

# Hybrid statistical and mechanistic mathematical model guides mobile health intervention for chronic pain

Sara M. Clifton\*

*Department of Engineering Sciences and Applied Mathematics,  
Northwestern University, Evanston, IL 60208, USA*

Chaeryon Kang

*Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261, USA*

Jingyi Jessica Li

*Department of Statistics, University of California Los Angeles, Los Angeles, CA 90095, USA*

Qi Long

*Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104, USA*

Nirmish Shah

*Department of Medicine, Duke University, Durham, NC 27710, USA*

Daniel M. Abrams

*Department of Engineering Sciences and Applied Mathematics,  
Northwestern University, Evanston, IL 60208, USA*

*Northwestern Institute for Complex Systems, Northwestern University, Evanston, IL 60208, USA and  
Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA*

(Dated: June 9, 2017)

Nearly a quarter of visits to the Emergency Department are for conditions that could have been managed via outpatient treatment; improvements that allow patients to quickly recognize and receive appropriate treatment are crucial. The growing popularity of mobile technology creates new opportunities for real-time adaptive medical intervention, and the simultaneous growth of “big data” sources allows for preparation of personalized recommendations. Here we focus on the reduction of chronic suffering in the sickle cell disease community. Sickle cell disease is a chronic blood disorder in which pain is the most frequent complication. There currently is no standard algorithm or analytical method for real-time adaptive treatment recommendations for pain. Furthermore, current state-of-the-art methods have difficulty in handling continuous-time decision optimization using big data. Facing these challenges, in this study we aim to develop new mathematical tools for incorporating mobile technology into personalized treatment plans for pain. We present a new hybrid model for the dynamics of subjective pain that consists of a dynamical systems approach using differential equations to predict future pain levels, as well as a statistical approach tying system parameters to patient data (both personal characteristics and medication response history). Pilot testing of our approach suggests that it has significant potential to predict pain dynamics given patients’ reported pain levels and medication usages. With more abundant data, our hybrid approach should allow physicians to make personalized, data driven recommendations for treating chronic pain.

## INTRODUCTION

In the fields of physics, chemistry, and engineering, models are often derived from mechanistic fundamental laws expressed in the form of differential equations. Resulting “dynamical systems” models can be used both to gain intuition into the expected behavior of the system, and to make specific predictions about results of experiments (e.g., see [1]). In fields such as social sciences, bioinformatics, and medicine, models are often constructed from data via statistical inference, without direct derivation from fundamental principles (e.g. [2]).

The mechanistic and statistical approaches to mathematical modeling have different advantages. The former allows prior knowledge to be introduced and validated or rejected based on the success of the model. The latter requires almost no a-priori information about how the system is expected to behave.

In this paper, we present a hybrid approach to mathematical modeling that incorporates both mechanistic and statistical elements, with the goal of gaining a deeper understanding of the human experience of subjective pain. Specifically, we hope to predict how patient-reported pain levels vary over time based on medication dosage information and other patient characteristics.

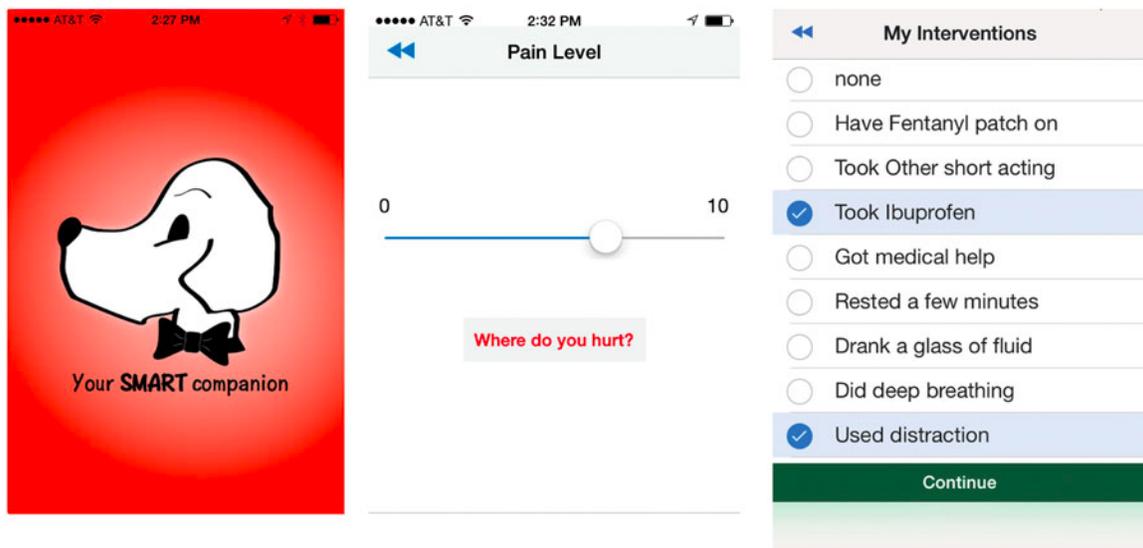


FIG. 1. **Smartphone app.** Sample images of SMART app for iPhone/Android smartphone devices.

### Application to pain

Sickle cell disease (SCD) is a chronic illness associated with frequent medical complications and hospitalizations. Approximately 90% of acute care visits are for pain events, and 30-day hospital re-utilization rates are alarmingly high [3]. While factors influencing these high re-utilization rates are poorly understood, close follow-up and continued use of pain medications has been shown to decrease re-hospitalization rates. Mobile technology has become an integral part of health care management, and our recently self-developed mobile application (Sickle cell Mobile Application to Record symptoms via Technology, or SMART app—see Figure 1) for SCD assists with documentation and intervention of pain.

Pain in particular is difficult to quantify and has never before been monitored at the temporal scale we report here across so many patients. It is known that subjective pain, though indeed subjective, is correlated with objective measurable stimuli qualities in experiments (see, e.g., [4–6]). Thus there is reason to believe that subjective pain may follow understandable dynamics in time, especially when mitigated by opioid or non-opioid drugs. Our approach to the problem is motivated by the hope that a reasonable model for pain dynamics will yield some level of predictive power, despite the clear expectation that there will also be significant noise within and across patients. We can attribute the stochastic variation to sources like patient mood, temporal changes in patient state, weather, etc. In contrast, we hope that patient attributes like age, gender, SCD disease type, etc. will remain roughly constant on the time scale of the experiment and allow us to explore possible correlation of these attributes with model parameters.

### Data source: mobile health application

We seek to understand the temporal dynamics of chronic pain as experienced by SCD patients. To that end, we have developed a mobile phone application that allows patients to record medication usage and subjective pain levels (measured on a 0-10 scale) in real time [7, 8].

Figure 1 shows several images of the app interface, while Figure 2 shows a typical data set resulting from a single patient’s use of the app over the course of several weeks.

## MATERIALS AND METHODS

### Data

As of October 2016, data were available from 47 patients using the SMART app. Data sets from 8 of those patients were excluded because of excessive sparsity based on the following criteria: (1) total number of reports  $\leq 5$ ; or (2) pain reports never exceeded zero during the period under consideration. See Table I for demographic details of included patients. We denote the sample size  $n = 39$ .

