A Machine-Learning Based Drug Repurposing Approach Using Baseline Regularization

Zhaobin Kuang*, zkuang@wisc.edu, The University of Wisconsin, Madison
Yujia Bao, yujia@csail.mit.edu, The Massachusetts Institute of Technology
James Thomson, JThomson@morgridge.org, The Morgridge Institute for Research
Michael Caldwell, caldwell.michael@marshfieldclinic.org, The Marshfield Clinic
Peggy Peissig, Peissig.Peggy@mcrf.mfldclin.edu, The Marshfield Clinic
Ron Stewart, RStewart@morgridgeinstitute.org, The Morgridge Institute for Research
Rebecca Willett, willett@discovery.wisc.edu, The University of Wisconsin, Madison
David Page, page@biostat.wisc.edu, The University of Wisconsin, Madison

Abstract

We present the baseline regularization model for computational drug repurposing using electronic health records (EHRs). In EHRs, drug prescriptions of various drugs are recorded throughout time for various patients. In the same time, numeric physical measurements (e.g. fasting blood sugar level) are also recorded. Baseline regularization uses statistical relationships between the occurrences of prescriptions of some particular drugs and the increase or the decrease in the values of some particular numeric physical measurements to identify potential repurposing opportunities.

Keywords

Electronic health records, computational drug repurposing, longitudinal data, self-controlled case series, silico repurposing.

Running Head

Baseline Regularization for Drug Repurposing

1 Introduction

With the increasing availability of electronic health record (EHR) data, there is an emerging interest in using EHRs from various patients for computational drug repurposing (CDR). Specifically, in EHRs, drug prescriptions of various drugs are recorded throughout time for various patients. In the same time, numeric physical measurements, such as fasting blood sugar (FBG) level, blood pressure, and low density lipoprotein are also recorded. By designing machine learning algorithms

*corresponding author

that can establish relationships between the occurrences of prescriptions of some particular drugs and the increase or the decrease in the values of some particular numeric physical measurements, we might be able to identify drugs that can be potentially repurposed to control certain numeric physical measurements. This chapter describes such a machine learning algorithm called *baseline regularization* [12] for CDR.

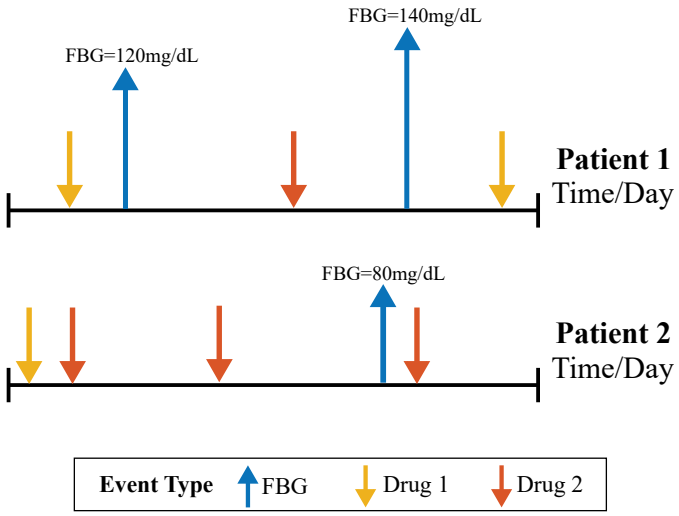

Figure 1: Visualization of electronic health records (EHRs) from two patients. Fasting blood sugar (FBG) level measurements as well as drug prescriptions of various drugs are observed for the two patients over time.

2 Materials

Figure 1 visualizes a set of electronic health records from two patients. Drug prescriptions of different types enter the EHRs of the two patients at different times. Fasting blood sugar (FBG) level measurements are also recorded at various times. In this chapter, we will consider how to identify drugs that can be potentially repurposed to control FBG level as an example to illustrate the use of baseline regularization. The idea is to formulate this problem as a machine learning problem by considering an FBG record as a response variable and using the drug prescriptions that occur before the FBG record as features to predict the value of the FBG record. If through the predictive model we notice that the prescription of a particular drug is associated with the decrease of FBG, then we can consider this drug as a potential candidate to be repurposed for glucose control. It should be noticed that while we are using FBG level control as an example for the ease of presentation, the proposed algorithm can also be used to identify drugs that can be potentially repurposed to control other numeric physical measurements.

