrelation is found in residuals of BDLM-AR models. We also found a second peak of treatment effect within patient 7706 two days after the immediate effect. For patient 7708, we observe a similar estimation between different models in terms of total effect. Treatment total effect estimated from BDLM-AR(7) model with weekend effect is −1.41 (90% CI: −3.58, 0.27). Extra information obtained from BDLM-AR model is that the majority of treatment effect lasts for around 2 days. The results are similar when weekend effect was added to the model. For RegAR(1) model, statistically significant autocorrelation is found in residuals, indicating an inadequacy of model fitting. Analysis results using fatigue symptom outcome can be found in Table A3 in the online Supporting Information.

TABLE 4 Lag coefficient estimates for light therapy study using depressive symptom outcome. 90% credible intervals/confidence intervals are in brackets. *P*-value of Ljung-Box test for each model is on the last row.

	Subject ID: 7706					
	BDLM-AR (1)	**BDLM-AR (7)**	**BDLM-AR(1) with weekend effect**	**BDLM-AR(7) with weekend effect**	**RegAR(1)**	**RegAR (7)**
μ	2.49 (2.09,2.85)	1.95 (−0.27,2.88)	2.47 (2.04,2.87)	1.89 (−0.31,2.85)	2.58 (2.21,2.95)	3.14 (2.05,4.23)
b_1	-	-	0.05 (−0.32,0.43)	0.15 (−0.19,0.50)	-	-
Total effect	0.03 (−0.38,0.46)	−0.39 (−1.05,0.10)	0.04 (−0.36,0.47)	−0.37 (−1.04,0.12)	−0.02 (−0.52,0.49)	−0.36 (−0.93,0.21)
Total carryover effect	0.02 (−0.31,0.39)	−0.27 (−1.01,0.14)	0.02 (−0.31,0.38)	−0.27 (−1.04,0.11)	-	-
β_0	0.02 (−0.41,0.43)	−0.13 (−0.59,0.37)	0.03 (−0.39,0.45)	−0.10 (−0.54,0.37)	−0.02 (−0.52,0.49)	−0.36 (−0.93,0.21)
β_1	0 (−0.26,0.26)	−0.05 (−0.4,0.29)	0 (−0.25,0.26)	−0.04 (−0.38,0.29)	-	-
β_2	0.01 (−0.14,0.18)	−0.09 (−0.41,0.11)	0.01 (−0.15,0.17)	−0.09 (−0.42,0.09)	-	-
β_3	0.01 (−0.08,0.13)	−0.07 (−0.35,0.07)	0.01 (−0.08,0.12)	−0.07 (−0.37,0.05)	-	-
β_4	0.01 (−0.05,0.09)	−0.03 (−0.22,0.07)	0.01 (−0.05,0.09)	−0.03 (−0.24,0.06)	-	-
β_5	0 (−0.05,0.04)	−0.02 (−0.16,0.06)	0 (−0.05,0.03)	−0.02 (−0.16,0.06)	-	-
β_6	0 (−0.03,0.02)	−0.01 (−0.10,0.06)	0 (−0.03,0.02)	−0.01 (−0.10,0.06)	-	-
β_7	0 (−0.02,0.01)	−0.01 (−0.08,0.04)	0 (−0.02,0.01)	−0.01 (−0.08,0.04)	-	-
ϕ_1	0.44 (0.24,0.66)	0.09 (−0.17,0.38)	0.44 (0.23,0.66)	0.10 (−0.17,0.39)	0.52 (0.34,0.67)	0.31 (0.13,0.49)
ϕ_2	-	0.16 (−0.08,0.41)	-	0.18 (−0.08,0.43)	-	0.13 (−0.06,0.32)
ϕ_3	-	0.06 (−0.16,0.26)	-	0.03 (−0.23,0.25)	-	−0.05 (−0.25,0.15)
ϕ_4	-	0.20 (0.01,0.39)	-	0.18 (−0.03,0.37)	-	0.20 (0.01,0.40)
ϕ_5	-	0.10 (−0.12,0.31)	-	0.12 (−0.09,0.33)	-	0.09 (−0.14,0.32)
ϕ_6	-	0.07 (−0.13,0.26)	-	0.08 (−0.12,0.26)	-	0.07 (−0.15,0.29)
ϕ_7	-	0.05 (−0.14,0.24)	-	0.05 (−0.14,0.24)	-	0.06 (−0.17,0.28)
P-value	0.76	0.96	0.77	0.78	<0.001	0.97

(Continues)

TABLE 4 (Continued)