### Predictive model

In order to develop a hybrid model that incorporates both a mechanistic a-priori knowledge-driven component and a statistical data-driven component, we divide tasks

Demographic characteristics		N	(%)
Institution	A	14	(35.9)
	B	17	(43.6)
	C	8	(20.5)
Gender	Male	16	(41.0)
	Female	23	(59.0)
Age at baseline (years)	18-34	24	(61.5)
	> 34	15	(38.5)
SCD disease type	Hemoglobin SC	8	(20.5)
	Hemoglobin SS	22	(56.4)
	Hemoglobin SB+ (Beta) Thalassemia	5	(12.8)
	Beta-Zero Thalassemia	3	(7.7)
	SO-Ara	1	(2.6)
Hydroxyurea user		27	(69.2)
Folic acid vitamin user		26	(66.7)
Long-acting opioid user		29	(74.4)
Short-acting opioid user		35	(89.7)
Non-opioid user		29	(74.4)
		Mean	SD (Min, Max)
Number of pain reports		67.2	60.4 ( 9.0, 257.0 )
Days of pain reports		164.6	109.6 ( 10.3, 435.1 )
Within-patient average VAS score		4.7	2.1 ( 0.3, 9.4 )
		Mean	SD (Min, Max)
Number of pain reports (first 2 weeks)		13.2	9.6 ( 2.0, 45.0 )
Number of long-acting opioid doses (first 2 weeks)		6.0	8.4 ( 0.0, 35.0 )
Number of short-acting opioid doses (first 2 weeks)		7.2	7.5 ( 0.0, 35.0 )
Number of non-opioid doses (first 2 weeks)		2.1	3.1 ( 0.0, 12.0 )

TABLE I. Patient demographic information and the number of pain reports supplied by patients across entire study.

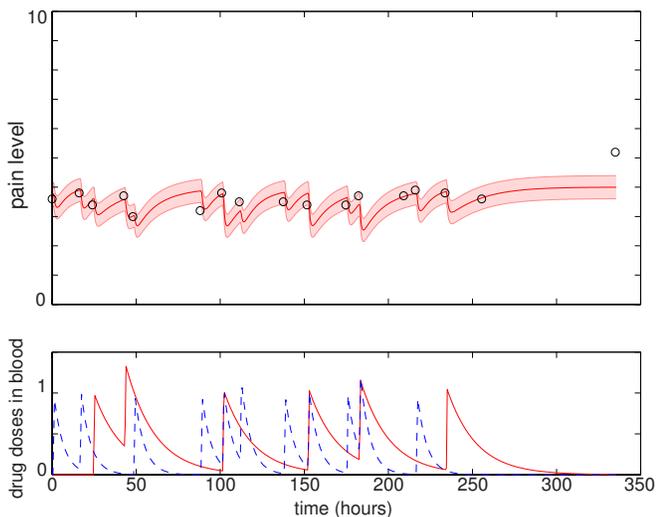


FIG. 2. **Sample pain and medication data from a single patient.** Upper panel: patient reported pain (black circles) and model fit (red solid line); red shading indicates model fit plus/minus one standard deviation. Lower panel: long-acting methadone (red solid line) and short-acting oxycodone (blue dashed line) medication concentrations in patient bloodstream as inferred from medication usage reported via the SMART application.

into two disjoint sets that fit these two categories; see the Discussion section for more context. We begin with a “dynamical systems” model for subjective pain mo-

tivated by the hypothesis that human sensory systems function on a roughly “return to setpoint” basis [9–12]. Any model of human pain response, however, will inevitably require specification of a variety of parameters determining the time scale(s) and degree of severity of the response. The statistical modeling tasks employ patient data to infer parameters (1) from patient characteristics and population distributions and (2) from patient-specific pain and medication response history.

To make this more concrete, in Figure 3 we present a flow chart summarizing our approach to the hybrid modeling problem. Steps  $I_2$  and A comprise the statistical modeling component; steps B and C comprise the mechanistic modeling component. A further optimization step D builds on the predictions of the hybrid model to allow for a balance between competing demands of pain reduction and medication usage minimization. This paper details steps  $I_1$ ,  $I_2$ , A and B. We leave the remaining steps for future work.

#### *Mechanistic component (for every patient)*

We propose and evaluate two related mechanistic models based on a set of coupled ordinary differential equations (ODEs), either (a) deterministic or (b) stochastic. The stochastic differential equation (SDE) model comprises a Langevin equation, which can be converted into a Fokker-Planck partial differential equation (PDE) for the

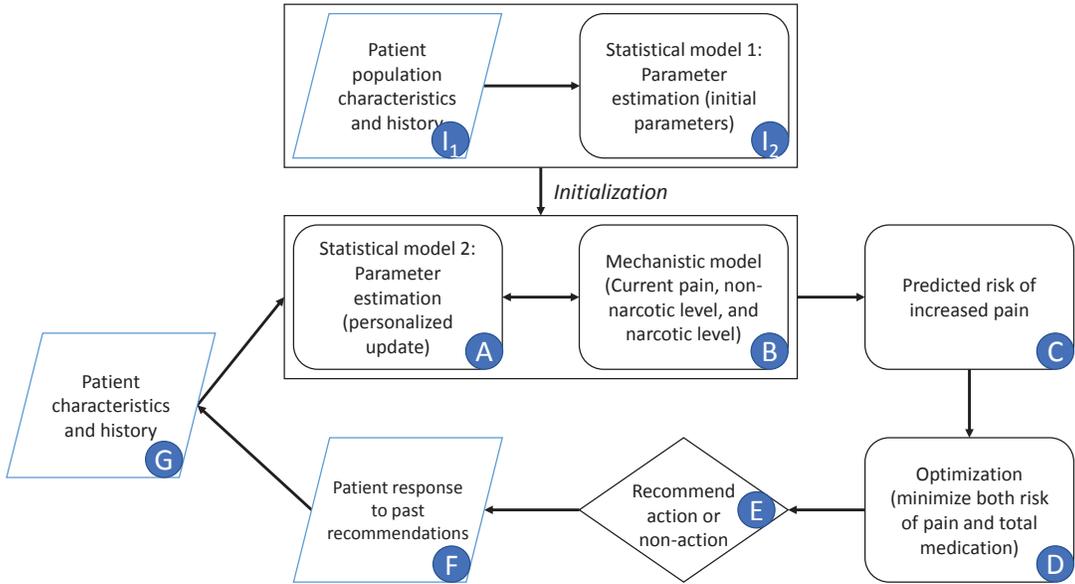


FIG. 3. **Schematic flowchart showing model framework.** Rounded rectangles represent modeling or computation steps, rhombuses represent data inputs or outputs, and diamond represents decision step. Items  $I_1$  and  $I_2$  are only necessary for initialization of the model. Items A through E are the focus of this paper.

evolution of the probability distribution for pain  $\rho(P, t)$  [13]. This allows for prediction of both the expected pain level for a patient at any point in the future *and* an assessment of the confidence in (and a confidence interval for) that prediction.

Mathematically, the deterministic mechanistic model we propose is the following, for a single patient:

$$\begin{aligned}
 \frac{dP}{dt} &= -(k_0 + k_1 D_1 + k_2 D_2 + k_3 D_3)P + k_0 u \\
 \frac{dD_1}{dt} &= -k_{D_1} D_1 + \sum_{j=1}^{N_1} \delta(t - \tau_{1,j}) \\
 \frac{dD_2}{dt} &= -k_{D_2} D_2 + \sum_{j=1}^{N_2} \delta(t - \tau_{2,j}) \\
 \frac{dD_3}{dt} &= -k_{D_3} D_3 + \sum_{j=1}^{N_3} \delta(t - \tau_{3,j}),
 \end{aligned} \tag{1}$$

where  $P$  is the patient pain level (on a scale of 1–10),  $k_0$  is the pain relaxation rate without drugs,  $k_i$  is the marginal effect on the pain relaxation rate due to drug  $i$  ( $i = 1, 2, 3$ ),  $u$  is the unmitigated pain level (i.e. without drug intervention),  $D_i$  is the amount of standard drug  $i$  doses within the patient,  $k_{D_i}$  is the elimination rate of drug  $i$  within the patient,  $\{\tau_{i,j}\}_{j=1}^{N_i}$  are the drug  $i$  dosage times, and  $N_i$  is the number of doses of drug  $i$  taken.  $\delta$  represents the Dirac delta function. Note that the parameters and variables will in general need to be indexed with distinct values for each patient in a population, though we omit those indices here for clarity. For convenience and clarity we also here omit “hatted” notation (e.g.,

$\hat{P}(t)$ ) sometimes used for model predictions.

Tables II and III summarize the meanings of model variables and parameters, respectively.