2.1 Notation

Without loss of generality, we assume that only drug prescription records and FBG records are available for each patient. And we consider only patients with at least one FBG record throughout their observations. Let there be N patients and p drugs under consideration in total. Suppose that for the i^{th} patient, there are n_i drug prescription records and m_i FBG records in total, where $i \in \{1, 2, \ldots, N\}$. We can use a 2-tuple (x_{ij}, t_{ij}) to represent the j^{th} drug prescription record of

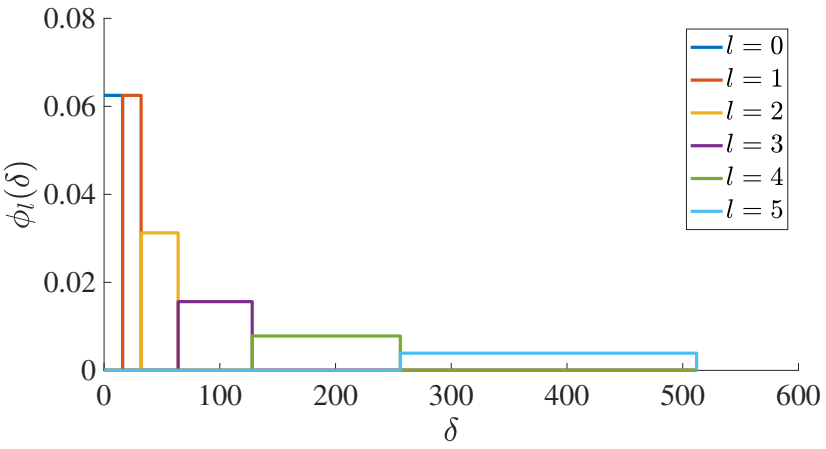

Figure 2: Dyadic influence functions for $S = 512$ and $L = 6$.

the i^{th} patient, where $j \in \{1, 2, \ldots, n_i\}$, $x_{ij} \in \{1, 2, \ldots, p\}$ represents which drug among the p drugs is prescribed, and t_{ij} represents the timestamp of the drug prescription. Similarly, we can also use a 2-tuple (y_{ik}, τ_{ik}) to represent the k^{th} FBG measurement record from the i^{th} patient, where $k \in \{1, 2, \ldots, m_i\}$, y_{ik} denotes the value of the FBG measurement, and τ_{ik} represents the measurement timestamp. Note that given i, $t_{i1} \leq t_{i2} \leq \cdots \leq t_{in_i}$ and $\tau_{i1} \leq \tau_{i2} \leq \cdots \leq \tau_{in_i}$. In this way, we can represent the EHR of each patient as a set of the aforementioned 2-tuples.

3 Methods

We first present how the potential influence of various drugs over time on the value of FBG measurements can be ascertained via the use of dyadic influence functions, directly from raw EHR data. We then present our baseline regularization model that combines the effects of time-varying patient-specific baselines and the effects from various drugs throughout time to predict FBG levels for CDR.