	Subject ID: 7708					
	BDLM-AR (1)	**BDLM-AR (7)**	**BDLM-AR(1) with weekend effect**	**BDLM-AR(7) with weekend effect**	**RegAR(1)**	**RegAR (7)**
μ	2.90 (1.67,4.12)	3.19 (0.31,6.88)	2.89 (1.51,4.22)	3.23 (0.16,7.25)	2.76 (1.91,3.62)	4.64(2.38,6.89)
b_1	-	-	0.02 (−1.16,1.18)	−0.01 (−1.36,1.34)	-	-
Total effect	−1.20 (−2.75,0.17)	−1.39 (−3.47,0.28)	−1.19 (−2.78,0.20)	−1.41 (−3.58,0.27)	−1.42 (−2.64,−0.19)	−1.72 (−3.14,−0.31)
Total carryover effect	−0.45 (−1.97,0.55)	−0.57 (−2.46,0.67)	−0.45 (−1.96,0.50)	−0.58 (−2.49,0.64)	-	-
β_0	−0.75 (−2.14,0.55)	−0.82 (−2.50,0.70)	−0.74 (−2.16,0.54)	−0.83 (−2.59,0.69)	−1.42 (−2.64,−0.19)	−1.72 (−3.14,−0.31)
β_1	−0.27 (−1.31,0.53)	−0.35 (−1.63,0.62)	−0.27 (−1.27,0.50)	−0.36 (−1.67,0.59)	-	-
β_2	−0.09 (−0.78,0.45)	−0.13 (−0.99,0.53)	−0.09 (−0.76,0.42)	−0.13 (−1.00,0.55)	-	-
β_3	−0.05 (−0.54,0.30)	−0.08 (−0.72,0.36)	−0.06 (−0.53,0.27)	−0.09 (−0.74,0.35)	-	-
β_4	−0.03 (−0.35,0.21)	−0.05 (−0.49,0.27)	−0.03 (−0.34,0.20)	−0.04 (−0.49,0.28)	-	-
β_5	−0.01 (−0.21,0.17)	0.01 (−0.24,0.30)	−0.01 (−0.21,0.16)	0.02 (−0.23,0.32)	-	-
β_6	0 (−0.12,0.13)	0.02 (−0.15,0.25)	0 (−0.12,0.12)	0.02 (−0.15,0.26)	-	-
β_7	0 (−0.09,0.08)	0.01 (−0.12,0.15)	0 (−0.08,0.08)	0.01 (−0.12,0.15)	-	-
ϕ_1	0.44 (0.26,0.62)	0.34 (0.10,0.57)	0.30 (0.08,0.52)	0.33 (0.09,0.57)	0.43 (0.24,0.59)	0.36 (0.17,0.54)
ϕ_2	-	−0.07 (−0.34,0.20)	-	−0.07 (−0.35,0.19)	-	−0.01 (−0.21,0.19)
ϕ_3	-	0.03 (−0.23,0.30)	-	0.03 (−0.23,0.30)	-	0 (−0.20,0.21)
ϕ_4	-	0.06 (−0.21,0.32)	-	0.06 (−0.21,0.33)	-	0.13 (−0.07,0.33)
ϕ_5	-	0.08 (−0.18,0.33)	-	0.07 (−0.18,0.32)	-	0.11 (−0.10,0.31)
ϕ_6	-	0.02 (−0.25,0.32)	-	0.03 (−0.25,0.31)	-	0.06 (−0.16,0.28)
ϕ_7	-	0.03 (−0.26,0.32)	-	0.04 (−0.25,0.33)	-	−0.09 (−0.29,0.12)
P-value	0.36	0.63	0.35	0.67	<0.001	0.91

6 | DISCUSSION

In this paper, we introduce a novel method to analyze data from N-of-1 trials. The method handles temporal correlation between measurements and carryover effects via distributed lag structure and parameters are estimated using Bayesian approach with (fused) ridge type regularization. From the design perspective, N-of-1 trial can be viewed as a multi-period crossover trial in an individual. Traditional crossover trial requires physical washout period to eliminate carryover effects, resulting in pauses of study intervention and potentially lengthening study duration. Instead of using physical washout period, our proposed method provides an alternative to address the carryover effects analytically, which can be applied to N-of-1 trial even without washout period. This is specifically suited to applications where outcomes are measured continuously over the study period. Our simulation studies show that the proposed BDLM-AR model generally outperforms Koyck DLM, BDLagM, BR-DLM and NB-DLM in estimating carryover effects while comparable in estimating the total effects. Furthermore, we showed that BDLM-AR can simultaneously estimate autoregressive error. The advantage of BDLM-AR model increases when strong serial correlation exists.