Variable	Meaning	Units
$P(t)$	Instantaneous pain level on 0–10 scale	[pain]
$D_1(t)$	Concentration of drug 1 (long-acting opioid) in the body	[standard doses]
$D_2(t)$	Concentration of drug 2 (short-acting opioid) in the body	[standard doses]
$D_3(t)$	Concentration of drug 3 (non-opioid) in the body	[standard doses]
$\rho(P, t)$	Instantaneous probability distribution of pain level $P$	[probability]

TABLE II. Variables in mechanistic models.

In this very simple model for pain dynamics (1), pain is expected to relax at rate  $k_0$  to unmitigated level  $u$  set by aggravating factors (like sickle cell disease) in the absence of intervention through opioids (drugs 1 and 2) or non-opioids (drug 3). When drugs are present in the patient’s body, pain drops at a faster rate and the short-term equilibrium pain level (not the unmitigated pain level  $u$ ) is reduced. Note that we treat all parameters as constant over the time period of interest, which we take to be two weeks (based on clinical heuristic experience).

In the model for drug concentrations, medication in the body is assumed to be metabolized at a constant

Parameter	Meaning	Units
$u$	unmitigated pain level	[pain]
$k_0$	rate of decrease of pain in the absence of drugs or acute sources of pain	$[T^{-1}]$
$k_1$	effect of drug 1 (long-acting opioid) on pain relaxation rate	$[T^{-1}]$
$k_2$	effect of drug 2 (short-acting opioid) on pain relaxation rate	$[T^{-1}]$
$k_3$	effect of drug 3 (non-opioid) on pain relaxation rate	$[T^{-1}]$
$k_{D_1}$	rate of decay of drug 1 (long-acting opioid) in body due to metabolism	$[T^{-1}]$
$k_{D_2}$	rate of decay of drug 2 (short-acting opioid) in body due to metabolism	$[T^{-1}]$
$k_{D_3}$	rate of decay of drug 3 (non-opioid) in body due to metabolism	$[T^{-1}]$
$\varepsilon$	amplitude of intrinsic variability in human subjective pain reports	$[T^{1/2}]$
$N_i$	number of standard drug $i$ doses taken	[count]
$\{\tau_{i,j}\}$	drug $i$ dose times (indexed by $j$ )	$[T]$

TABLE III. Parameters in mechanistic models.

rate. Rates can be determined from existing substantiated pharmacokinetic models (e.g., [14, 15]); Dirac delta function onset of medication serum concentration is a good approximation to the fast rise typical of the medications under consideration. See Fig. 2 for a sample deterministic model output.

Note that we deliberately chose to employ an extremely simple conceptual model for pain dynamics. More sophisticated versions might be developed to incorporate higher order dynamics for  $P$ , or to include nonlinear or nonautonomous effects (e.g., allowing for explicit parameter variation with time of day or year), but currently available data are insufficient to constrain a model of greater complexity.

The stochastic differential (Langevin) equation version of our mechanistic model is as follows:

$$\begin{aligned}
 dP &= -(k_0 + k_1 D_1 + k_2 D_1)P dt + k_0 u (dt + \varepsilon dW) \\
 dD_1 &= \left( -k_{D_1} D_1 + \sum_{j=1}^{N_1} \delta(t - \tau_{1,j}) \right) dt, \\
 dD_2 &= \left( -k_{D_2} D_2 + \sum_{j=1}^{N_2} \delta(t - \tau_{2,j}) \right) dt \\
 dD_3 &= \left( -k_{D_3} D_3 + \sum_{j=1}^{N_3} \delta(t - \tau_{3,j}) \right) dt,
 \end{aligned} \tag{2}$$

where a hypothesis of uncorrelated additive white noise has been made. From this we derive the Fokker-Planck equation for the probability distribution of pain over time  $\rho(P, t)$ :

$$\begin{aligned}
 \frac{\partial \rho}{\partial t} &= -\frac{\partial}{\partial P} \left[ \left( -(k_0 + k_1 D_1 + k_2 D_2 + k_3 D_3)P + \right. \right. \\
 &\quad \left. \left. k_0 u \right) \rho(P, t) \right] + \frac{\partial^2}{\partial P^2} \left[ \frac{1}{2} (\varepsilon k_0 u)^2 \rho(P, t) \right].
 \end{aligned} \tag{3}$$

Absent any pain medication, this Fokker-Planck equation implies the steady-state pain distribution

$$\rho^*(P) = \sqrt{\frac{1}{\pi k_0 u^2 \varepsilon^2}} \exp \left[ -\frac{(P - u)^2}{k_0 u^2 \varepsilon^2} \right], \tag{4}$$

a Gaussian distribution with mean  $u$  and standard deviation  $u\varepsilon\sqrt{k_0}/2$ . See Figure 4 for a sample stochastic model output.

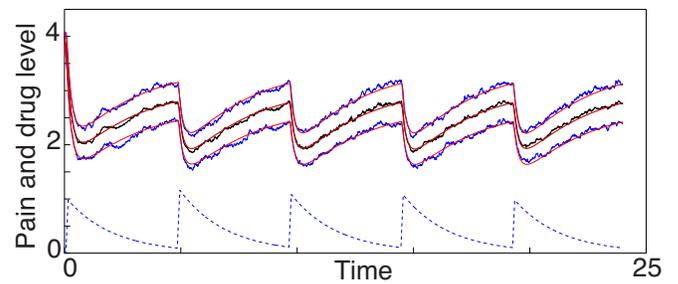


FIG. 4. **Sample output from stochastic differential equation model (2).** Red thick line: theoretical mean pain; red thin lines:  $\pm$  one theoretical standard deviation; black thick line: mean of pain distribution in ensemble of 100 stochastic simulations; blue thin lines:  $\pm$  one standard deviation in ensemble of 100 stochastic simulations; blue dashed line: drug 1 dose in bloodstream. Spikes occur when patient takes recommended dosage.

*Statistical component (for all patients)*

In order to account for the variation among patients and improve prediction of the unmitigated pain level, we associate patient characteristics and history with the unmitigated pain level  $u$  (an  $n$ -dimensional vector with  $u_j$  corresponding to the  $j$ -th patient's unmitigated pain level) using a linear model.

Let  $X$  be an  $n \times p$  design matrix containing the covariates of patients (i.e., patient characteristics). We write  $X = (X_1, \dots, X_n)^T$ , with  $X_j$  corresponding to the  $p$ -dimensional covariates of patient  $j$ . Then we formulate the relationship between between  $u$  and the  $p$  predictors as:

$$u = X\beta + \epsilon, \tag{5}$$

where  $\beta$  is a  $p$ -dimensional coefficient vector, and  $\epsilon$  is an  $n$ -dimensional vector of zero-mean random errors. When

$p$  is small, the estimate for  $\beta$  is obtained using the ordinary least squares procedure:  $\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \|u - X\beta\|_2^2$ ,

where  $\|\cdot\|_q$  denotes the  $\ell_q$  norm. Then the unmitigated pain level  $u_j$  is updated by  $u_j^{new} = X_j^T \hat{\beta}$ ,  $j = 1, \dots, n$ .

Since the unmitigated pain levels are not observable from patient pain reports, the initial  $u_j$ 's are independently sampled from a uniform distribution between 0 and 10, i.e.  $u_j^0 \sim U(0, 10)$ . After using  $\{u_j^0\}$  as initial values to fit the mechanistic model, the resulting estimated  $\{u_j\}$  will be updated by the linear model (5) as  $\{u_j^{new}\}$ , which will then be used as initial values in the next round of fitting of the mechanistic model. See Section for more detail on the hybridization of the statistical component with the mechanistic component.

Given a high-dimensional set of patient characteristics, we need to select a subset of patient characteristics that are significantly associated with  $u$  by minimizing the penalized loss function. In this study, we select patient characteristics using the LASSO (Least Absolute Shrinkage and Selection Operator) [16], by minimizing the penalized loss function  $\Gamma(\beta) = \|u - X\beta\|_2^2 + \lambda\|\beta\|_1$  with respect to  $\beta$ . The penalty parameter  $\lambda$  is determined using 5-fold cross-validation. The selected  $p$  features are then used to fit the linear model (5) by ordinary least squares.