3.1 Dyadic Influence

We assume that drug prescriptions in the EHR of a patient have certain influences on the values of the FBG measurements that occur after the prescriptions. Since drug prescriptions occur throughout time for various patients, given an FBG measurement record, an intuition is that a drug prescription record that occurs long before has less effect, if any, on the value of the FBG measurement in question, compared with a more recent drug prescription occurrence. Based on this intuition, for $t_{ij} \leq \tau_{ik}$, we represent the effect of a drug prescription (x_{ij}, t_{ij}) on an FBG measurement (y_{ik}, τ_{ik}) through a weighted sum of a pre-defined set of dyadic influence functions $\{\phi_l(\cdot)\}_{l=0}^{L-1}$ [3]. Specifically, let $S > 0$ and $L \in \mathbb{N}^+$ be given. For $l \in \{0, 1, 2, \ldots, L-1\}$, we define

$$\alpha_l \triangleq \begin{cases} 2^{L-1}/S, & l = 0 \\ 2^{L-l}/S, & l = 1, 2, \ldots, L-1 \end{cases};$$

and the half-closed-half-open intervals,

$$I_l \triangleq \begin{cases} [0, 1/\alpha_l), & l = 0 \\ [1/\alpha_l, 2/\alpha_l), & l = 1, 2, \ldots, L-1 \end{cases}.$$

Then we define

$$\phi_l(\delta) \triangleq \alpha_l \mathbb{I}(\delta \in I_l),$$

where $\delta = \tau_{ik} - t_{ij}$ is the time difference between the drug prescription and the FBG measurement, and $\mathbb{I}(\cdot)$ is the indicator function. Note that these $\phi_l(\cdot)$'s all integrate to one and are orthogonal to one another.

Figure 2.1 visualizes the set of dyadic influence functions when $S = 512$ and $L = 6$. As can be seen, when the time difference between two events δ increases, the influence decays in exponential order. For $\delta \geq S$, the previous drug prescription is assumed not to have any influence on the value of the FBG measurement in question. Dyadic influence functions provide a flexible approach to ascertain influences of various drug prescriptions in the past on the value of FBG measurement records. This is in contrast to the drug era construction that is prevalent in the pharmacovigilance literature [15, 21, 20, 14], where ad-hoc heuristics are used to generate a consecutive time period during which the value of an FBG measurement is assumed to be under unattenuated influence.

3.2 Baseline Regularization

Baseline regularization assumes that an observed FBG value is due to the influences of various drug prescriptions that occur in the past as well as a hidden, intrinsic baseline FBG value that represents the FBG level that would have been observed if the patient were not under any other influences. Specifically, baseline regularization considers solving the optimization problem in (1):

$$\begin{aligned} \hat{\boldsymbol{b}}, \hat{\boldsymbol{\beta}} \triangleq \arg\min_{\boldsymbol{b},\boldsymbol{\beta}} \frac{1}{2M} \sum_{i=1}^{N} \sum_{k=1}^{m_i} \left(y_{ik} - b_{ik} - \sum_{j=1}^{n_i} \sum_{q=1}^{p} \sum_{l=0}^{L-1} \beta_{ql} \phi_l(\tau_{ik} - t_{ij}) \cdot \mathbb{I}(x_{ij} = q) \right)^2 \\ + \lambda_1 \sum_{i=1}^{N} \sum_{k=1}^{m_i - 1} \left| b_{ik} - b_{i(k+1)} \right| + \lambda_2 \|\boldsymbol{\beta}\|_1, \end{aligned} \tag{1}$$

where $M = \sum_{i=1}^{N} m_i$ is the total number of FBG measurements under consideration, $\lambda_1 > 0$ and $\lambda_2 > 0$ are regularization parameters, and

$$\begin{aligned} \boldsymbol{b} &\triangleq \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m_1} & \cdots & b_{N1} & b_{N2} & \cdots & b_{Nm_N} \end{bmatrix}^\top \text{ and} \\ \boldsymbol{\beta} &\triangleq \begin{bmatrix} \beta_{10} & b_{11} & \cdots & \beta_{1(L-1)} & \cdots & \beta_{p0} & \beta_{p1} & \cdots & \beta_{p(L-1)} \end{bmatrix}^\top \end{aligned}$$

are the parameters that we need to estimate. The baseline regularization problem is a regularized least square problem with a fused lasso penalty (controlled by λ_1) and a lasso penalty (controlled by λ_2).

