Measuring effects of medication adherence on time-varying health outcomes using Bayesian dynamic linear models

LUIS F. CAMPOS*, MARK E. GLICKMAN, KRISTEN B. HUNTER

Department of Statistics, Harvard University, One Oxford Street, Cambridge, MA 02138 USA
soyluiscampos@gmail.com

SUMMARY

One of the most significant barriers to medication treatment is patients' non-adherence to a prescribed medication regimen. The extent of the impact of poor adherence on resulting health measures is often unknown, and typical analyses ignore the time-varying nature of adherence. This article develops a modeling framework for longitudinally recorded health measures modeled as a function of time-varying medication adherence. Our framework, which relies on normal Bayesian dynamic linear models (DLMs), accounts for time-varying covariates such as adherence and non-dynamic covariates such as baseline health characteristics. Standard inferential procedures for DLMs are inefficient when faced with infrequent and irregularly recorded response data. We develop an approach that relies on factoring the posterior density into a product of two terms: a marginal posterior density for the non-dynamic parameters, and a multivariate normal posterior density of the dynamic parameters conditional on the non-dynamic ones. This factorization leads to a two-stage process for inference in which the non-dynamic parameters can be inferred separately from the time-varying parameters. We demonstrate the application of this model to the time-varying effect of antihypertensive medication on blood pressure levels for a cohort of patients diagnosed with hypertension. Our model results are compared to ones in which adherence is incorporated through non-dynamic summaries.

Keywords: Antihypertensive medication; Bayesian dynamic linear models; Hypertension; Medication adherence; Time-varying outcomes; Two-stage inference.

1. INTRODUCTION

Over 85 million American adults, or about one-third of the population over 20 years old, suffer from hypertension (Benjamin and others, 2018). Approximately 16% of adults in the United States are unaware that they have hypertension (Benjamin and others, 2018). Hypertension, when left untreated, can lead to a range of severe and costly health concerns such as cardiovascular disease, stroke, and renal disease (Amery and others, 1985; Probstfield, 1991). Among the many factors associated with uncontrolled blood pressure (BP), poor adherence to prescribed antihypertensive medications is of significant concern to clinicians, health care systems, and other stakeholders (Choo and others, 1999; Morisky and others,

*To whom correspondence should be addressed.

1986; Osterberg and Blaschke, 2005). Patients who adhere poorly to their prescribed medication are at risk for worse BP outcomes (Vitolins and others, 2000; Lam and Fresco, 2015). Given the considerable variation in adherence, both across patients and temporally within a patient's treatment period, it is an open question of how to accurately measure the impact of varying adherence patterns on BP levels. Furthermore, the variation in adherence patterns creates difficulties in accurately measuring the effects of socio-demographic and health risk factors on BP levels.

This article develops a Bayesian dynamic linear model (DLM) (West and Harrison, 1997; Durbin and Koopman, 2001; Petris and others, 2009) for health measures recorded over time as a function of time-varying adherence, with a particular application to the effects of antihypertensive medication on BP levels. Bayesian DLMs have a long history as a statistical framework for forecasting and measuring trajectories in many domains, including real-time missile tracking and financial securities forecasting. We apply DLMs to describe time-varying health measures (like BP levels) as a function of detailed adherence (or other time-varying covariates) and individual demographics and comorbidities measured typically at study baseline. The application of DLMs to time-varying adherence and health measures is novel but fits naturally into the DLM framework because these measures can be tracked over time as adherence data accumulate in parallel. Because the DLM framework permits the inclusion of patient-level predictors, we can apply the model to measure the effects of socio-economic or racial characteristics or the effects of different comorbid conditions. Our model can control for differential medication adherence patterns, resulting in a more accurate measurement of the effects of other covariates.

To estimate the effects of time-varying adherence on health measures, we need access to three types of data. First, we need adherence data for each study participant; we can collect these through electronic monitoring devices. These devices electronically time-stamp each time the pill container is opened, providing an accurate recording of when a patient took their medication. Second, baseline health and socio-demographic information are typically recorded at the start of medication adherence studies and are often non-dynamic. Finally, we need health measures that might be impacted by differential medication adherence. We can often measure these during clinic visits, the timing of which may be determined by the patient and are, therefore, available only intermittently. Thus it is quite common for the number of health measures per study patient to be much smaller than the number of days on which we have medication adherence information.

Inference using the DLM framework is challenging in this case because there are considerably fewer health measures observed compared to the number of days of adherence measurements. In traditional uses of DLMs, we measure both time-varying covariates and responses at every time point, and inference for the time-varying state parameters can be accomplished using standard Bayesian updating algorithms (West and Harrison, 1997, Chapter 16). With adherence data, when we measure the responses at irregular and infrequent intervals, the usual updating approaches can be demonstrably inefficient. Instead, we develop an inferential approach in a Bayesian setting that takes advantage of factoring the posterior density into a product of two terms. The first term involves the exact marginal likelihood of the DLM, marginalizing out the dynamic state process. The second is the conditional posterior density for the state process parameters. The two-stage procedure allows us to determine the posterior distribution for the non-dynamic parameters using standard Markov chain Monte Carlo (MCMC) techniques and the state process parameters without resorting to needlessly complicated computational tools. With a DLM that has normally distributed responses and a stochastic process for the latent states that have normal innovations, the factorization is the product of two multivariate normal (MVN) densities and a prior. This factored posterior density also easily allows inference for the non-dynamic parameters, which in the setting of medication adherence is likely of interest because the researcher may want to understand the effect of baseline health characteristics on the health measures controlling for differential adherence.

We organize the remainder of this article as follows. In Section 2, we introduce a motivating example and details of the study cohort. We specify the DLM for our framework in Section 3. The model, which

assumes an autoregressive structure, accounts for possibly multivariate health measures, which may or may not be measured simultaneously. We then develop our computational approach for inference in Section 4. We apply our methods to modeling BP in Section 5, where we compare our methods to typical models used to measure the effects of adherence. We conclude in Section 6 with a discussion on the limitations and potential extensions of our methods.

2. DATA

The data we analyzed were obtained from the baseline pre-randomization period of a trial that studied the effects of a provider–patient communication skill-building intervention on adherence to antihypertensive medication and BP control (clinicaltrials.gov ID: NCT00201149). Recruitment occurred at seven outpatient primary care clinics at Boston Medical Center, an inner-city safety-net hospital affiliated with the Boston University School of Medicine. Patients enrolled in the study from August 2004 to June 2006, had to meet several eligibility criteria. Patients were of white or black race (African or Caribbean descent), were at least 21 years old, had an outpatient diagnosis of hypertension on at least three different occasions before study enrollment, and were currently on antihypertensive medication. Because all of the patients were continuing users of antihypertensive medications, a large proportion of the patients had their BP under control at the beginning of the baseline period. The cohort size was 869 patients, but not all patients had BP measured during this period. We restrict our analysis to the 503 patients who had at least one BP measurement and had available adherence data. The study involved measuring antihypertensive medication taking using Medication Event Monitoring System (MEMS) caps, a particular type of electronic pill-top monitoring device. Patients were given a bottle with a MEMS cap containing their most frequently taken antihypertensive medications. They were instructed to open the bottle each time they took a dose. MEMS caps contain a microprocessor that records the date and time the bottle was opened. The timing information was then downloaded through a wireless receiver after the patient returned the MEMS cap. Our study focused on medication-taking behavior during the entire pre-randomization baseline period of the study. A patient was considered adherent to the prescribed medication on a day if the opened the MEMS cap as many times as the prescribed dosing frequency, and not adherent otherwise. These measurements were recorded daily, but the duration over which adherence information was measured varied by patient.

BP measurements were recorded less frequently, as they were obtained as part of routine clinical care. The BP readings were measured using manual or electronic devices and were obtained by clinical staff, including physicians, nurses, and medical assistants. In cases where multiple readings were obtained on a single day (typically at the same clinic visit), the individual values were recorded. Diastolic and systolic BP values (DBP and SBP, respectively) were recorded separately.

In addition to detailed time-varying adherence and BP readings, other patient-specific baseline information was collected. Race (White versus African American), gender, and age at the start of the study were recorded for each patient. Additionally, the following comorbidities (as binary variables) were investigated in our study, given their potential impact on overall BP levels: presence of cerebrovascular disease, congestive heart failure, chronic kidney disease, coronary artery disease, diabetes mellitus, hyperlipidemia, peripheral vascular disease, and obesity (defined as body mass index greater than 30 kg/m²). These comorbidities were recorded at the start of the study using electronic medical records.

To motivate our modeling framework, consider the data from one of the patients in the study cohort displayed in Figure 1. The figure shows the DBP and SBP measurements on the 3 days the patient had their BP measured and daily indications of whether the patient was adherent. This patient was adherent for 29 of the 82 days in the study period. However, this simple summary masks the time-varying pattern of the patient's adherence. The patient began the study by being fully adherent to their prescribed medication, during which time their BP remained under control. After 20 days, the patient started becoming less adherent, and starting around day 38, the patient discontinued taking their medication altogether. Over

Fig. 1. Example BP and adherence data for a study patient. Blue squares are DBP, red dots are SBP, and the black dots near the horizontal axis indicate whether the patient was adherent on a particular day (filled dots indicate adherence, open dots indicate non-adherence).

this latter period, the patient's BP increased. By the end of the study period, the patient had DBP and SBP values that were not under control. Besides suggesting a potential link between adherence and BP measures, this example motivates using a DLM for health measures that accounts for time-varying adherence.