If time-varying unmitigated pain levels and time-varying covariates are present, the regression model (5) can be extended to the linear mixed model [17, 18]:  $u = X\beta + Z\delta + \epsilon$ , where  $Z = (Z_1, \dots, Z_n)^T$  is an  $n \times r$  design matrix for  $r$  random effect factors and  $\delta = (\delta_1, \dots, \delta_r)^T$  is a vector of random effects. Patient characteristics can be selected by maximizing the penalized log-likelihood:  $\ell_{\text{pen}}(\beta, \delta) = \ell(\beta, \delta) - \lambda\|\beta\|_1$  [19]. Such an extension of the proposed hybrid model to allow for time-varying unmitigated pain levels and covariates will be considered in a future study with more data available.

### Model fitting

We fit our model to real patient data by minimizing the residual sum of squares between model predictions and patient reports provided within the first two weeks of reporting. We expect that the assumption of constant model parameters breaks down after approximately two weeks (clinical heuristics). Minimization over parameters  $u$ ,  $k_1$ ,  $k_2$ , and  $k_3$  was done via the Nelder-Mead simplex algorithm [20]. Parameter  $k_0$  was fixed at  $2\ln(2) \approx 1.4$  corresponding to a pain equilibration half-life time scale of 30 minutes in the absence of medication. If a patient did not take all three classes of drugs, the model and fitting only included the consumed drugs.

We initialize the parameter optimization in  $n$  mechanistic models (one per patient) with random values during a first iteration, then we feed the optimization output

into the statistical model (for all patients). Once the statistical model is run, it results in a new set of parameter estimates that can then be employed as initial parameter seeds for a second round of optimization in  $n$  mechanistic models (to minimize the residual sum of squares). Proceeding iteratively in this fashion (see Fig. 3), we find convergence to a consistent set of parameters for each patient (details below).

### Method verification

Before applying our hybrid model to real-world patient data, we verify the soundness of the approach with synthetic data constructed to resemble real-world data, but generated by the model itself with high sampling frequency. The synthetic data used for verification of the method are generated directly from the mechanistic model with an assumed parameter set generated in the following way: unmitigated pain  $u = (\text{patient age})/10$ , initial pain level  $P(t=0) = u - 2$ , and drug parameters  $k_1, k_2, k_3 \sim N(0.75, 0.25)$ . Each patient reports pain every 1/2 hour for 336 hours (two weeks). At each report time, the probability of the patient taking a particular drug (among three drug classes) is 1/16; in other words, the patients took each drug on average every 8 hours. White noise of magnitude 1 is added to each pain report.

As an illustration using *real* patient drug times (specifically those of Patient A3), we create synthetic data generated using  $u = 5$ ,  $k_1 = 3$ , and  $k_2 = 2$ : see Figure 5. When the initial parameter search is seeded with random parameter values, the mechanistic model fit can lead to convergence to either the ‘‘true’’ optimum (5, 3, 2) or to other ‘‘spurious’’ optima with incorrect values of  $u$ ,  $k_1$ , and  $k_2$ .

In this illustrative example, the method converges to  $u = 5.01$ ,  $k_1 = 3.19$ , and  $k_2 = 1.84$ . The relevant root-mean-square (RMS) error is 1.01; this is close to the lowest possible expected error given the unit magnitude noise added to the synthetic data. This numerical experiment shows that the mechanistic model fitting method can converge even in the presence of significant amounts of noise. However, with *only* the mechanistic model it can be quite difficult to find a good set of initial parameter seeds<sup>1</sup>: that is one motivation for introducing the statistical model.

To test our hybrid method using both the mechanistic model for fitting and the statistical model for parameter estimation, we create a synthetic patient database of 39 patients as described above. We then iterate rounds of fitting between mechanistic and statistical models, start-

<sup>1</sup> The seeding problem becomes exponentially harder as the dimension of the parameter space increases.

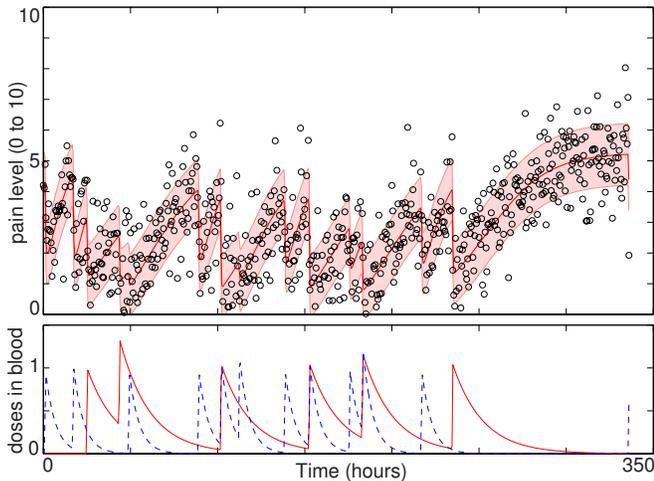


FIG. 5. **Model fitting demonstration for densely reported noisy synthetic data.** Upper panel: hypothetical densely-reported patient pain (black circles) and model fit (red solid line); red shading indicates model fit plus/minus one standard deviation. Lower panel: long-acting opioid (red solid line) and short-acting opioid (blue dashed line) medication concentration in patient bloodstream.

ing with uniform random guesses for all patient parameters ( $u, k_1, k_2, k_3$ ). Figure 6 demonstrates how the parameter  $u$  converges to a value with small error after just a few iterations steps, even in the presence of significant noise. In order to evaluate the performance of the model on new data, we use the hold-out validation method by splitting the dataset into a training set (first week) and a test set (second week). Model fit error and hold-out validation error, as well as other parameters values, converge similarly: see Figure 7.

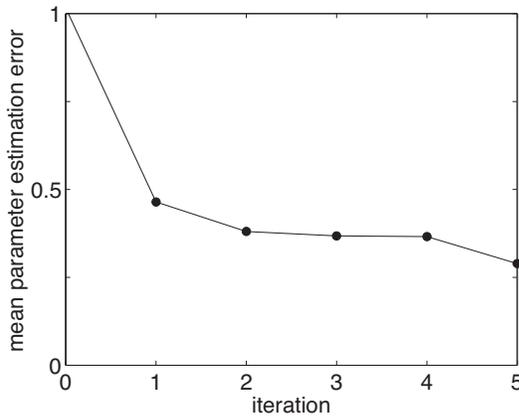


FIG. 6. **Hybrid model fitting demonstration for ensemble of densely reported noisy synthetic data.** For an ensemble of 39 synthetic patient data sets, the average absolute error in  $u$  gradually decreases. Iteration 0 indicates one fit to the mechanistic model alone. Subsequent iterations indicate the number of hybrid model (statistical + mechanistic) fits.

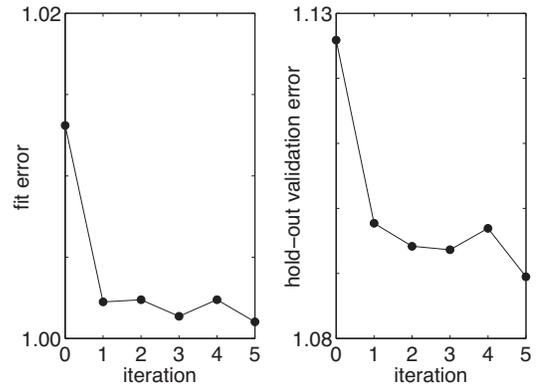


FIG. 7. **Hybrid model fitting demonstration for ensemble of densely reported noisy synthetic data.** For an ensemble of 39 synthetic patient data sets, the average root-mean-squared (RMS) error in patient pain levels gradually decreases. Iteration 0 indicates one fit to the mechanistic model alone. Subsequent iterations indicate the number of hybrid model (statistical+mechanistic) fits. Training error (or fit error) is on the left; test error (or validation error) is on the right. Due to the additive white noise of magnitude 1, the smallest testing or training error we could expect is 1.