The parameter $\boldsymbol{b}$ is a baseline parameter vector whose components represent the potentially different baseline FBG levels throughout time for different patients. Such time-varying and patient specific baselines are of great importance to provide flexibility to describe the intricate data generation process in reality. For example, diabetic patients tend to have higher FBG levels compared to a healthy person. Therefore, the fact that the baselines used are patient-specific helps to model

such heterogeneity among different individuals in the data. Even for a particular patient, the FBG levels can also change dramatically over the years as the patient ages. Therefore, the time-varying nature of the baseline parameters also helps to capture the heterogeneity of the FBG levels over time. The baseline parameter $\boldsymbol{b}$ is regularized by a fused lasso penalty, without which $\boldsymbol{b}$ is flexible enough to explain any given FBG level observations. The intuition of using a fused lasso penalty is to minimize the difference between two adjacent baseline parameters. Since baseline parameters represent the FBG values that would have been observed if the patient were not under other influences, it is reasonable to assume that these baseline values are usually relatively stable over a certain period of time, and hence we encourage such stability via the use of fused lasso penalties.

The parameter $\boldsymbol{\beta}$ represents the effects of every drug on the value of the FBG level depending on the time difference between the drug prescription and the FBG measurement. A lasso penalty is used to encourage sparsity over the effect parameter $\boldsymbol{\beta}$ as we assume that only a small portion of drugs can have some effect on the value of an FBG measurement during a certain period of time.

The least square objective is hence to minimize the differences between the observed FBG values and the values given by the model that take into consideration both the time-varying patient-specific baseline parameters that change stably and the sparse effect parameters that describe effects of various drugs during various periods of time.

For the q^{th} drug, let $\left\{\hat{\beta}_{q0}, \hat{\beta}_{q1}, \hat{\beta}_{q2}, \ldots, \hat{\beta}_{q(L-1)}\right\}$ be the set of effects learned from the baseline regularization model. We measure the overall effect of o_q on the FBG level as the average of the elements in the set: $o_q \triangleq \frac{1}{L}\sum_{l=0}^{L-1} \hat{\beta}_{ql}$.

Algorithm 1 Baseline Regularization

Require: $\mathbf{y}$, $\mathbf{Z}$, $\mathbf{D}$, λ_1,and λ_2.
Ensure: $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{\beta}}$.

Initialize $\boldsymbol{\beta}^{(0)}$.
$u \leftarrow 0$.
while true **do**
  $\check{\mathbf{y}}^{(u+1)} \leftarrow \mathbf{y} - \mathbf{Z}\boldsymbol{\beta}^{(u)}$.
  $\boldsymbol{b}^{(u+1)} \leftarrow \arg\min_{\boldsymbol{b}} \frac{1}{2M}\|\check{\mathbf{y}}^{(u+1)} - \boldsymbol{b}\|_2^2 + \lambda_1\|\mathbf{D}\boldsymbol{b}\|_1$. ▷ $\boldsymbol{b}$-step
  $\tilde{\mathbf{y}}^{(u+1)} \leftarrow \mathbf{y} - \boldsymbol{b}^{(u+1)}$.
  $\boldsymbol{\beta}^{(u+1)} \leftarrow \arg\min_{\boldsymbol{\beta}} \frac{1}{2M}\|\tilde{\mathbf{y}}^{(u+1)} - \mathbf{Z}\boldsymbol{\beta}\|_2^2 + \lambda_2\|\boldsymbol{\beta}\|_1$. ▷ $\boldsymbol{\beta}$-step
  if Stopping criteria met **then**
    $\hat{\boldsymbol{b}} \leftarrow \boldsymbol{b}^{(u+1)}$ and $\hat{\boldsymbol{\beta}} \leftarrow \boldsymbol{\beta}^{(u+1)}$.
    return $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{\beta}}$.
  else
    $u \leftarrow u + 1$.
  end if
end while