3. A DYNAMIC MODEL FOR MULTIVARIATE HEALTH MEASURES

We propose a statistical framework for multivariate time-varying health measures as a function of medication adherence that is a member of the class of Bayesian DLMs. Let y_{itk} be the value of the k th health measure, with $k = 1, \dots, K$, for patient i at time t , for $t = 1, \dots, T_i$. We assume that time is discretized and equally spaced (e.g., days), and that health measures are observed only at times $t_1, t_2, \dots, t_{m_i}$. Let $\mathbf{y}_{it\cdot}$ indicate a vector collecting all outcomes across the dotted index into a column vector. Then the vector $\mathbf{y}_{it\cdot}$ collects the outcomes for patient i at time t , i.e., $\mathbf{y}_{it\cdot} = (y_{it1}, \dots, y_{itK})^T$. Similarly, $\mathbf{y}_{i\cdot k}$ collects the k th outcome observed at times $(t_1, \dots, t_{m_i})$ for patient i , i.e., $\mathbf{y}_{i\cdot k} = (y_{it_1 k}, \dots, y_{it_{m_i} k})^T$. We will use this notation generally. Our framework assumes that $\mathbf{y}_{it\cdot}$ is an observed measurement generated from a distribution with mean $\boldsymbol{\mu}_{it\cdot}$, which follows a stochastic process. The framework recognizes that observed health measurements on a given day could vary around a mean level due to various influences including emotional state, activity level, recent alcohol consumption, ambient temperature, and other unobserved factors that might affect the observed health outcomes.

We model the mean process at time t with the sum of the contributions of a non-dynamic term and a dynamic term. Specifically, we assume

$$\mathbf{y}_{it\cdot} = \boldsymbol{\mu}_{it\cdot} + \boldsymbol{\varepsilon}_{it} = \boldsymbol{\beta}\mathbf{x}_i + \boldsymbol{\alpha}_{it\cdot} + \boldsymbol{\varepsilon}_{it}, \quad (3.1)$$

where $\mathbf{x}_i$ is a vector of p non-dynamic covariates typically measured at baseline and $\boldsymbol{\beta}$ are the corresponding linear coefficients. Furthermore, $\boldsymbol{\alpha}_{it}$ is a stochastic process that may depend on dynamic covariates, such as time-varying adherence. Here $\mathbf{y}_{it}$, $\boldsymbol{\mu}_{it}$, $\boldsymbol{\alpha}_{it}$, and $\boldsymbol{\varepsilon}_{it}$ are all vectors of length K corresponding to the observation, mean outcome process, latent state model, and sampling error, respectively. The non-dynamic covariate effects $\boldsymbol{\beta}$ are of dimension $K \times p$, and we use $\boldsymbol{\beta}_k$ to denote the k th row of $\boldsymbol{\beta}$, i.e., the non-dynamic covariate effects on the k th outcome. The error term for subject i at time t , $\boldsymbol{\varepsilon}_{it}$, accounts for typical sampling variability, measurement error, and possible correlation of the outcomes within patient and time. We further assume that $\boldsymbol{\varepsilon}_{it} \sim N(0, \Sigma_\varepsilon)$ with possibly non-diagonal covariance matrix Σ_ε .

We denote c_{it} as patient i 's adherence on day t . If patient i 's adherence is tracked for T_i consecutive days, we indicate this measurement set as $\{1, \dots, T_i\}$, and denote the collection of time-varying adherence measures as $\mathbf{c}_i = \{c_{it}\}_{t=1}^{T_i}$. We let $c_{it} = 1$ if patient i was adherent on day t and -1 otherwise. As we describe below, we extend the definition of c_{it} to allow the inclusion of additional time-varying covariates, we denote this vector as $\mathbf{c}_{it}$. We further let $\mathcal{T}_i = \{t_1, \dots, t_{m_i}\} \subset \{1, 2, \dots, T_i\}$ denote the set of m_i days on which the outcomes were measured for patient i . The set of outcome measurements for patient i is denoted by $\mathbf{y}_{i\cdot} = \{\mathbf{y}_{it}\}_{t \in \mathcal{T}_i}$. We typically have that $m_i < T_i$ for all i .

The latent process $\boldsymbol{\alpha}_{it}$ captures day-to-day variability of the underlying outcome measure and is assumed to be influenced by whether a patient takes the prescribed medication. In reality, the relationship between medication taking and outcome can be complex. However, for our application we assume an $AR(1)$ model on the latent states that also include the $\mathbf{c}_{it}$ as additive time-varying covariates as in (3.2). This $AR(1)$ model is consistent with exponential decay of the impact of antihypertensive medications on the latent BP process. Most antihypertensive medications are known to metabolize quickly, hence the need for consistent medication taking. For example, the effect of labetalol, a popular beta-blocker used to treat hypertension, decays exponentially (McNeil and others, 1979), and has a half-life in plasma of 3.0–3.5 h (McNeil and Louis, 1984). Our modeling framework permits other types of latent models, including $AR(k)$, as well as the inclusion of multiple time-varying covariates.

If $\mathbf{c}_{it}$ is of dimension r , then we assume a process on $\boldsymbol{\alpha}_{it}$ given by

$$\boldsymbol{\alpha}_{it} = \boldsymbol{\rho} \boldsymbol{\alpha}_{i,t-1} + \boldsymbol{\phi} \mathbf{c}_{it} + \mathbf{v}_{it}, \quad (3.2)$$

where $\boldsymbol{\rho}$ is the K -dimensional diagonal matrix of first-order autoregressive parameters, that is, $\boldsymbol{\rho} = \text{diag}(\rho_1, \rho_2, \dots, \rho_K)$. We assume $|\rho_k| < 1$ to ensure stationarity. The time-varying outcome effects $\boldsymbol{\phi}$ is a $K \times r$ matrix. We let $\boldsymbol{\phi}_k$ denote the k th row of $\boldsymbol{\phi}$, i.e., the time-varying covariate effects on the k th outcome. We further assume in (3.2) that the innovations $\mathbf{v}_{it}$ have zero-mean MVN distributions with diagonal covariance matrix $\Sigma_v = \text{diag}(\sigma_{v1}^2, \sigma_{v2}^2, \dots, \sigma_{vK}^2)$. Additionally, we assume the initial state $\boldsymbol{\alpha}_{i1} \sim N_K(0, \Sigma_0)$ with $\Sigma_0 = \text{diag}(\sigma_{01}^2, \sigma_{02}^2, \dots, \sigma_{0K}^2)$. Let $\boldsymbol{\theta}$ denote the non-dynamic parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \Sigma_\varepsilon, \boldsymbol{\rho}, \boldsymbol{\phi}, \Sigma_v, \Sigma_0)$. Equations (3.1) and (3.2) define the distributions $p(\mathbf{y}_{i\cdot} | \boldsymbol{\alpha}_{i\cdot}, \mathbf{x}_i, \boldsymbol{\theta})$ and $p(\boldsymbol{\alpha}_{i\cdot} | \boldsymbol{\alpha}_{i,t-1}, \mathbf{c}_{i\cdot}, \boldsymbol{\theta})$, respectively.

Figure 2 contains a simulated example of a patient with a scalar outcome over a 30-day period corresponding to the model in (3.1) and (3.2). As evidenced in Figure 2, the mean process generally decreases on days when a patient adheres to their medication. However, this is not always the case, and an increase can occur when the corresponding innovation v_t is large and positive, offsetting the impact of the patient taking their medication. The observed health measure is normally distributed around the mean for the day on which we record the outcome measure. On day 1, for example, the health measure is higher than the mean, and on day 15, the health measure is lower than the mean.

This model is attractive for relating health measures to time-varying adherence for several reasons. With it, we can consider the impact of non-dynamic baseline variables separately from the impact of time-varying adherence. This breakdown allows us to disentangle the effects of adherence and patient-specific covariates. Furthermore, in settings where we observe repeated outcome measures at time t , inference

Fig. 2. A 30-day trajectory for a simulated patient from a DLM with a scalar outcome. The bottom of the figure displays the adherence indicator simulated independently with a 90% probability of adherence. The adherence effect is simulated to be large and negative ($\phi = -0.5$). The contribution of the non-dynamic covariates is assumed to be $\mathbf{x}_t^T \boldsymbol{\beta} = 130$.

for the sampling variability, Σ_ε , can be made more precise and can be separated from the innovation variability Σ_v .

Additionally, the $AR(1)$ model component has a well-known asymptotic mean under full adherence and full non-adherence. Specifically, the impact of being fully adherent converges to $\boldsymbol{\alpha}_{it} \rightarrow (I - \rho)^{-1} \boldsymbol{\phi}$ for increasing values of t . We can therefore predict that if patient i were to continue to be fully adherent, their mean outcomes would tend to $\boldsymbol{\mu}_{it} \rightarrow \boldsymbol{\beta} \mathbf{x}_i + (I - \rho)^{-1} \boldsymbol{\phi}$. We can perform an analogous calculation when the patient is fully non-adherent. This property permits estimating the best-case or worst-case health measure means for perfect adherence or perfect non-adherence.