## RESULTS

### General results

One key result is that our model for chronic pain does indeed have some predictive value (see Fig. 8). This is an improvement over the current state of the art, since no other predictive model exists of which we are aware. Furthermore, fitted parameter values correlate significantly with patient characteristics, suggesting that meaningful information is captured by this minimal plausible model. It may be possible to motivate new clinical insight on the basis of the observed correlations, perhaps leading to differential treatment of SCD sufferers with differing characteristics.

### Statistical results

We use the following baseline patient characteristics to predict the unmitigated pain levels in the statistical modeling step: age, gender, SCD disease type, hydroxyurea use, folic acid vitamin use, long-acting opioid use, short-acting opioid use, and non-opioid use. We explore the marginal effects of these characteristics and their possible pairwise two-way interactions using the LASSO. The model (5) can be extended to include time-varying covariates such as temperature, weather, patient’s walking/social activities, and patient’s mood at time  $t$ , once these data become available in a future study.

The statistical model that resulted from the LASSO

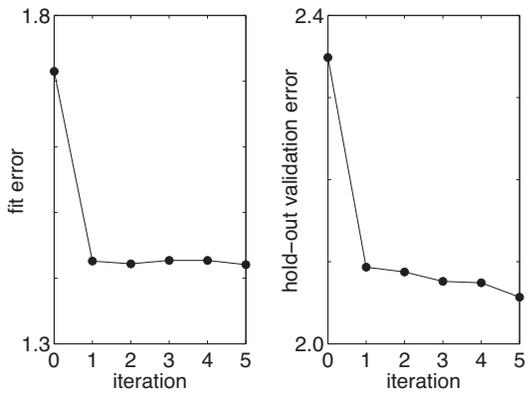


FIG. 8. **Hybrid model fitting on real patient data.** For an ensemble of 39 real patient data sets, the average root-mean-squared (RMS) error in patient pain levels gradually decreases. Iteration 0 indicates one fit to the mechanistic model alone. Subsequent iterations indicate the number of hybrid model (statistical+mechanistic) fits. Training error (or fit error) is on the left; test error (or validation error) is on the right.

variable selection is given by

$$\begin{aligned}
 u = & \beta_0 + \beta_1 \mathbf{1}\{\text{SCD disease type} = \text{HgbSC}\} \\
 & + \beta_2 \mathbf{1}\{\text{SCD disease type} = \text{SB+Thal or SO-Ara}\} \\
 & + \beta_3 (\text{age} - 18) + \beta_4 \mathbf{1}\{\text{Hydroxyurea user}\} \\
 & + \beta_5 \mathbf{1}\{\text{Non-opioid user}\} \\
 & + \beta_6 \mathbf{1}\{\text{SCD disease type} = \text{SB+Thal or So-Ara}\} \\
 & \times (\text{age} - 18) + \epsilon,
 \end{aligned} \tag{6}$$

where  $\epsilon_j \sim N(0, \sigma^2)$ ,  $j = 1, \dots, n$ , and  $\mathbf{1}\{\cdot\}$  is the indicator function.

Table IV summarizes the results from one round of fitting of the regression model (6). Adjusting for the effect of other terms in the regression model, SCD disease type of SB+Thal or So-Ara (with coefficient  $\beta_2$ ), non-opioid use (with coefficient  $\beta_5$ ), and the interaction term between SCD disease type of SB+Thal or So-Ara and age (with coefficient  $\beta_6$ ) are important predictors of the unmitigated pain levels at the significant level of 0.05. Using non-opioid medication is associated with decreased unmitigated pain levels. Unmitigated pain levels increase with patients' age for SB+Thal or So-Ara patients.

### Mechanistic model validation

With such sparse data and up to four fitting parameters, one may worry that the model (1) is being overfitted. To test this concern, we propose 6 related alternative models with fewer fitting parameters, and we compare

cross-validation error and Akaike information criterion (AIC) among the models. See Table V for model descriptions. Neither measure selected a best-fit model across all patients, but none of these simple models is overfitting the data. See Figure 9 for AIC results and Figure 10 for cross validation results.

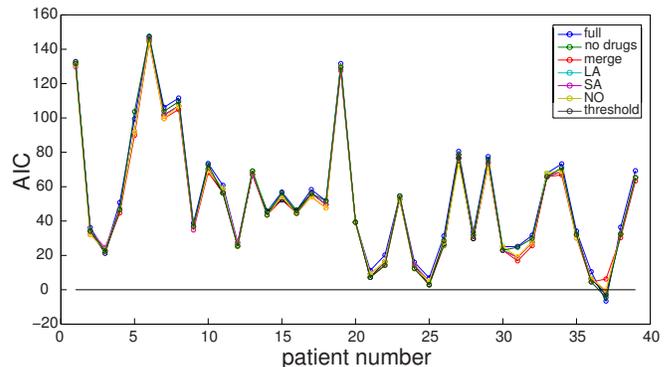


FIG. 9. Akaike information criterion (AIC) for alternative models listed in Table V. Most models perform equally well; among patients with differing model performance, there exists no clear ‘best’ model for all patients.

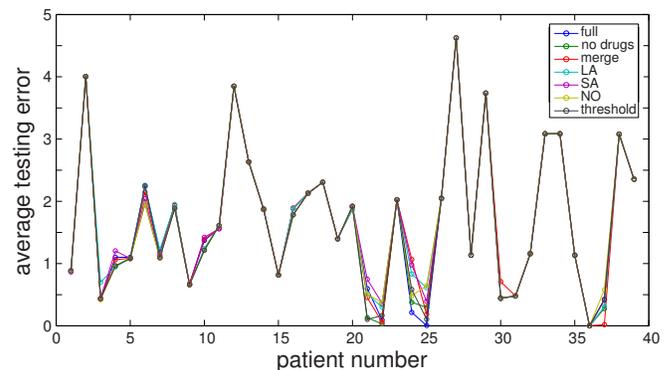


FIG. 10. Two-fold cross validation testing error for alternative models listed in Table V. For every patient, each model was independently fitted to the first and second half of the time series pain report data (training). Then the fitted models were used to test the other half of the data. This figure shows the average root-mean-square testing error for the two tests. Most models perform equally well; among patients with differing model performance, there exists no clear ‘best’ model for all patients. Note that patient #36 did not have enough data to fit any models, so zero error is misleading.

### BIASED PAIN REPORTING

Perhaps the most significant limitation of our model lies in a potential bias in our data set. Patients typically

Variable	Estimate	Std Err	T-value	P-value
Intercept	7.646	1.228	6.228	0.000 ***
HgbSC	-1.566	0.890	-1.761	0.088
SB+Thal or So-Ara	-5.479	2.332	-2.349	0.025 *
Age at baseline -18	0.001	0.034	0.290	0.773
Hydroxyurea user	-1.205	0.839	-1.437	0.160
Non-opioid user	-2.523	0.842	-2.995	0.005 **
(SB+Thal or So-Ara) $\times$ (Age at baseline -18)	0.241	0.010	2.419	0.021 *

TABLE IV. Result of the prediction model of the unmitigated pain using the linear regression model. Significance \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ )

Model name	Description	Fitting parameters
Full model	Include all drugs taken (model (1))	up to 4
No drugs	Include no drug dosing information	1
Merge drugs	Combine all drugs into one drug class with same response	up to 2
LA only	Include only long-acting opioid doses	up to 2
SA only	Include only short-acting opioid doses	up to 2
NO only	Include only non-opioid doses	up to 2
Threshold	Include drug class only if drug is taken at least $n$ times*	up to 4

TABLE V. Mechanistic model variations. Fitting parameters include unmitigated pain level  $u$  and drug response parameters  $k_i$  for all drugs consumed. Therefore some patients have fewer fitting parameters than listed if they consumed fewer than three types of drugs. \*In our tests,  $n = 5$  was the drug dose threshold.

report pain levels when taking medication, but many of them only take medication when pain levels rise. Thus we suspect a selection bias of unknown significance, causing higher pain levels to be reported at a disproportionately high rate. To test this concern, we compare the unbiased model (1) with a similar model incorporating biased pain reporting.