3.3 Optimization for Baseline Regularization

The baseline regularization problem in (1) is a convex optimization problem. Furthermore, $\boldsymbol{b}$ and $\boldsymbol{\beta}$ are separable in the optimization problem. Therefore, we can perform a blockwise minimization procedure that alternates between the minimization of $\boldsymbol{b}$ and $\boldsymbol{\beta}$ to achieve optimality [25]. When $\boldsymbol{b}$ is fixed, the optimization problem with respect to $\boldsymbol{\beta}$ is a lasso linear regression problem [22]. When $\boldsymbol{\beta}$ is fixed, the optimization problem with respect to $\boldsymbol{b}$ is a blockwise fused lasso signal

approximator problem [24]. Both problems can be solved efficiently. The blockwise minimization algorithm is summarized in Algorithm 1. To see the two subproblems, let

$$z_{iql} \triangleq \sum_{j=1}^{n_i} \phi_l(\tau_{ik} - t_{ij}) \cdot \mathbb{I}(x_{ij} = q).$$

Then (1) can be rewritten as:

$$\hat{\boldsymbol{b}}, \hat{\boldsymbol{\beta}} \triangleq \arg\min_{\boldsymbol{b},\boldsymbol{\beta}} \frac{1}{2M} \|\mathbf{y} - \boldsymbol{b} - \mathbf{Z}\boldsymbol{\beta}\|_2^2 + \lambda_1 \|\mathbf{D}\boldsymbol{b}\|_1 + \lambda_2 \|\boldsymbol{\beta}\|_1, \tag{2}$$

where

$$\mathbf{y} \triangleq \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m_1} & \cdots & y_{N1} & y_{N2} & \cdots & y_{Nm_N} \end{bmatrix}^\top,$$

$\mathbf{Z}$ is an $M \times (p \times L)$ data matrix whose i^{th} row is:

$$\begin{bmatrix} z_{i10} & z_{i11} & \cdots z_{i1(L-1)} & \cdots & z_{ip0} & z_{ip1} & \cdots z_{ip(L-1)} \end{bmatrix},$$

and $\mathbf{D}$ is the blockwise first difference matrix:

$$\mathbf{D} \triangleq \begin{bmatrix} \mathbf{D}_{m_1} & & & \\ & \mathbf{D}_{m_2} & & \\ & & \ddots & \\ & & & \mathbf{D}_{m_N} \end{bmatrix},$$

with an $(m-1) \times m$ first difference matrix defined as $\mathbf{D}_1 = 0$ and:

$$\mathbf{D}_m \triangleq \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & & \\ & & & -1 & 1 \end{bmatrix}.$$

Therefore, from (2), when $\boldsymbol{\beta}$ is fixed, let $\check{\mathbf{y}} \triangleq \mathbf{y} - \mathbf{Z}\boldsymbol{\beta}$; then the blockwise fused lasso signal approximator problem with respect to $\boldsymbol{b}$ is:

$$\arg\min_{\boldsymbol{b}} \frac{1}{2M} \|\check{\mathbf{y}} - \boldsymbol{b}\|_2^2 + \lambda_1 \|\mathbf{D}\boldsymbol{b}\|_1.$$

On the other hand, from (2), when $\boldsymbol{b}$ is fixed, let $\tilde{\mathbf{y}} \triangleq \mathbf{y} - \boldsymbol{b}$, then the lasso linear regression problem with respect to $\boldsymbol{\beta}$ is:

$$\arg\min_{\boldsymbol{\beta}} \frac{1}{2M} \|\tilde{\mathbf{y}} - \mathbf{Z}\boldsymbol{\beta}\|_2^2 + \lambda_2 \|\boldsymbol{\beta}\|_1. \tag{3}$$

In Algorithm 1 the two most computationally-intensive steps are Step 5 and Step 7. The former involves solving a fused lasso signal approximator problem, whose solution can be computed exactly by the dynamic programming algorithm proposed in [11]. The latter involves solving a lasso linear regression problem, which is achieved by the cyclic coordinate descent algorithm with variable screening proposed in [9] and [23].