It is worth noting that there are several computational tools available in aiding the selection of the latent parameter model in the Bayesian context. For example, we could employ Bayes factors (Jeffreys, 1939; Weinberg, 2012), cross-validatory approaches (Vehtari and others, 2016), or modern methods explicitly developed for model selection in DLMs (Shao and others, 2018). For application to our modeling approach, these tools could help with the decision to include multiple lags or additional covariates. We consider some of these model selection criteria when comparing the model specified above to other models that account for medication adherence. However, we do not explore these approaches more generally for latent variable model selection in this article, as this is beyond the scope of our work.

4. MARGINAL DLMs

Inference for the model in Section 3 is challenging given the large number of parameters (both $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ are of high dimension). In particular, if the number of recorded adherence indicators per patient is large, then the size of $\boldsymbol{\alpha}$ is similarly large. Standard Bayesian inferential methods through posterior simulation approaches for such a highly parameterized model can result in slow convergence and unreliable inferences.

Advances in Bayesian computation have made highly parameterized dynamic models tractable even when we cannot compute the likelihood directly, or that the model involves non-linear relationships. The setting described in Section 3, however, contains a tractable and straightforward structure.

With recent developments in sequential Monte Carlo (SMC), including software packages like Libbi (Murray, 2015), sampling the high-dimensional latent states of the DLM has become computationally feasible and accessible. These advances focus on situations where the likelihood cannot be computed directly and instead must be approximated. Marginal sampling schemes, like the particle marginal Metropolis–Hastings algorithm (Andrieu and others, 2010), alternately sample the structural parameters and latent process parameters and accept or reject them with an adjusted Metropolis–Hastings step accounting for the approximation of the likelihood needed in sampling the latent space. Recent work (Bhattacharya and Wilson, 2018) approximates the posterior of the structural parameters on a grid of points, reducing the possible sampling values to a discrete set. Chopin and others (2012) extended SMC to situations where there is a known sequential structure of otherwise intractable likelihoods. In this section, we first compute the marginal likelihood of the data given the structural parameters exactly. Then we compute the exact posterior distribution of the latent process parameters conditional on structural parameters relying on only simple probability calculations.

Specifically, our approach takes advantage of factoring the posterior joint density of θ and α as follows

$$p(\alpha, \theta | y) = p(\theta | y) p(\alpha | \theta, y), \quad (4.1)$$

where we omit the dependence on both $\{c_{it}\}$ and x_i . The first factor in (4.1) is discussed below, while the second factor, the conditional posterior density of the latent process parameters, can be derived exactly and is discussed in Section 4.2.

Marginal inference about θ can be accomplished by integrating the left side of (4.1) with respect to α , yielding the first factor in the expression. This factor can be expanded using Bayes' Theorem as

$$p(\theta | y) = \frac{p(y | \theta) p(\theta)}{p(y)}. \quad (4.2)$$

The marginal likelihood, $p(y | \theta)$, in the numerator is determined from

$$p(y | \theta) = \int p(y | \alpha, \theta) p(\alpha | \theta) d\alpha. \quad (4.3)$$

As we derive in Section 4.1, the marginal distribution in (4.3) is MVN with a mean and variance that depend only on the fixed model parameters θ , the adherence measures $\{c_{it}\}$, and individual covariates x_i . This marginalization is made possible by the normality of the latent state innovations v_{it} and sampling error ε_{it} . Inference for θ can therefore be obtained directly from this marginal distribution.

4.1. Marginal likelihood of the DLM

We begin our calculation of the marginal likelihood $p(y | \theta)$ by deriving it in the case of a DLM with a single outcome variable ($k = 1$) and for one individual, and assume a single baseline covariate ($p = 1$) and single time-varying covariate ($r = 1$), which is assumed to be an adherence indicator. The parameters in θ are therefore $(\beta, \sigma_\varepsilon^2, \rho, \phi, \sigma_v^2, \sigma_0^2)$. For now we assume that both adherence indicators and *potential* time-varying outcomes are measured over T consecutive days; the outcomes are $(y_1, y_2, \dots, y_T)$ and the adherence indicators are $c = (c_1, c_2, \dots, c_T)$. Later, we will subset the outcomes to the *actual* observation times. We

can write the vector of outcomes as the sum of a shared non-time-varying component, a time-varying component and an error term as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} \beta x \\ \beta x \\ \vdots \\ \beta x \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_T \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix} = \beta x \mathbb{1}_T + \boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \quad (4.4)$$

where we use the $\mathbb{1}_T$ notation to indicate a column vector of length T consisting of all ones. The following development is conditional on $\boldsymbol{\theta}$ and covariates (x, c) unless noted otherwise. Thus the first term on the right-hand side of (4.4), $\beta x \mathbb{1}_T$, is treated as a constant vector. The third term is distributed as $\boldsymbol{\varepsilon} \sim N_T(\mathbf{0}, \sigma_\varepsilon^2 I)$, where $\mathbf{0}$ is a column vector of zeros, and I is a T -dimensional identity matrix.

Because the initial latent variable α_1 is normally distributed, and α_t is a sequentially defined linear combination of normal random variables, $(\alpha_1, \dots, \alpha_T)$ is itself MVN. The AR(1) structure of the latent states admits a recursive mean and variance calculation

$$\mathbb{E}[\alpha_t] = \rho \mathbb{E}[\alpha_{t-1}] + \phi c_t \quad \text{and} \quad \mathbb{V}ar[\alpha_t] = \rho^2 \mathbb{V}ar[\alpha_{t-1}] + \sigma_v^2.$$

The recursion and initial conditions imply a general formula for the mean and variance of α_t

$$E_t = \mathbb{E}[\alpha_t] = \phi \sum_{k=2}^t \rho^{t-k} c_k \quad (4.5)$$

$$V_t = \mathbb{V}ar[\alpha_t] = \sigma_v^2 \sum_{k=2}^t \rho^{2(t-k)} + \rho^{2(t-1)} \sigma_0^2 \quad (4.6)$$

for $t \in \{2, \dots, T\}$ with initial mean $E_1 = \mathbb{E}[\alpha_1] = 0$ and initial variance $V_1 = \mathbb{V}ar[\alpha_1] = \sigma_0^2$. We collect the mean terms into a vector

$$\mathbf{E} = (E_1, \dots, E_T). \quad (4.7)$$

Applying a similar recursion to the covariance, we obtain that $\mathbb{C}ov(\alpha_t, \alpha_{t-k}) = \rho^k V_{t-k}$ for $t \in \{2, \dots, T\}$ and $k \in \{0, \dots, t-1\}$. More compactly, the covariance matrix is

$$\boldsymbol{\Sigma} = \begin{pmatrix} V_1 & \rho V_1 & \rho^2 V_1 & \dots & \rho^{T-1} V_1 \\ \rho V_1 & V_2 & \rho V_2 & \dots & \rho^{T-2} V_2 \\ \rho^2 V_1 & \rho V_2 & V_3 & \dots & \rho^{T-3} V_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} V_1 & \rho^{T-2} V_2 & \rho^{T-3} V_3 & \dots & V_T \end{pmatrix}. \quad (4.8)$$

The distribution of the terms in (4.4) can be written explicitly as

$$(y_1, y_2, \dots, y_T)^T \sim N_T(\beta x \mathbb{1}_T + \mathbf{E}, \boldsymbol{\Sigma} + \sigma_\varepsilon^2 I). \quad (4.9)$$

The distribution in (4.9) is the marginal likelihood for T sequentially observed outcomes, marginalizing out the latent parameters.

Let $\mathcal{T} = \{t_1, t_2, \dots, t_m\}$ be the *actual* observation times, and let $\mathbf{E}_{\mathcal{T}} = (E_{t_1}, E_{t_2}, \dots, E_{t_m})^T$. By properties of the MVN distribution, the outcome vector $\mathbf{y}_{\cdot} = (y_{t_1}, y_{t_2}, \dots, y_{t_m})^T$ is also MVN with mean and covariance, respectively

$$\mathbb{E}[\mathbf{y}_{\cdot}] = \beta x \mathbf{1}_{|\mathcal{T}|} + \mathbf{E}_{\mathcal{T}} \quad \text{and} \quad \mathbb{V}ar[\mathbf{y}_{\cdot}] = \Sigma_{\mathcal{T} \times \mathcal{T}} + \sigma_{\varepsilon}^2 I_{|\mathcal{T}|}. \quad (4.10)$$

The notation $\Sigma_{\mathcal{T} \times \mathcal{T}}$ indicates subsetting the covariance matrix Σ to the corresponding rows and columns designated by $\mathcal{T}$. If, for example, $\mathcal{T} = \{1, 7, 9\}$ the first, seventh, and ninth rows and columns are taken from Σ , making $\Sigma_{\mathcal{T} \times \mathcal{T}}$ a 3×3 matrix. One consequence of the marginalization is that if we observe multiple measurements of the same outcome on a single day, they will only vary according to σ_{ε}^2 .