Suppose the probability density function of pain at a particular time is a normal distribution with mean  $\mu$  and variance  $\sigma^2$ :

$$\rho(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right]. \quad (7)$$

Integrating model (1) gives the expected pain value  $\mu$  at any point in time.

If higher pain is disproportionately reported through the mobile health application, then we will be much more likely to see higher pain levels from this normal distribution. As a first approximation, we assume the reporting bias is linear:

$$\rho_r(x | \mu, \sigma) = \alpha(ax + b) \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right] H(x), \quad (8)$$

where  $\alpha$  normalizes the distribution,  $a, b$  tune the probabilities of reporting a pain value  $x$ , and the Heaviside function  $H(x)$  prevents negative pain values. Figure 11 shows both real and reported pain distributions at a particular time.

We need a way to connect these distributions because we want to control real pain described by (7), but the patient only provides data from the reported distribution (8). In other words, real pain is important but invisible,

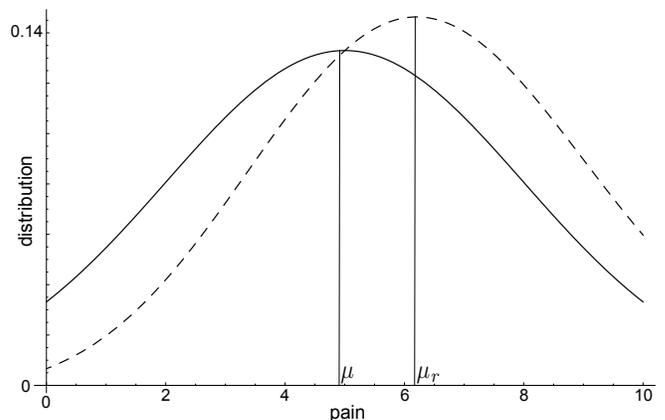


FIG. 11. Probability density distributions for unbiased (solid) and biased (dashed) pain reporting at a particular time  $t_i$ . The standard deviation (here,  $\sigma = 3$ ) has been exaggerated for illustrative purposes. In real data, the typical standard deviation is around  $\sigma = 1.8$ .

and reported pain is unimportant but visible. One way to connect the distributions is through their means and variances<sup>2</sup>:

$$\mu_r(\mu, \sigma) = \int_{-\infty}^{\infty} x \rho_r(x | \mu, \sigma) dx \quad (9)$$

$$\sigma_r^2(\mu, \sigma) = \int_{-\infty}^{\infty} x^2 \rho_r(x | \mu, \sigma) dx. \quad (10)$$

<sup>2</sup> Note that the means and variances also change in time. We omit time dependence for notational clarity.

Assuming  $\sigma_r$  is approximately constant in time<sup>3</sup>, we can also estimate  $\sigma_r^2$  using the definition of variance:

$$\sigma_r^2 = \frac{1}{M} \sum_{i=1}^M (P_i - \mu_r(t_i))^2, \quad (11)$$

where  $M$  is the number of pain reports,  $P_i$  is the  $i$ th reported pain value, and  $\mu_r(t_i)$  is the expected reported pain given distribution (8) at time  $t_i$ .

Given a proposed model for real pain, we can solve this system of three equations for the three unknowns:  $\mu_r(t_i)$ ,  $\sigma(t_i)$ , and  $\sigma_r$ . We can then compute the likelihood of the reported pain using

$$\mathcal{L} = \prod_{i=1}^M \rho_r(P_i | \mu(t_i), \sigma(t_i)). \quad (12)$$

Because we can compute the likelihood of the supplied data given any proposed model for real pain, we can tune the model parameters to maximize likelihood (technically, we minimize the negative logarithm of likelihood). This results in a best-fit model under the assumption of biased pain reporting. We compare the best-fits of the model under both biased and unbiased reporting assumptions, and find that neither model is a better fit for most patients. See Figure 12.

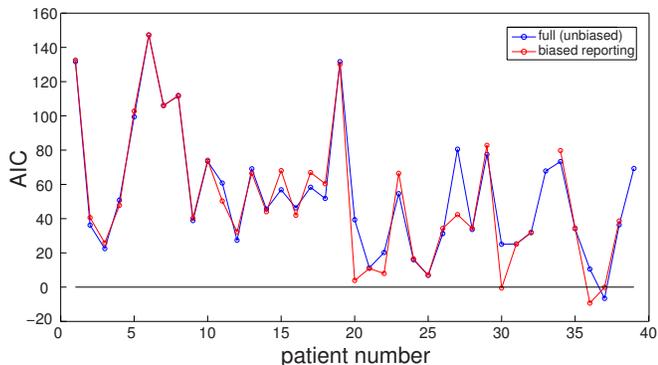


FIG. 12. Akaike information criterion for unbiased and biased pain reporting. Because both models have an equal number of fitting parameters, AIC is a proxy for model likelihood (lower AIC implies higher likelihood). Again, it is not clear that one model performs universally better than the other. Note that missing biased reporting model fits indicate that the fitting algorithm did not converge ( $a = 0.08, b = 0.1$ ).

<sup>3</sup> It is not possible to verify this with data because patients only report one pain value at a particular time. However, a Kolmogorov-Smirnov normality test on residuals over the first two weeks of data rejects normality ( $p < 0.05$ ) for only 2 of the 39 patients.

## PAIN AND MEDICATION OPTIMIZATION

A key goal of the modeling of human pain dynamics is to develop predictions that allow optimized treatment: both pain and medication use should be minimized. Excess medication carries particular long-term risks for chronic pain sufferers [21–24], but pain mitigation is also a primary goal of SCD treatment. How can these contradictory objectives be balanced?

Our model allows us to forecast the probability distribution of pain for a patient at a point in the near future, given past data and future drug dosage protocol. This information may be useful to a physician, allowing him or her to make an optimized, data-driven decision balancing medication and pain for the patient in real time.

We propose several tools that may be useful to a physician. First, we find the optimal drug timing given that a certain amount of each drug will be taken over a certain time period (say, within 24 hours). For instance, if the patient will take two short-acting opioids and one long-acting opioid within 24 hours, then the algorithm will offer the best times to take those three drug doses in order to minimize the expected average pain. We provide the physician with the expected optimal average pain for all drug combinations up to a certain maximum number of safe drug doses. See Figure 13.

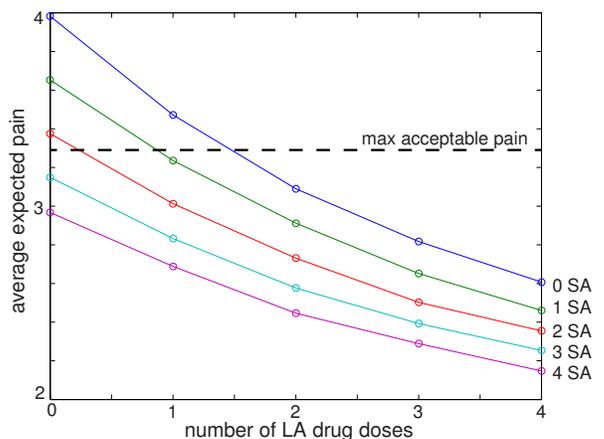


FIG. 13. Example expected pain given optimal drug dosage protocol (in this case, for Patient A3). For each set of drug dosage protocols, from no drugs to 4 doses of each drug, a physician can see the expected average pain over a certain time period. Given a patient’s maximum acceptable pain level, the physician can select the best compromise between drug doses and expected pain. In this case, the physician may tell the patient to take no long-acting (LA) drugs but take 3-4 short-acting (SA) drugs. Alternatively, the physician might tell the patient to take one LA and 1-2 doses of SA medication. The timing of those LA and SA drugs is provided by the optimization algorithm.