We adopt a Bayesian estimation framework, which facilitates modeling and inference of N-of-1 trial data in several ways. First, a key in modeling carryover effects in N-of-1 trial data is to address multicollinearity in the lag treatment effects. Under a Bayesian framework, we achieve this by postulating a prior precision matrix on the lag coefficients to provide the appropriate constraints on the lag coefficients. Specifically, the designed form of this prior precision matrix is motivated by and connected to a fused ridge penalized estimation procedure (9), which imposes shrinkage and smoothness of the lag coefficients. Second, while cross validation is often a method of choice in tuning the penalty terms (λ_i and λ_i^* via γ_1 and γ_2) in penalized estimation, it is not feasible to split the sample at random time points because of the temporal order in N-of-1 trial data. Bayesian formulation provides a natural way to tune the penalties in a data-driven manner via posterior inference. Third, we have applied our model to a real application of using white light therapy for depressive symptoms, along with other Bayesian approaches (BDLagM, BR-DLM, NB-DLM). These approaches allow for using posterior credible intervals of individual lag coefficients as inferential tools. The posterior intervals offer additional insights on the whole time course of treatment effect.

ACKNOWLEDGEMENTS

Ziwei Liao and Ying-Kuen Cheung's research are supported in part by NIH Grant R01MH109496. Min Qian's research is supported in part by NSF Grant DMS-2112938 and NIH Grant R01MH109496. Ian Kronish's research is supported in part by NCI HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services or NIH nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

DATA AVAILABILITY STATEMENT

R code for our Bayesian distributed lag model and corresponding simulations is available via the following GitHub repository, https://github.com/williammomo/BDLM-AR.

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Ying Kuen Cheung https://orcid.org/0000-0003-0882-7056

REFERENCES

1. Kravitz RL, Duan N. *Design and Implementation of N-of-1 Trials: a user's guide. Agency for healthcare research and quality*. Rockville, MD: Agency for Healthcare Research and Quality; 2014.
2. Topol EJ. Transforming medicine via digital innovation. *Sci Transl Med*. 2010;2(16):16cm4.
3. Cheung YK, Wood D, Zhang K, et al. Personal preferences for Personalised Trials among patients with chronic diseases: an empirical Bayesian analysis of a conjoint survey. *BMJ Open*. 2020;10(6):e036056.
4. Kronish IM, Cheung YK, Julian J, et al. Clinical usefulness of bright white light therapy for depressive symptoms in cancer survivors: results from a series of personalized (N-of-1) trials. *Healthcare*. 2020;8(1).
5. Johnson JA, Garland SN, Carlson LE, et al. Bright light therapy improves cancer-related fatigue in cancer survivors: a randomized controlled trial. *J Cancer Surviv*. 2018;12(2):206-215.
6. Gabler NB, Duan N, Vohra S, Kravitz RL. N-of-1 trials in the medical literature: a systematic review. *Med Care*. 2011;49(8):761-768.
7. Koyck LM. *Distributed lags and investment analysis*. Vol 4. Amsterdam: North-Holland Publishing Company; 1954.

8. Almon S. The distributed lag between capital appropriations and expenditures. *Econometrica*. 1965;33(1):178-196.
9. Bass FM, Clarke DG. Testing distributed lag models of advertising effect. *J Market Res*. 1972;9(3):298-308.
10. Welty LJ, Peng R, Zeger S, Dominici F. Bayesian distributed lag models: estimating effects of particulate matter air pollution on daily mortality. *Biometrics*. 2009;65(1):282-291.
11. Zanobetti A, Wand M, Schwartz J, Ryan L. Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics*. 2000;1(3):279-292.
12. Tibshirani R, Wang P. Spatial smoothing and hot spot detection for CGH data using the fused lasso. *Biostatistics*. 2008;9(1):18-29.
13. Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K. Sparsity and smoothness via the fused lasso. *J R Stat Soc Series B Stat Methodology*. 2005;67(1):91-108.
14. Chib S. Bayes regression with autoregressive errors: A Gibbs sampling approach. *J Econom*. 1993;58(3):275-294.
15. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal*. 2006;1(3):515-534.
16. Gelman A, Roberts GO, Gilks WR, et al. Efficient Metropolis jumping rules. *Bayesian Stat*. 1996;5(599-608):42.
17. Vats D, Knudson C. Revisiting the gelman–rubin diagnostic. *Stat Sci*. 2021;36(4):518-529.
18. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat*. 1998;7(4):434-455.
19. Gelman A, Carlin JB, Stern HS, Vehtari A, Rubin DB, Dunson DB. *Bayesian Data Analysis*. New York: Chapman and Hall/CRC; 2013.
20. Ljung GM, Box GE. On a measure of lack of fit in time series models. *Biometrika*. 1978;65(2):297-303.

SUPPORTING INFORMATION

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

How to cite this article: Liao Z, Qian M, Kronish IM, Cheung YK. Analysis of N-of-1 trials using Bayesian distributed lag model with autocorrelated errors. *Statistics in Medicine*. 2023;42(13):2044-2060. doi: 10.1002/sim.9676