It is worth noting that even though this marginal distribution does not depend on the time-varying parameters, the autoregressive structure remains in both the mean and covariance. For example, the centered marginal mean of y_t is

$$\mathbb{E}[y_t] - \beta x = \rho (\mathbb{E}[y_{t-1}] - \beta x) + \phi c_t.$$

This is the same recursive relationship of the dynamic component of our DLM in (3.2). Even though we marginalize out the time-varying parameters, their sequential structure is retained. Now that that we have marginalized the likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ in this special case, we generalize this calculation to include the model described in Section 3.

The calculations above are for one study participant and a single outcome variable but extend to multivariate outcome measurements. The marginal mean and covariance are both individual- and outcome-dependent, as both $\boldsymbol{\phi}_k$ and α_{ik} vary by outcome and individual. When referring to multiple study participants and multiple outcome variables, we can make the dependency explicit by denoting the quantities in (4.7) and (4.8) for outcome k of individual i as $\mathbf{E}_{ik}$ and Σ_{ik} , respectively.

Because the majority of the structure is contained within each outcome (recall $\boldsymbol{\rho}$ is a diagonal matrix), we organize the multivariate outcomes as follows. We collect the outcome k observed through time $t = 1, \dots, T_i$ into a $T_i \times 1$ vector $\mathbf{y}_{i \cdot k} = (y_{i1k}, y_{i2k}, \dots, y_{iT_i k})^T$. The marginal likelihood of a vector of complete outcome measurements can be written as

$$\begin{pmatrix} \mathbf{y}_{i \cdot 1} \\ \mathbf{y}_{i \cdot 2} \\ \vdots \\ \mathbf{y}_{i \cdot k} \end{pmatrix} \mid \boldsymbol{\theta} \sim N_{kT} \left(\mathbf{E}_i + \boldsymbol{\beta} \mathbf{x}_i \otimes \mathbf{1}_{T_i}, \Sigma_i + \Sigma_{\varepsilon} \otimes I_{T_i} \right), \quad (4.11)$$

where $\otimes$ is the Kronecker product, and where

$$\mathbf{E}_i = \begin{pmatrix} \mathbf{E}_{i1} \\ \mathbf{E}_{i2} \\ \vdots \\ \mathbf{E}_{ik} \end{pmatrix} \quad \text{and} \quad \Sigma_i = \begin{pmatrix} \Sigma_{i1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Sigma_{i2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Sigma_{ik} \end{pmatrix}. \quad (4.12)$$

As in (4.10), we can derive the marginal distribution of the outcome measures $\mathbf{y}_{i \cdot} \mid \boldsymbol{\theta}$ by subsetting the appropriate rows and columns of the quantities in (4.12).

If we assume that outcomes between study participants are independent conditional on $(\boldsymbol{\theta}, \mathbf{x}, \mathbf{c})$, the marginal distribution of outcomes unconditional on the time-varying parameters is given by

$$p(\mathbf{y}_{1 \cdot}, \dots, \mathbf{y}_{n \cdot} \mid \boldsymbol{\theta}, \mathbf{x}, \mathbf{c}) = \prod_{i=1}^n p(\mathbf{y}_{i \cdot} \mid \boldsymbol{\theta}, \mathbf{x}_i, \mathbf{c}_{i \cdot}). \quad (4.13)$$

This simple marginalization drastically reduces the number of parameters in the model and simplifies computation. We can conduct inference on our non-dynamic parameters in a Bayesian setting by introducing a prior distribution for θ and using the marginal likelihood in (4.13). The posterior density for the non-dynamic parameters is given by

$$p(\theta | \{y_{i..}\}_{i=1}^n, \mathbf{x}, \mathbf{c}) \propto p(\theta) \prod_{i=1}^n p(y_{i..} | \theta, \mathbf{x}_i, \mathbf{c}_{i..}). \quad (4.14)$$

With this re-expression of the posterior, inference can be performed in a straightforward and efficient way via MCMC, or using Hamiltonian Monte Carlo (Homan and Gelman, 2014) sampling to obtain draws from the posterior distribution, as implemented in STAN (Carpenter and others, 2017). Alternatively, posterior samples could be obtained using SMC, which takes advantage of the sequential likelihoods $p(y_{it_j} | y_{it_1}, \dots, y_{it_{j-1}}, \theta)$, which are easily available in our framework because they are conditional distributions of the MVN distribution in (4.11).

4.2. Inference for the latent process

Posterior inference for α can be determined once $p(\theta | \mathbf{y})$ has been obtained by exploiting the factorization of the posterior density in (4.1). To do so, we note that the joint distribution of α and $\mathbf{y}$ conditional on θ is MVN. For our development, we consider the case in which outcomes are observed for T consecutive days. With a k -dimensional outcome variable, α contains kT elements and $\mathbf{y}$ contains kT (scalar) outcomes. Using the notation in (4.12), the joint distribution of outcomes and time-varying parameters is

$$\left(\begin{array}{c} \alpha_{i,1} \\ \vdots \\ \alpha_{i,k} \\ \mathbf{y}_{i,1} \\ \vdots \\ \mathbf{y}_{i,k} \end{array} \right) \bigg| \theta \sim N_{2kT} \left(\left(\begin{array}{c} \mathbf{E}_i \\ \mathbf{E}_i + \beta \mathbf{x}_i \otimes \mathbf{1}_{T_i} \end{array} \right), \left(\begin{array}{cc} \Sigma_i & \Sigma_i \\ \Sigma_i & \Sigma_i + \Sigma_\varepsilon \otimes I_{T_i} \end{array} \right) \right). \quad (4.15)$$

The quantities on the diagonal of the covariance matrix in (4.15) were determined in Section 4.1. The covariance of $\mathbf{y}_{i,k}$ and $\alpha_{i,\ell}$ (for $k \neq \ell$), conditional on θ , can be shown to be $\mathbf{0}$ because of independence assumptions made in Section 3 (with conditioning on θ suppressed).

$$\begin{aligned} \text{Cov}(\alpha_{i,\ell}, \mathbf{y}_{i,k}) &= \mathbb{E} [\text{Cov}(\alpha_{i,\ell}, \mathbf{y}_{i,k} | \alpha_{i,k})] + \text{Cov}(\mathbb{E}[\alpha_{i,\ell} | \alpha_{i,k}], \mathbb{E}[\mathbf{y}_{i,k} | \alpha_{i,k}]) \\ &= \mathbb{E} [\text{Cov}(\alpha_{i,\ell}, \varepsilon_{i,k})] + \text{Cov}(\alpha_{i,\ell}, \alpha_{i,k}) = \mathbf{0}. \end{aligned}$$

At this point, we can subset the mean vector and covariance matrix of (4.15) as in (4.10) to account for the sporadically observed outcomes, but for simplicity we continue assuming that the response values are observed at each time period. The conditional posterior distribution of the latent parameters is determined from the joint distribution in (4.15) by conditioning on $\mathbf{y}$ as

$$\left(\begin{array}{c} \alpha_{i,1} \\ \vdots \\ \alpha_{i,k} \end{array} \right) \bigg| \mathbf{y}_{i,1}, \dots, \mathbf{y}_{i,K}, \theta \sim N_{kT}(\tilde{\alpha}_i, \tilde{\Sigma}_i), \quad (4.16)$$

where

$$\tilde{\alpha}_i = \mathbf{E}_i + \Sigma_i (\Sigma_i + \Sigma_\varepsilon \otimes I_{T_i})^{-1} \left[\begin{pmatrix} \mathbf{y}_{i,1} \\ \vdots \\ \mathbf{y}_{i,k} \end{pmatrix} - (\mathbf{E}_i + \beta \mathbf{x}_i \otimes \mathbb{1}_{T_i}) \right]$$

and

$$\tilde{\Sigma}_i = \Sigma_i - \Sigma_i (\Sigma_i + \Sigma_\varepsilon \otimes I_{T_i})^{-1} \Sigma_i.$$

We can use the joint distribution specified in (4.16) along with the posterior distribution on the structural parameters shown in (4.14) to sample from the latent process parameters by first simulating $\theta^* \sim p(\theta | \mathbf{y}_{1..}, \dots, \mathbf{y}_{n..}, \mathbf{x}, \mathbf{c})$ and then sampling from $p(\alpha_{i,1}, \dots, \alpha_{i,K} | \mathbf{y}_{1..}, \dots, \mathbf{y}_{n..}, \theta^*)$.

5. APPLICATION TO BP STUDY

We apply our modeling framework to time-varying BP as a function of adherence to antihypertensive medication and baseline comorbidities and demographics described in Section 2. We begin with a discussion of models often used in this task that incorporate adherence as non-dynamic information. We also address missing adherence measures for some patients.

5.1. Non-dynamic models incorporating adherence

Medication adherence is typically incorporated into outcome models as a non-dynamic linear covariate in one of two ways. It is either included as the overall average adherence or a binary indicator for being above a certain threshold, indicating “poor” versus “good” adherence (Rose and others, 2011; Lee and others, 2006; Schroeder and others, 2004). Either of these approaches includes adherence as a non-dynamic covariate in the model. Repeated outcome measures are modeled with patient-specific random effects.