Second, we select the best drug dosing protocol (both number of drugs and dose timings) given an objective

function balancing pain and medication. There are an unlimited number of possible objective functions that balance pain and medication, but we propose the following:

$$\begin{aligned}
 m(\bar{P}, d1, d2, d3) = & w_P \left( (P_{\max} - \bar{P})^{-\alpha} - P_{\max}^{-\alpha} \right) \\
 & + w_{d1} \left( (d1_{\max} - d1)^{-\beta} - d1_{\max}^{-\beta} \right) \\
 & + w_{d2} \left( (d2_{\max} - d2)^{-\gamma} - d2_{\max}^{-\gamma} \right) \\
 & + w_{d3} \left( (d3_{\max} - d3)^{-\eta} - d3_{\max}^{-\eta} \right),
 \end{aligned} \tag{13}$$

where  $\bar{P}$  is the average expected pain;  $\{d1, d2, d3\}$  are the number of drug doses of each type;  $\{w_P, w_{d1}, w_{d2}, w_{d3}\}$  are the weights of pain and drugs;  $\{P_{\max}, d1_{\max}, d2_{\max}, d3_{\max}\}$  are the maximum safe levels of pain and drugs; and  $\{\alpha, \beta, \gamma, \eta\}$  tune the steepness of the objective function near those dangerous levels.

See Figure 14 for the contributions of the pain and one drug component to the objective function  $m$ . The contribution to the objective function is zero if no pain exists or if no drugs are taken. As pain or drug doses approach dangerous levels, the contribution to the objective function blows up. After a physician has made sufficient recommendations to a patient, a machine learning algorithm could select the weight parameters for each physician/patient pair. At that point, the algorithm could propose the optimal dosing protocol without much effort on the physician’s part. See Figure 15.

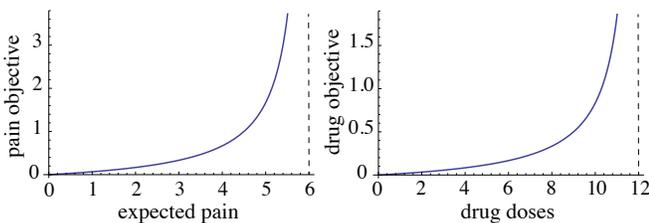


FIG. 14. Contribution of the pain and one drug component to the objective function (13). The contribution to the objective function is 0 if no pain exists or if no drugs are taken. As pain or drug doses approach dangerous levels, the contribution to the objective function blows up.

## DISCUSSION AND LIMITATIONS

### Reflection on hybrid modeling

Statistical models and mechanistic models have both been successfully applied to various aspects of human behavior. The inference of “black box” statistical models from empirical data has the advantage that it obviates the need for a-priori knowledge of system dynamics. However, mechanistic models (sometimes referred to as

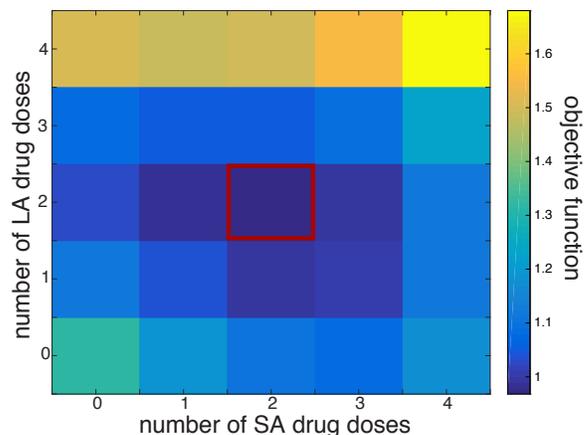


FIG. 15. Example patient intervention recommendation (in this case, for Patient A3). For a set of personalized optimization parameters (selected by a physician or machine learning algorithm), the optimal drug timing minimizes the objective function (13) for each number of drug doses per 24-hour period. In this case, the patient is advised to take two standard doses of the long-acting (LA) opioid and two standard doses of the short-acting (SA) opioid, indicated by the red box.

“white box” or “clear box”) can easily incorporate such knowledge when available.

Perhaps because of the often distinct educational backgrounds of practitioners or distinct typical applications, statistical and mechanistic approaches are not frequently combined in addressing a single problem. Compared with our work, the most similar hybrid modeling idea was developed by Sheiner and colleagues in the field of pharmacokinetics, where they proposed models to estimate population characteristics of pharmacokinetic parameters [25–27]. In their work, the pharmacokinetic models (i.e., mechanistic models) are well established, and the novelty and focus was the introduction of statistical models for pharmacokinetic parameter estimation. On the contrary, in our study the mechanistic model is not known before but developed by us based on clinical knowledge and reasonable assumptions, and our focus is the prediction of pain levels rather than parameter estimation.

Other attempts based on the hybrid modeling idea in the scientific literature have appeared in the context of neural networks (e.g., [28–30]) and chemical engineering (e.g., [29, 31, 32]), where they largely played a computational rather than analytical role. Some attempts have also been made with medical applications: Rosenberg et al. ([33]) and Adams et al. ([34]) developed a model by combining a dynamical systems approach with a statistical model to predict a patient’s CD4 cell counts and HIV viral load over time in an HIV study. Timms et al. ([35]) proposed a dynamical systems approach using ODEs to improve self-regulation in a smoking cessation study. Reinforcement learning techniques such as Q-learning (e.g.,

[36]) also share some commonalities with the hybrid approach.

In this work we make our own attempt at a novel incorporation of statistical inference together with mechanistic dynamical systems modeling to produce a hybrid mathematical model for predicting and explaining human behavior. We apply the new approach specifically to the problem of predicting the dynamics of subjective pain in a population of individuals suffering from sickle cell disease. The rationale behind our method development is that we have prior knowledge of pain trajectories with medication, making the problem suitable for mechanistic modeling; meanwhile, we do not know the relationship between patient health characteristics and pain levels, so we would like to investigate this using a statistical model.

### Limitations

The hybrid dynamical systems/statistical approach appears to have great potential. The low frequency of pain reporting currently limits its usefulness, but future addition of high-frequency pain correlates like blood pressures, heart rate, activity level, etc., via wearable medical devices (e.g. the “Fitbit”) may drastically improve on that. Furthermore, application of similar methods to more data-rich forecasting problems (e.g. insulin levels) may also expand the utility of our work.

Another important limitation to our current model lies in the mechanistic component. We presented here what we considered to be the simplest plausible model: pain fluctuates about an “unmitigated” equilibrium  $u$ , and medication reduces pain below that level, but pain returns as medication is metabolized and removed from the bloodstream. This simple model cannot capture long-term changes in the unmitigated pain level, and hence its forecast validity is likely limited to short time scales (days to weeks).

Society is clearly moving in the direction of an overwhelming amount of medical data. It is imperative that we learn to take advantage of this information to improve patient treatments beyond the traditional standard of care. The approach we report here not only addresses the specific challenge of chronic pain mitigation in SCD patients, but also provides a testbed for new ways of dealing with big, ever-growing data sets in real time.

### CONCLUSIONS

We have successfully demonstrated the hybrid application of statistical and mechanistic mathematical modeling with application to understanding the dynamics of subjective human pain. Our model explains real-world data on human pain and can generate predictions of future pain dynamics.

We expect that similar methods could be used to incorporate disease-specific knowledge and modeling with statistical inference in a variety of medical applications. Given the coming deluge of data from wearables (including clinical trial NCT02895841 already underway) and mobile health applications, there is a clear need for new mathematical methods to take advantage of the opportunity for personalizable, data-driven medical treatments.

### DATA ACCESSIBILITY

Data and code available at Northwestern University ARCH repository: <https://doi.org/10.21985/N24S3T> [37]. Real patient data is not included to protect patient privacy, but all methods may be tested on synthetic data.