Table 1: Top Thirty Drugs Selected by Baseline Regularization Associated with FBG Decrease

INDX	CODE	DRUG NAME	SCORE
1	4132	GLUCOPHAGE	-82.388
2	7470	PIOGLITAZONE HCL	-36.869
3	8437	ROSIGLITAZONE MALEATE	-29.046
4	5786	METFORMIN	-18.867
5	4184	GLYBURIDE	-16.664
6	6382	NEEDLES INSULIN DISPOSABLE	-15.233
7	5787	METFORMIN HCL	-9.910
8	4806	INSULIN GLARGINE HUM.REC.ANLOG	-8.523
9	4497	HUM INSULIN NPH/REG INSULIN HM	-7.336
10	160	ACTOS	-6.006
11	7768	PREMARIN	-4.879
12	4106	GLIMEPIRIDE	-4.028
13	6656	NPH HUMAN INSULIN ISOPHANE	-3.613
14	4971	ISOSORBIDE MONONITRATE	-3.229
15	4561	HYDROCORTISONE	-3.084
16	4107	GLIPIZIDE	-3.007
17	9379	THIAMINE HCL	-2.968
18	1573	CAPTOPRIL	-2.871
19	5368	LIPITOR	-2.819
20	9152	SYRING W-NDL DISP INSUL 0.5ML	-2.380
21	1988	CIPROFLOXACIN HCL	-2.367
22	3937	FOSINOPRIL SODIUM	-2.252
23	5390	LISINOPRIL	-2.004
24	9994	VERAPAMIL HCL	-1.965
25	1216	BLOOD SUGAR DIAGNOSTIC	-1.900
26	7760	PREGABALIN	-1.708
27	6803	ONDANSETRON HCL	-1.678
28	4970	ISOSORBIDE DINITRATE	-1.575
29	6540	NITROGLYCERIN	-1.496
30	5571	MAGNESIUM	-1.266

4 Results

To demonstrate the utility of baseline regularization, we run our algorithm on the Marshfield Clinic EHR to identify drugs that can be potentially used to control FBG level. We consider patients with at least one FBG measurement throughout their observations. This leads to a total number of 333,907 FBG measurements from 75,146 patients.

To ascertain influences from drug prescriptions, we choose S to be half a year and $L = 5$ for the dyadic influence function. We only consider drugs that have at least one drug prescription that is at most S amount of time prior to the occurrence of at least one FBG measurement, yielding a total number of 5147 different drugs for consideration. λ_1 and λ_2 are chosen such that roughly 200 drugs will be selected eventually by the model. This is because we do not know in advance whether the drugs returned by the algorithm could potentially control FBG level or not, and we need to examine the findings of the algorithm manually. Therefore, the regularization parameters need to be carefully chosen so that the number of drugs selected by the model can be feasibly examined. Table 1 reports the top thirty drugs ranked by their overall effects among the 180 drugs generated by the baseline regularization using $\lambda_1 = 86$ and $\lambda_2 = 2.841977 \times 10^{-4}$. For more information about choosing the regularization parameters, please see Section 5.

As shown in Table 1, the drugs in green are drugs that are prescribed to control blood sugar level. The drugs in white are not normally used to control blood sugar level. However, there might be some potentially interesting findings based on a literature review. For example, thiamine HCL is reported to reduce the adverse effect of hyperglycemia by inhibiting certain biological pathways [27], and deficiency of thiamine is observed in diabetic patients [17]. Ciprofloxacin HCL could lead to hypoglycemia, according to the medication guide from the Food and Drug Administration (FDA) [5]. Lisinopril is also associated with hypoglycemia, according to the drug label from the FDA [7]. Verapamil HCL is reported to decrease blood sugar level as well as to have some hope in preventing pancreatic β cell loss. Such a loss is considered a pathological characteristic for diabetes [18]. Cases of hypoglycemia associated with the use of pregabalin have been reported [1, 19]. Premarin, fosinopril sodium, and hydrocortisone are potential false positives for our method, since they have been linked to hyperglycemia [4]. Drugs with mixed evidence are also found. For example, according to [4], both Lipitor and captopril are linked to hyperglycemia. Studies that suggest otherwise are also seen in the literature [6, 10, 16].