We present these alternative models in the context of BP outcomes, a bivariate measure. We label the outcomes 1 for SBP and 2 for DBP. These models take the form

$$\begin{pmatrix} y_{it1} \\ y_{it2} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_i \beta_1 \\ \mathbf{x}_i \beta_2 \end{pmatrix} + \begin{pmatrix} \delta_{i1} \\ \delta_{i2} \end{pmatrix} + \begin{pmatrix} \bar{c}_i \gamma_1 \\ \bar{c}_i \gamma_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{it1} \\ \varepsilon_{it2} \end{pmatrix}. \quad (5.1)$$

We consider two different possible adherence measures, $\bar{c}_i$. We first consider average adherence, $\bar{c}_i = T_i^{-1} \sum_{t=1}^{T_i} (c_{it} + 1)/2$, the fraction of days adherent. In this case, we would interpret the adherence effect parameters (γ_1, γ_2) as the differences in BP of being fully adherent relative to being fully non-adherent, controlling for the baseline covariates. We also consider a dichotomized summary $\bar{c}_i = \mathbb{1} \left\{ \left(T_i^{-1} \sum_{t=1}^{T_i} c_{it} \right) > 2p - 1 \right\}$, an indicator of overall adherence. In this case, (γ_1, γ_2) would be interpreted as the differences in BP for those with “good” ($\bar{c}_i = 1$) versus “poor” ($\bar{c}_i = 0$) adherence. Several values of p were considered to assess the sensitivity to this choice, but we present results for $p = 0.8$ which is a conventional choice (Schroeder and others, 2004). The model in (5.1) includes patient-specific random effects $\delta_{i1} \sim N(0, \sigma_{\delta_1}^2)$ and $\delta_{i2} \sim N(0, \sigma_{\delta_2}^2)$. The introduction of these random effects, two for each additional patient, has implications when calculating quantities used for formal model comparison. However, the hierarchical nature of the random effects permits straightforward inferences of the posterior distribution.

We compare the fit of the above two non-dynamic models to our dynamic model framework. The non-dynamic approach has a potential advantage of being more robust to model misspecification relative

to our dynamic model, particularly with the choice of the specific model for the evolution of the health measures as a function of daily adherence. However, incorporating average adherence may mask significant time-varying effects of detailed adherence. We explore this trade-off in Section 5.3.

5.2. Missing adherence indicators

Some patients in our cohort have missing adherence indicators. Among the 503 patients in our analyses, 70 have at least 1 day of missing adherence. Adherence could be missing due to MEMS cap malfunctions, hospital inpatient stays in which the MEMS containers were not used, or other causes. Fifty of the 70 patients had only one or two missing adherence values. In the most extreme case, one patient was missing 48 out of 102 adherence measures.

For patient i , let $\mathbf{c}_i^{\text{obs}} = \{c_{it}\}_{t \in \mathcal{T}_{\text{obs}}}$ be the set of observed adherence values and $\mathbf{c}_i^{\text{mis}} = \{c_{it}\}_{t \in \mathcal{T}_{\text{mis}}}$ be the set of missing adherence values. Letting η_i represent the parameters describing a potential adherence model for patient i , the posterior density in (4.14) conditional on the observed adherence data only is

$$\begin{aligned} p(\boldsymbol{\theta} | \mathbf{y}_{i..}, \mathbf{x}_i, \mathbf{c}_i^{\text{obs}}) &= \int p(\boldsymbol{\theta}, \mathbf{c}_i^{\text{mis}}, \eta_i | \mathbf{y}_{i..}, \mathbf{x}_i, \mathbf{c}_i^{\text{obs}}) d\mathbf{c}_i^{\text{mis}} d\eta_i \\ &= \int p(\boldsymbol{\theta} | \mathbf{y}_{i..}, \mathbf{x}_i, \mathbf{c}_i^{\text{obs}}, \mathbf{c}_i^{\text{mis}}, \eta_i) p(\mathbf{c}_i^{\text{mis}} | \mathbf{y}_{i..}, \mathbf{x}_i, \mathbf{c}_i^{\text{obs}}, \eta_i) p(\eta_i | \mathbf{y}_{i..}, \mathbf{x}_i, \mathbf{c}_i^{\text{obs}}) d\mathbf{c}_i^{\text{mis}} d\eta_i \\ &\stackrel{(a)}{=} \int p(\boldsymbol{\theta} | \mathbf{y}_{i..}, \mathbf{x}_i, \mathbf{c}_i^{\text{obs}}, \mathbf{c}_i^{\text{mis}}, \eta_i) p(\mathbf{c}_i^{\text{mis}} | \eta_i) p(\eta_i | \mathbf{c}_i^{\text{obs}}) d\mathbf{c}_i^{\text{mis}} d\eta_i. \end{aligned} \quad (5.2)$$

The last line in (5.2) labeled (a) involves a set of modeling choices. First, we assume that the adherence model parameters do not depend on other covariates or the outcomes (conditional on the adherence values), and this is reflected by assuming $p(\eta_i | \mathbf{y}_{i..}, \mathbf{x}_i, \mathbf{c}_i^{\text{obs}}) = p(\eta_i | \mathbf{c}_i^{\text{obs}})$. Other approaches, such as those described by [Naranjo and others \(2013\)](#), provide a framework for modeling missing time-varying covariates in DLMS with complex models. Our assumption is more conservative since it does not use additional potentially informative data, but it is also likely more robust to model misspecification. Second, $p(\mathbf{c}_i^{\text{mis}} | \mathbf{y}_{i..}, \mathbf{x}_i, \mathbf{c}_i^{\text{obs}}, \eta_i) = p(\mathbf{c}_i^{\text{mis}} | \eta_i)$ implies that the missing adherence values can be simulated directly from our adherence model, represented by η_i , without regard to individual-level characteristics. Again, this is a conservative choice because adding information to the imputation model could help improve predictions if properly modeled.

We simulate the missing adherence measures from Beta-Bernoulli distributions that depend only on adherence measures for each patient separately. Letting η_i represent patient i 's average adherence, the unobserved adherence values are drawn from a Bernoulli distribution

$$\{c_{it}^*\}_{t \in \mathcal{T}} | \eta_i \stackrel{i.i.d.}{\sim} \text{Bernoulli}(\eta_i),$$

where the $c_{it}^* = 2c_{it} - 1$ take on values $\{-1, 1\}$. Under a uniform prior distribution on patient i 's adherence rate, the posterior distribution for η_i is given by

$$\eta_i | \{c_{it}\}_{t \in \mathcal{T}_{\text{obs}}} \sim \text{Beta}(n_{i,1} + 1, n_{i,-1} + 1)$$

when they were observed to be adherent $n_{i,1}$ days and non-adherent $n_{i,-1}$ days.

We repeated the simulation process 20 times and combined the posterior samples for a Bayesian multiple-imputation analysis ([Little and Rubin, 2002](#)) for our non-dynamic models. For our fully Bayesian model, we incorporate the simulated adherence values into our posterior simulation analyses. We found that in both cases, the inference of our non-dynamic parameters was not sensitive to the missing adherence values.

Table 1. *Number of BP readings during the study period*

Number of BP readings	1	2	3	4	5	6	7	8+
Number of patients	226	125	72	28	16	12	15	9
Percent of patients	44.90	24.90	14.30	5.60	3.20	2.40	3.00	1.79

5.3. Analysis of BP measures

Table 1 displays the number of BP measurements among our 503 patients during the study period. While a majority, 351, of patients had either one or two BP readings, the rest, 152, had three or more. The number of BP readings totaled 1152, averaging 2.29 readings per patient.

For the period of the study, the proportion of days that patients took their medication varied widely. The proportions ranged from 8.9% to 100%, with a median of 95.1% and a mean of 88.9%. These adherence rates on average were high, with many patients being fully adherent throughout the study period. Only 19.1% of patients had below 80% adherence. Though poor medication adherence is a known challenge in the general population, some studies have also shown a high degree of adherence in trial settings (Krousel-Wood and others, 2004). The average number of days in the study was 98 days per patient (minimum and maximum of 21 and 395 days, respectively, with an interquartile range of 14), and 75% of patients were followed between 84 and 112 days.

Baseline summaries of the non-dynamic covariates appear in Table 2. A majority of the cohort consisted of women, more patients of black race (African or Caribbean descent) than white, and a significant fraction of low-income patients. The cohort also consisted of mostly obese patients and had a moderately high comorbidity burden (89% had at least one comorbidity). Based on 503 patients with BP readings within 14 days of enrollment, the cohort on average had relatively well-controlled hypertension at baseline. Again, this may not be surprising because the cohort eligibility criteria required patients to be continuing users of antihypertensive medications. Many patients had relatively controlled BP measures at the beginning of the study. The cohort consisted of only 26.2% having DBP > 80 mm Hg and SBP > 130 mm Hg, and 7.8% having DBP > 90 mm Hg, and SBP > 140 mm Hg.