### FUNDING

The authors gratefully acknowledge NSF support through grants No. DMS-1557727 and No. DGE-1324585.

---

\* E-mail me at: [scifton@u.northwestern.edu](mailto:scifton@u.northwestern.edu)

- [1] S. H. Strogatz, *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press, 2014.
- [2] D. A. Freedman, *Statistical models: theory and practice*. cambridge university press, 2009.
- [3] O. S. Platt, B. D. Thorington, D. J. Brambilla, P. F. Milner, W. F. Rosse, E. Vichinsky, and T. R. Kinney, “Pain in sickle cell disease: rates and risk factors,” *New England Journal of Medicine*, vol. 325, no. 1, pp. 11–16, 1991.
- [4] Y. Granovsky, M. Granot, R.-R. Nir, and D. Yarnitsky, “Objective correlate of subjective pain perception by contact heat-evoked potentials,” *The Journal of Pain*, vol. 9, no. 1, pp. 53–63, 2008.
- [5] A. Hughes, A. Macleod, J. Growcott, and I. Thomas, “Assessment of the reproducibility of intradermal administration of capsaicin as a model for inducing human pain,” *Pain*, vol. 99, no. 1, pp. 323–331, 2002.
- [6] S. Stevens, A. Carton, and G. Shickman, “A scale of apparent intensity of electric shock.,” *Journal of Experimental Psychology*, vol. 56, no. 4, p. 328, 1958.
- [7] C. R. Jonassaint, N. Shah, J. Jonassaint, and L. De Castro, “Usability and feasibility of an mhealth intervention for monitoring and managing pain symptoms in sickle cell disease: The sickle cell disease mobile application to record symptoms via technology (smart),” *Hemoglobin*, no. ahead-of-print, pp. 1–7, 2015.
- [8] N. Shah, J. Jonassaint, and L. De Castro, “Patients welcome the sickle cell disease mobile application to record symptoms via technology (smart),” *Hemoglobin*, vol. 38, no. 2, pp. 99–103, 2014.

- [9] D. T. McRuer and E. S. Krendel, "Mathematical models of human pilot behavior," tech. rep., DTIC Document, 1974.
- [10] U. G. Fors, L. G. Edwall, and G. A. Haegerstam, "The ability of a mathematical model to evaluate the effects of two pain modulating procedures on pulpal pain in man," *Pain*, vol. 33, no. 2, pp. 253–264, 1988.
- [11] N. Britton, S. Skevington, and M. Chaplain, "Mathematical modelling of acute pain," *Journal of Biological Systems*, vol. 3, no. 04, pp. 1119–1124, 1995.
- [12] G. Stepan, "Delay effects in the human sensory system during balancing," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1891, pp. 1195–1212, 2009.
- [13] C. W. Gardiner *et al.*, *Handbook of stochastic methods*, vol. 3. Springer Berlin, 1985.
- [14] F. Yang, X. Tong, D. G. McCarver, R. N. Hines, and D. A. Beard, "Population-based analysis of methadone distribution and metabolism using an age-dependent physiologically based pharmacokinetic model," *Journal of pharmacokinetics and pharmacodynamics*, vol. 33, no. 4, pp. 485–518, 2006.
- [15] P. Poulin and F.-P. Theil, "Prediction of pharmacokinetics prior to in vivo studies. ii. generic physiologically based pharmacokinetic models of drug disposition," *Journal of pharmaceutical sciences*, vol. 91, no. 5, pp. 1358–1370, 2002.
- [16] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [17] P. Diggle, *Analysis of longitudinal data*. Oxford University Press, 2002.
- [18] G. M. Fitzmaurice, N. M. Laird, and J. H. Ware, *Applied longitudinal analysis*, vol. 998. John Wiley & Sons, 2012.
- [19] A. Groll and G. Tutz, "Variable selection for generalized linear mixed models by l1-penalized estimation," *Statistics and Computing*, vol. 24, no. 2, pp. 137–154, 2014.
- [20] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.
- [21] B. Bannwarth, "Risk-benefit assessment of opioids in chronic noncancer pain," *Drug safety*, vol. 21, no. 4, pp. 283–296, 1999.
- [22] R. J. Gatchel, "A biopsychosocial overview of pretreatment screening of patients with pain," *The Clinical journal of pain*, vol. 17, no. 3, pp. 192–199, 2001.
- [23] S. Savage, "Opioid therapy of chronic pain: assessment of consequences," *Acta Anaesthesiologica Scandinavica*, vol. 43, no. 9, pp. 909–917, 1999.
- [24] D. Brookoff, "Chronic pain: 2. the case for opioids," *Hospital Practice*, vol. 35, no. 9, pp. 69–84, 2000.
- [25] L. B. Sheiner, B. Rosenberg, and V. V. Marathe, "Estimation of population characteristics of pharmacokinetic parameters from routine clinical data," *Journal of pharmacokinetics and biopharmaceutics*, vol. 5, no. 5, pp. 445–479, 1977.
- [26] L. B. Sheiner and S. L. Beal, "Evaluation of methods for estimating population pharmacokinetic parameters. i. michaelis-menten model: routine clinical pharmacokinetic data," *Journal of pharmacokinetics and biopharmaceutics*, vol. 8, no. 6, pp. 553–571, 1980.
- [27] J. W. Mandema, D. Verotta, and L. B. Sheiner, "Building population pharmacokinetic-pharmacodynamic models. i. models for covariate effects," *Journal of pharmacokinetics and biopharmaceutics*, vol. 20, no. 5, pp. 511–528, 1992.
- [28] D. C. Psychogios and L. H. Ungar, "A hybrid neural network-first principles approach to process modeling," *AIChE Journal*, vol. 38, no. 10, pp. 1499–1511, 1992.
- [29] M. L. Thompson and M. A. Kramer, "Modeling chemical processes using prior knowledge and neural networks," *AIChE Journal*, vol. 40, no. 8, pp. 1328–1340, 1994.
- [30] H.-T. Su, N. Bhat, P. Minderman, and T. McAvoy, "Integrating neural networks with first principles models for dynamic modeling," in *Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes (DYCORD+'92): Selected Papers from the 3rd IFAC Symposium, Maryland, USA, 26-29 April 1992*, p. 327, Elsevier, 2014.
- [31] J. Schubert, R. Simutis, M. Dors, I. Havlík, and A. Lübbert, "Hybrid modelling of yeast production processes—combination of a priori knowledge on different levels of sophistication," *Chemical engineering & technology*, vol. 17, no. 1, pp. 10–20, 1994.
- [32] B. P. Duarte and P. M. Saraiva, "Hybrid models combining mechanistic models with adaptive regression splines and local stepwise regression," *Industrial & engineering chemistry research*, vol. 42, no. 1, pp. 99–107, 2003.
- [33] E. S. Rosenberg, M. Davidian, and H. T. Banks, "Using mathematical modeling and control to develop structured treatment interruption strategies for hiv infection," *Drug and alcohol dependence*, vol. 88, pp. S41–S51, 2007.
- [34] B. Adams, H. Banks, M. Davidian, and E. Rosenberg, "Estimation and prediction with hiv-treatment interruption data," *Bulletin of mathematical biology*, vol. 69, no. 2, pp. 563–584, 2007.
- [35] K. P. Timms, D. E. Rivera, L. M. Collins, and M. E. Piper, "A dynamical systems approach to understanding self-regulation in smoking cessation behavior change," *nicotine & tobacco research*, vol. 16, no. Suppl 2, pp. S159–S168, 2014.
- [36] L. G. Jaimes, M. Llofriú, and A. Raij, "A stress-free life: just-in-time interventions for stress via real-time forecasting and intervention adaptation," in *Proceedings of the 9th International Conference on Body Area Networks*, pp. 197–203, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014.
- [37] S. M. Clifton, C. Kang, J. J. Li, Q. Long, N. Shah, and D. M. Abrams, "Data from: Hybrid statistical and mechanistic mathematical model guides mobile health intervention for chronic pain."