The baseline regularization algorithm is implemented with `R`. The blockwise fused lasso signal approximator problem is solved using a subroutine in the `R` package `glmgen` [2]. The lasso linear regression problem is solved using the `R` package `glmnet` [8].

5 Notes

5.1 Splitting Patient Records

In (1), we try to control the differences between two adjacent baseline parameters via the use of the fused lasso penalty. Consider the pair b_{ik} and $b_{i(k+1)}$ that indicates the baseline FBG levels corresponding to two adjacent physical measurements. Although the two measurements are adjacent to each other in time, the actual time difference between the two measurements could be large, i.e. $\tau_{ik} \ll \tau_{i(k+1)}$. In this case, it might not be reasonable any more to regularize the difference between the two baselines as the FBG level could go through substantial changes during such a long period of time. Therefore, we consider splitting the records from the same patient into various subsets within which the records are close to each other in time, and just regularize the

differences between adjacent baselines within the same subset. It remains to determine how far apart two adjacent records should be for us to consider them belonging to distinct subsets. We take a data-driven approach to determine this threshold. In detail, we compute the time differences of all adjacent pairs of FBG measurements for all patients. We then use Tukey's method of outlier identification [26] to determine the smallest outlier. The distribution of the differences is heavy-tailed, and most of the differences are small. Therefore, the smallest outlier is a relatively large time difference value, and we set this value as our threshold. After splitting the FBG records of a patient into various subsets, each subset of FBG records can be considered as data from an independent patient. Therefore, the previously established formulation of the baseline regularization model can be naturally extended to handle this situation by simply modifying $\mathbf{D}$ in (2) accordingly. The threshold value identified in our dataset is 4.1 years.

5.2 Model Selection

Since in CDR, we do not know a priori what drugs returned by the algorithm can actually decrease or increase FBG levels, we manually review the drug list to identify potential repurposing opportunities. Therefore, model selection for baseline regularization not only needs to identify a model that explains the data well but also needs to generate a drug list of moderate size so that subsequent reviewing efforts are feasible.

To determine an appropriate λ_1, we start from identifying the minimum λ_1^* such that all the baseline parameters are fused to its average in the following fused lasso signal approximator problem, where we only use the baseline parameter $\boldsymbol{b}$ to model the FBG measurements $\mathbf{y}$:

$$\arg\min_{\boldsymbol{b}} \frac{1}{2M}\|\mathbf{y}-\boldsymbol{b}\|_2^2+\lambda_1\|\mathbf{D}\mathbf{b}\|_1.$$

Define $\mathbf{T}_m$ as an $m\times m$ upper triangular matrix whose upper part and the diagonal are all ones, and whose entries are otherwise zeros. Then according to [28],

$$\lambda_1^*=\max_{i\in\{1,2,\ldots,N\}}\|\mathbf{T}_{m_i}\left(\mathbf{y}_i-\bar{y}_i\mathbf{1}_{m_i}\right)\|_\infty, \tag{4}$$

where $\mathbf{1}_m$ is an $m\times 1$ vector of all ones, and $\bar{y}_i$ is the mean of all the FBG measurements from the i^{th} patient. Upon the determination of λ_1^* in (4), we can choose $\lambda_1=\gamma\lambda_1^*$, where $\gamma\in(0,1)$ can vary to generate different models. The results reported in Table 1 are given by $\lambda_1=0.05\lambda_1^*$.