We assume the same prior distributions for all models when possible. For $k = 1, 2$, indicating SBP and DBP measures separately, we assume the following.

$$\begin{aligned}
 \rho_k &\sim U(-1, 1), \phi_k \sim N(0, 25), \gamma_k \sim N(0, 25) \\
 \sigma_{\varepsilon k} &\sim U(0, 30), \rho_{\varepsilon} \sim U(-1, 1) \\
 \sigma_{v_k} &\sim U(0, 10), \sigma_{0k} \sim U(0, 30) \\
 \beta_{11} &\sim N(120, 400), \beta_{12} \sim N(80, 400) \\
 \beta_{jk} &\sim N(0, 400), j = 2, \dots, p \\
 \sigma_{Uk} &\sim U(0, 30)
 \end{aligned}$$

The prior components were selected to be vague but proper. The intercepts β_{1k} have distributions centered near the typical SBP and DBP but had variances that were sufficiently large to acknowledge the uncertainty in the effects. We assume a uniform prior component with compact support for the standard deviation parameters, as recommended by Gelman (2006). The correlation and autocorrelation parameters were also assumed to have uniform priors as in the dynamic model parametrization. Sensitivity to the choice of prior distribution may depend generally on a number of factors, including cohort size, number of outcomes per patient, predictability of the covariates, and so on. A more detailed discussion of approaches to sensitivity

Table 2. *Baseline socio-demographic and health characteristics of the patients in the study cohort*

	Mean (Std Dev)
Age (years)	60.3 (11.1)
DBP at enrollment	79.7 (11.5)
SBP at enrollment	133.5 (19.6)
	Percent
Female	67.5
African-American	54.9
Income below \$20 000	44.3
Obese	60.6
Cerebral vascular disease	5.6
Congestive heart failure	3.8
Renal insufficiency	6.8
Coronary artery disease	14.9
Diabetes	36.2
Hyperlipidemia	57.1
Peripheral vascular disease	6.2

analyses can be found in Sections 2.8 and 2.9 of [Gelman and others \(2013\)](#). Such analyses are beyond the scope of this article. The convergence of the MCMC simulated values was inspected with trace plots of multiple chains and using the Gelman–Rubin convergence statistic ([Gelman and Rubin, 1992](#)) and Bulk and Tail Effective Sample Sizes ([Vehtari and others, 2019](#)). We include detailed summaries of the diagnostics in the [Supplementary materials](#) available at *Biostatistics* online.

Table 3 presents posterior means and 95% central posterior intervals for the non-dynamic covariate effects using the DLM. We report the point estimates and intervals for the DLM only; we indicate the effects in the alternative models that were significant at the 0.05 level with a † (for the average adherence model) or a ‡ (for the dichotomized adherence model). We mark coefficient effects that are significantly different from 0 (at the 0.05 level) with an asterisk (*). The estimated covariate effects are similar across all models, with the point estimates agreeing in magnitude and sign. Even though the DLM covariate effects tended to have narrower intervals on average (3% reduction), the significance of the findings tended to agree as well. In particular, the effect of race (white versus non-white) was significantly negative, indicating that whites tended to have lower BP controlling for all other variables and time-varying adherence. This effect was not captured by the non-DLM models for DBP. Patients who were obese at the beginning of the study tended to have significantly higher DBP and SBP. These findings are consistent with the results of previous studies ([Kressin and others, 2010](#); [Rose and others, 2011](#)) in their significance and direction of the effects. These agreed with the alternative models in terms of significance and directionality of the effect.

Here, we include age as a categorical variable instead of a numerical coefficient because BP is known to change non-linearly with age [Franklin \(1999\)](#). We divide the patients by their age at enrollment into four groups: ages 50 and younger (group 0), 51 to 60 (group 1), 61 to 70 (group 2), and 71 and older (group 3). The effects given in Table 3 are relative to the youngest group. According to [Franklin \(1999\)](#), SBP increases with age while DBP increases on average until age 50–60, after which it decreases with age. This trend is evidenced in our analysis with the covariate effect of age for SBP increasing with age while decreasing for DBP.

Table 3. *Summaries of covariate effects for the bivariate dynamic linear model*

Variable	Systolic	95% CI	Diastolic	95% CI
Intercept	132.97	128.8 to 137.2* ^{†‡}	85.68	83.25 to 88.11* ^{†‡}
Sex (male)	-1.1	-4.06 to 1.86	0.53	-1.15 to 2.21
Age (group 1)	-0.01	-3.93 to 3.91	-1.5	-3.76 to 0.76
Age (group 2)	1.87	-2.09 to 5.83	-3.54	-5.82 to -1.26* ^{†‡}
Age (group 3)	5.22	0.86 to 9.58* ^{†‡}	-6.73	-9.23 to -4.22* ^{†‡}
White	-3.31	-6.05 to -0.56* ^{†‡}	-1.58	-3.14 to -0.01*
Obese	3.07	0.32 to 5.8* ^{†‡}	1.68	0.09 to 3.26* ^{†‡}
Nicotine dependence	-0.9	-6.02 to 4.24	1.3	-1.64 to 4.24
Hyperlipidemia	-1.44	-4.17 to 1.26	-1.34	-2.89 to 0.21
Diabetes	1.44	-1.4 to 4.31	-2.98	-4.6 to -1.35* ^{†‡}
Peripheral vascular disease	-1.28	-6.62 to 4.05	-2.6	-5.62 to 0.45
Renal insufficiency	-0.67	-5.69 to 4.31	-2.73	-5.59 to 0.12
Benign prostatic hypertrophy	2.82	-4.97 to 10.49	-1.59	-6.07 to 2.86
Coronary artery disease	-1.66	-5.39 to 2.04	-2.76	-4.89 to -0.64* ^{†‡}
Congestive heart failure	-0.78	-7.03 to 5.48	0.59	-3.03 to 4.18
Cerebral vascular disease	2.45	-3.08 to 7.93	-0.17	-3.33 to 2.99

*Effect with 95% posterior interval not containing 0 in DLM.

[†]Significant effect at the 0.05 level in average adherence model.

[‡]Significant effect at the 0.05 level in dichotomized adherence model.

Table 4. *Measurement error estimates for three models*

Variable	Systolic	95% CI	Diastolic	95% CI
Dynamic linear model				
Standard errors	13.37	12.27–14.42	8.11	7.6–8.62
Correlation	0.62	0.56–0.68		
Average adherence model				
Standard errors	14.53	13.74–15.36	8.34	7.9–8.8
Correlation	0.58	0.53–0.62		
Dichotomized adherence model				
Standard errors	14.5	13.74–15.28	8.34	7.9–8.79
Correlation	0.58	0.53–0.62		

Table 4 reports inferences for the standard error and correlations of (ϵ_{it}) for all three models. Comparing the sampling standard deviation estimates across models indicates the gains in modeling the adherence effects as time-varying. The standard error estimates for the alternative adherence models tend to be slightly larger than those given by the DLM. Indeed, the values for the sampling standard deviations are more consistent with studies where they measure this variability in detail (*Rose and others, 2011*). The time-varying adherence explicitly captured in the DLM may account for the extra variation in the outcomes of the alternative adherence models through a reduction in the estimated measurement error variance.

Table 5 contains the adherence effects estimated from our models. The adherence effects across the three models are not directly comparable because of the different approaches to incorporating adherence. The average adherence model indicates that the difference between those who were fully adherent and

Table 5. Adherence effects for the models considered

Variable	Systolic	95% CI	Diastolic	95% CI
Dynamic linear model				
Adherence effect	−0.48	−0.92 to −0.16*	−0.24	−0.48 to −0.08*
Asymptotic adherence effect	−3.87	−6.41 to −1.43*	−3.15	−4.64 to −1.71*
Alternative adherence model:				
average adherence				
Adherence effect	−9.24	−17.42 to −0.89*	−9.46	−14.3 to −4.67*
Alternative adherence model: dichotomized				
adherence ($p = 0.8$)				
Adherence effect	−5.49	−8.85 to −2.33*	−3.78	−5.61 to −1.91*

*Significant effect at the 0.05 level.

those who were fully non-adherent, controlling for other covariates, is about −9.2 and −9.5 for systolic and diastolic, respectively. In other words, a 10% additive increase in adherence corresponds to a 0.92 and 0.95 reduction in SBP and DBP, respectively. However, this effect size assumes that the relationship between average adherence and BP is linear and holds throughout the entire range of adherence. Given the limited range of average adherence observed in the data (90% of patients have average adherence above 70.5%), interpreting this effect beyond this range is not recommended because it involves extrapolating beyond the data. The dichotomized adherence model shows similar results. In particular, this approach involves comparing those with relatively good overall adherence (above 80% adherent) to everyone else. Based on the dichotomized adherence model, patients in the former group tend to have a lower BP of −5.49 and −3.78 on average for SBP and DBP. This substantial difference, however, may be due to influence from the few patients with very poor adherence. Nevertheless, both models seem to indicate a sizable and significant benefit of lower BP to patients who adhere to antihypertensive medication.

The DLM gives similar results. However, we can infer a day-to-day effect of taking medication. The results of the model fit suggest a small but significant reduction of BP from following a prescribed regimen on 1 day, −0.24 mm Hg, and −0.12 mm Hg on average for SBP and DBP, respectively. These estimates imply a patient who is fully adherent over many consecutive days would experience a long-term reduction in SBP by −3.9 mm Hg, and a long-term reduction in DBP by −3.15 mm Hg. The magnitudes of an increase in long-term BP for continued non-adherence is the same but in the opposite direction. The daily adherence effects tend to agree in terms of the significance and direction for the different models. However, the DLM provides a more precise interpretation of these parameters that is consistent with the time-varying nature of the data.

Model selection of our DLM and two models involving aggregate adherence measures can proceed in a variety of ways. We consider an approach using a predictive criterion (Vehtari and others, 2016) through a leave-one-out cross-validation measure, and evidence-based criteria (Jeffreys, 1939; Weinberg, 2012) from Monte Carlo estimates of the marginal likelihood. We show the results of these selection criteria in Table 6 for the two aggregate adherence models introduced in this section and for the DLM. The predictive criteria shown are the Bayesian leave-one-out estimate of the out-of-sample predictive fit, denoted elpd_{loo} , the effective number of parameters p_{loo} , and the relative differences in elpd_{loo} across models. These quantities are calculated using the `loo` package in the R programming language (Vehtari and others, 2019). The evidence criteria shown is the Monte Carlo estimate of the marginal likelihood and is calculated using bridge sampling (Meng and Wong, 1996) as implemented in the `bridgesampling` package (Gronau and others, 2019). The predictive criteria include standard error estimates in parentheses.

Table 6. *Model-selection criteria for comparing DLM to random-effects models using predictive criteria and evidence-based criteria with the `loo` and the `bridgesampling` R packages, respectively*

Model	Predictive criteria			Evidence criteria
	elpd_{100}	p_{100}	elpd diff	$\log p(y \text{Model})$
Dichotomized adherence	−8893.2 (325.7)	469.5 (17.3)	0.0 (0.0)	−9457.33
Average adherence	−8899.2 (326.0)	474.8 (18.3)	−6.0 (4.6)	−9459.33
Dynamic linear model	−8981.8 (330.8)	53.7 (6.5)	−88.6 (14.0)	−6787.64

For the predictive criteria, we show both estimates and standard error estimates (in parentheses).

Fig. 3. The Pareto k diagnostics are shown for each of the three models considered in Section 5. Largest values of $\hat{k}$ indicate a large degree of sensitivity of the posterior to the individual data points. The posterior for the alternative adherence models (average adherence and dichotomized adherence) seem to be highly sensitive to the leave-one-out procedure as $\hat{k} > 0.7$ for a large number of data points.

The predictive criteria select the dichotomized adherence model to be the best among the three models considered, having the highest elpd_{100} among the models. While the average adherence model does not seem to perform much worse (the difference in elpd_{100} is within 2 standard errors of zero), the DLM does seem to perform much worse than the dichotomized model. The difference between the dichotomized model and the DLM is more than 6 standard errors away from zero. This difference indicates that the DLM would give worse predictions than either of the aggregate models.

As discussed in [Vehtari and others \(2016\)](#), we must take care when interpreting these predictive criteria for model selection. In short, the individual contributions of each data point to the overall likelihood can be leveraged to approximate the leave-one-out predictive fit using importance weighting. These importance weights also give us a measure of the influence of the individual data points to the overall posterior. In particular, the importance weights for the individual contributions provide a measure of the sensitivity

of the posterior distribution to each data point. Figure 3 shows that the posterior for the DLM model is indeed less sensitive to this leave-one-out procedure and that the other models are highly sensitive. Indeed, Vehtari and others (2016) recommend that all $\hat{k}$ should be less than 0.7. The large values of $\hat{k}$ for the two alternative models indicate that the results of elpd_{loo} should be suspect.

As an alternative to the predictive selection criterion, the final column of Table 6 shows the estimated log of the marginal likelihood of a model, $\log p(y|\text{Model})$, as calculated using bridge sampling (Meng and Wong, 1996). The ratio of the marginal likelihoods (or the difference in log-marginal likelihoods) is directly interpretable as the relative plausibility of two competing models. Here, we select the DLM over the two competing models. However, as discussed in Shao and others (2018), these quantities are highly sensitive to prior specification and parameter dimensionality. In particular, the introduction of priors for $(\delta_{i1}, \delta_{i2}), i = 1, \dots, N$ make the interpretation of the smaller log-marginal likelihoods for the average and dichotomized adherence models suspect.

Summing up, the models under consideration perform better or worse under different selection criteria. These criteria, however, are known to be sensitive to the prior specification and dimensionality of the parameter space. The models we consider here exhibit both of these characteristics, limiting our ability to interpret model selection straightforwardly. However, the introduction of a time-dependence structure of our DLM helps interpret the time-varying nature of our health outcomes and their relationship to time-varying covariates like medication adherence.

Another benefit of the DLM is our ability to infer the latent BP for unobserved days using the procedure discussed in Section 4.2. Figure 4 contains point-wise posterior credible intervals of the mean process $\mu_{it} = \beta x_i + \alpha_{it}$ for the patient presented in Figure 1. The solid colored lines indicate the posterior latent BP process while the dashed lines indicate a 95% credible envelope across time. That is, for each time t , a 95% credible interval of $\alpha_{i1}, \alpha_{i2} | \{y_{i..}\}_{i=1}^n$ is shown. The estimated mean DBP and SBP processes vary about βx_i gradually over time. These shifts vary naturally due to the autocorrelation parameters (ρ_1, ρ_2) , but they are also seen to vary directionally by the influence of medication taking.

6. DISCUSSION

In this article, we propose a multivariate DLM that we use to model time-varying outcome measures as a function of medication adherence and other time-varying covariates. Typical analyses in this setting ignore the time-varying structure of medication adherence and instead examine the relationship between adherence and outcomes via correlations with non-dynamic adherence measures (e.g., time-averaged adherence). Our framework explicitly provides a measure of the daily impact of medication taking and meaningful bounds for mean BP while controlling for baseline covariates. Furthermore, while DLMs are of frequent use, our particular setting is unique in that it benefits from a two-stage computational approach made possible by factorizing the posterior density into the dynamic and non-dynamic parameters.

Our modeling approach is sufficiently flexible to permit different assumptions from those in our hypertension application. For example, our assumption of an AR(1) structure for the mean outcome process was justifiable in our analysis because the effect of antihypertensive medication decays exponentially (McNeil and others, 1979). In other settings, alternative mean processes can be considered, including higher-order AR processes or growth curves. We model our baseline covariates linearly, but our framework permits non-linear inclusion of covariate information for both baseline non-dynamic covariates, as well as the dynamic predictors in the mean process component of our model. Our approach can be extended to perform non-dynamic covariate selection within a Bayesian setting using any of a number of different approaches, including placing a Laplace prior on the covariate effects (Park and Casella, 2008), or assuming spike-and-slab prior distributions on the covariate effects (Ishwaran and Rao, 2005).

A crucial assumption of our framework is that the outcome distribution is MVN, as this assumption allows for the marginalization strategy that leads to an efficient computational procedure. However, with

Fig. 4. For the patient in Figure 1, we impute the missing adherence on days 0 to 10 (grey dots) and predict the mean BP $\{\mu_{i,\cdot}\}$ using 95% central point-wise posterior intervals. The solid lines are the estimated mean DBP and SBP, and the dashed lines are the point-wise posterior intervals. Horizontal dotted lines are drawn at the baseline-estimated mean DBP and SBP.

some non-normal outcome distributions, various strategies can be employed that can potentially take advantage of the marginalization idea. For example, we can approximate non-normal outcome densities with normal distributions, after which we can perform the marginalization. Then, we can incorporate a Metropolis–Hastings algorithm into posterior sampling to account for the non-normality of the original sampling distribution. Such a procedure is likely to be far more efficient than sampling the dynamic parameters directly as part of model fitting.

The proposed modeling framework also permits other extensions that are straightforward to incorporate. First, we can accommodate multivariate outcomes where components are recorded at different intervals by marginalizing the distribution in (4.11) appropriately. For example, BP may only be recorded during regular clinic visits, but resting heart rate, a related outcome, can be recorded daily with the introduction of personal medical devices or smartwatches. Second, if we believe that the effect of adherence to medication varies from person to person, we can apply hierarchical priors to patient-specific adherence effects. This feature would allow for varying adherence effects across patients but also allows us to share information across individuals. Third, our analyses did not acknowledge differences among antihypertensive medications, but our framework easily permits distinguishing differential medication effects. The effects of different medications, perhaps grouped by relevant characteristics such as whether the medication is short-acting or long-lasting, or by medication type (e.g., for hypertension, diuretics, Angiotensin-converting-enzyme inhibitors), can be included as separate time-varying effects in the dynamic component of our model. Dosages and dosing frequencies can serve as covariates for the effects of particular medications.

The framework we developed can help establish answers to questions of interest to clinicians and medical researchers that have been difficult to assess through simpler models. In particular, our model can estimate the effects of different socio-demographic or health factors on health outcomes that control for detailed time-varying adherence to medication by allowing medication-taking behaviors to change over

time. From our framework, we can estimate the daily improvement in being adherent to one's medication, controlling for socio-demographic and health characteristics, but also the likely long-range achievable mean outcomes. We can also use the model to forecast health outcomes as a function of specified patterns of adherence, potentially serving as a tool for medical decision-making by clinicians. Our approach provides a robust framework for understanding the impacts of poor medication adherence as clinicians and patients work together to improve their medication treatment.

REPRODUCIBILITY

We include analysis and simulation code in a GitHub repository at <https://github.com/lfcampos/Medication-Adherence>.

SUPPLEMENTARY MATERIAL

Supplementary material is available at <http://biostatistics.oxfordjournals.org>.

ACKNOWLEDGMENTS

The authors would like to thank Pierre Jacob, Stephane Shao, Maxime Rischard, and the Health Statistics Working Group at Harvard University for their insightful discussions and comments. We also thank the reviewers and editors for their thoughtful comments and efforts towards enhancing our paper. Any errors remain our own. *Conflict of Interest:* The authors have no conflicts of interest to declare.

FUNDING

Support for this work was provided by funding from the National Institutes of Health grant R21 HL121366.

REFERENCES

- AMERY, A., BRIKKO, P., CLEMENT, D., DE SCHAEPRDYVER, A., FAGARD, R., FORTE, J., HENRY, J. F., LEONETTI, G., O'MALLEY, K. AND STRASSER, T. (1985). Mortality and morbidity results from the European Working Party on High Blood Pressure in the Elderly trial. *The Lancet* **325**, 1349–1354.
- ANDRIEU, C., DOUCET, A. AND HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**, 269–342.
- BENJAMIN, E. J., VIRANI, S. S., CALLAWAY, C. W., CHAMBERLAIN, A. M., CHANG, A. R., CHENG, S., CHIUVE, S. E., CUSHMAN, M., DELLING, F. N., DEO, R. and others (2018). Heart disease and stroke statistics—2018 update: a report from the American Heart Association. *Circulation* **137**, 67–492.
- BHATTACHARYA, A. AND WILSON, S. P. (2018). Sequential Bayesian inference for static parameters in dynamic state space models. *Computational Statistics & Data Analysis* **127**, 187 – 203.
- CARPENTER, B., GELMAN, A., HOFFMAN, M. D., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P. AND RIDDELL, A. (2017). Stan: a probabilistic programming language. *Journal of Statistical Software* **76**, 1–32.
- CHOO, P. W., RAND, C. S., INUI, T. S., LEE, M.-L. T., CAIN, E., CORDEIRO-BREAULT, M., CANNING, C. and PLATT, R. (1999). Validation of patient reports, automated pharmacy records, and pill counts with electronic monitoring of adherence to antihypertensive therapy. *Medical Care* **37**, 846–857.
- CHOPIN, N., JACOB, P. E. AND PAPASPILIOPOULOS, O. (2012). SMC 2: an efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**, 397–426.

- DURBIN, J. AND KOOPMAN, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- FRANKLIN, S. S. (1999). Ageing and hypertension: the assessment of blood pressure indices in predicting coronary heart disease. *Journal of Hypertension. Supplement : Official Journal of the International Society of Hypertension* **17**, S29–36.
- GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* **1**, 515–534.
- GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian Data Analysis*, 3rd edition. Chapman and Hall/CRC.
- GELMAN, A. AND RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–472.
- GRONAU, Q. F., SINGMANN, H., FORSTER, J. J., WAGENMAKERS, E.-J., THE JASP TEAM, GUO, J., GABRY, J., GOODRICH, B., MULDER, K. AND DE VALPINE, P. (2019). *bridgesampling: Bridge Sampling for Marginal Likelihoods and Bayes Factors*. R package version 0.7-2. <https://CRAN.R-project.org/package=bridgesampling>.
- HOMAN, M. D. AND Gelman, A. (2014). The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* **15**, 1593–1623.
- ISHWARAN, H. AND RAO, J. S. (2005). Spike and slab variable selection: frequentist and Bayesian strategies. *Annals of Statistics* **33**, 730–773.
- JEFFREYS, H. (1939). *Theory of Probability*, 1st ed. *Oxford Classic Texts in the Physical Sciences*. Oxford: Oxford Univ. Press.
- KRESSIN, N. R., ORNER, M. B., MANZE, M., GLICKMAN, M. E. AND BERLOWITZ, D. (2010). Understanding contributors to racial disparities in blood pressure control. *Circulation: Cardiovascular Quality and Outcomes* **3**, 173–180.
- KROUSEL-WOOD, M., THOMAS, S., MUNTNER, P. AND MORISKY, D. (2004). Medication adherence: a key factor in achieving blood pressure control and good clinical outcomes in hypertensive patients. *Current Opinion in Cardiology* **19**, 357–362.
- LAM, W. Y. AND FRESCO, P. (2015). Medication adherence measures: an overview. *BioMed Research International* **1**, 1–12.
- LEE, J. K., GRACE, K. A. AND TAYLOR, A. J. (2006). Effect of a pharmacy care program on medication adherence and persistence, blood pressure, and low-density lipoprotein cholesterol: a randomized controlled trial. *Journal of the American Medical Association* **296**, 2563–2571.
- LITTLE, R. J. A. AND RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*. Hoboken, New Jersey, USA: John Wiley & Sons, Inc.
- MCNEIL, J. J., ANDERSON, A. E., LOUIS, W. J. AND MORGAN, D. J. (1979). Pharmacokinetics and pharmacodynamic studies of labetalol in hypertensive subjects. *British Journal of Clinical Pharmacology* **8**, 157S–161S.
- MCNEIL, J. J. AND LOUIS, W. J. (1984). Clinical pharmacokinetics of labetalol. *Clinical Pharmacokinetics* **9**, 157–167.
- MENG, X.-L. AND WONG, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica* **6**, 831–860.
- MORISKY, D. E., GREEN, L. W. AND LEVINE, D. M. (1986). Concurrent and predictive validity of a self-reported measure of medication adherence. *Medical Care* **24**, 67–74.
- MURRAY, L. M. (2015). Bayesian state-space modelling on high-performance hardware using libbi. *Journal of Statistical Software* **67**, 1–26.
- NARANJO, A., TRINDADE, A. A. AND CASELLA, G. (2013). Extending the state-space model to accommodate missing values in responses and covariates. *Journal of the American Statistical Association* **108**, 202–216.

- OSTERBERG, L. AND BLASCHKE, T. (2005). Adherence to medication. *New England Journal of Medicine* **353**, 487–497.
- PARK, T. AND CASELLA, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association* **103**, 681–686.
- PETRIS, G., PETRONE, S. AND CAMPAGNOLI, P. (2009). *Dynamic Linear Models with R*. New York, NY: Springer.
- PROBSTFIELD, J. L. (1991). Prevention of stroke by antihypertensive drug-treatment in older persons with isolated systolic hypertension-final results of the Systolic Hypertension in the Elderly Program (SHEP). *Journal of the American Medical Association* **265**, 3255–3264.
- ROSE, A. J., GLICKMAN, M. E., D’AMORE, M. M., ORNER, M. B., BERLOWITZ, D. AND KRESSIN, N. R. (2011). Effects of daily adherence to antihypertensive medication on blood pressure control. *The Journal of Clinical Hypertension* **13**, 416–421.
- SCHROEDER, K., FAHEY, T. AND EBRAHIM, S. (2004). How can we improve adherence to blood pressure-lowering medication in ambulatory care? Systematic review of randomized controlled trials. *Archives of Internal Medicine* **164**, 722–732.
- SHAO, S., JACOB, P. E., DING, J. AND TAROKH, V. (2018). Bayesian model comparison with the Hyvärinen score: computation and consistency. *Journal of the American Statistical Association*, 1–24. DOI: 10.1080/01621459.2018.1518237.
- VEHTARI, A., GABRY, J., YAO, Y. AND GELMAN, A. (2019a). *loo: Efficient Leave-one-out Cross-validation and WAIC for Bayesian Models. R package version 2.1.0*. <https://CRAN.R-project.org/package=loo>.
- VEHTARI, A., GELMAN, A. AND GABRY, J. (2016). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* **27**, 1413–1432.
- VEHTARI, A., GELMAN, A., SIMPSON, D., CARPENTER, B. AND BÜRKNER, P.-C. (2019). *Rank-normalization, folding, and localization: An improved R-hat for assessing convergence of MCMC*. arXiv preprint arXiv:1903.08008.
- VITOLINS, M. Z., RAND, C. S., RAPP, S. R., RIBISL, P. M. AND SEVICK, M. A. (2000). Measuring adherence to behavioral and medical interventions. *Controlled Clinical Trials* **21**, 188S–194S.
- WEINBERG, M. D. (2012). Computing the Bayes factor from a Markov chain Monte Carlo simulation of the posterior distribution. *Bayesian Analysis* **7**, 737–770.
- WEST, M. AND HARRISON, J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd edition. New York, NY, USA: Springer New York, Inc.

[Received December 7, 2018; revised November 10, 2019; accepted for publication December 2, 2019]