To determine an appropriate λ_2, we first solve for the pathwise solution to a continuous self-controlled case series (CSCCS) problem [13], which is a lasso linear regression problem assuming a fixed baseline parameter for each patient:

$$\arg\min_{\boldsymbol{\beta}} \frac{1}{2M}\|\mathbf{y}-\mathbf{U}\bar{\mathbf{y}}-(\mathbf{X}-\mathbf{U}\bar{\mathbf{Z}})\boldsymbol{\beta}\|_2^2+\lambda_2\|\boldsymbol{\beta}\|_1,$$

where

$$\mathbf{U}\triangleq\begin{bmatrix}\mathbf{1}_{m_1} & & & \\ & \mathbf{1}_{m_2} & & \\ & & \ddots & \\ & & & \mathbf{1}_{m_N}\end{bmatrix},\quad \bar{\mathbf{y}}\triangleq(\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{U}^\top\mathbf{y},\quad \bar{\mathbf{Z}}\triangleq(\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{U}^\top\mathbf{Z}.$$

In our experiments, we are aiming at selecting about 200 drugs in the end. Therefore, from the solution path, we choose an λ_2 whose solution selects about 250 drugs and we use this λ_2

for the baseline regularization problem. The solution to the CSCCS problem can also be used to initialize $\boldsymbol{\beta}^{(0)}$ in baseline regularization in Algorithm 1. Given the same λ_2, we notice that the baseline regularization problem usually will select fewer drugs compared to the corresponding CSCCS problem. Intuitively, this is because the introduction of time-varying and patient-specific baseline parameters in the baseline regularization problem help to explain the changes in the FBG measurements better. Therefore, fewer drugs are needed in order to explain the changes of FBG levels in the dataset, yielding a sparser drug effect parameterization.

When multiple configurations of λ_1's and λ_2's are provided, we can use Akaike information criterion (AIC) or Bayesian information criterion (BIC) for model selection. The degree of freedom of the baseline regularization model needed in the calculation is the summation of the degree of freedom of the baseline parameter $\boldsymbol{b}$ and the degree of freedom of the drug effect parameter $\boldsymbol{\beta}$. The former is the total number of piecewise constant segments of $\boldsymbol{b}$ and the latter is the number of nonzero entries of $\boldsymbol{\beta}$.

Since the dimension of the parameterization in baseline regularization is larger than the sample size of the data, caution needs to be paid when we choose regularization parameters. Essentially, we would like to choose large λ_1 and λ_2 to impose strong regularization to avoid overfitting. The degree of freedom of the learned model also needs to be monitored and controlled so that it is smaller than the sample size of the data.

5.3 Stopping Criteria

Since the baseline regularization problem is a convex optimization problem, we can verify the convergence of the optimization procedure in Algorithm 1 by checking the violation of the Karush–Kuhn–Tucker (KKT) conditions of the current iterate. Since when $\boldsymbol{\beta}^{(u)}$ is given, the update to $\boldsymbol{b}^{(u+1)}$ can be carried out exactly by Step 4 and Step 5 of Algorithm 1, we are interested in knowing the violation due to $\boldsymbol{b}^{(u+1)}$ and $\boldsymbol{\beta}^{(u)}$ via the KKT conditions of (3):

$$\mathbf{s}^{(u)} = \frac{1}{n\lambda_2}\mathbf{Z}^\top \left(\mathbf{y} - \boldsymbol{b}^{(u+1)} - \mathbf{Z}\boldsymbol{\beta}^{(u)}\right),$$

where $\mathbf{s}^{(u)}$ is the subgradient of $\|\boldsymbol{\beta}\|_1$. If $\boldsymbol{b}^{(u+1)}$ and $\boldsymbol{\beta}^{(u)}$ are optimal, then

$$\hat{s}_d \begin{cases} = 1, & \beta_d^{(u)} > 0 \\ = -1, & \beta_d^{(u)} < 0 \\ \in [-1, 1], & \beta_d^{(u)} = 0 \end{cases}, \tag{5}$$

where $\hat{s}_d$ and $\beta_d^{(u)}$ are the d^{th} components of $\hat{\mathbf{s}}$ and $\boldsymbol{\beta}^{(u)}$, respectively. By measuring how much $\mathbf{s}^{(u)}$ violates the specification of $\hat{\mathbf{s}}$ in (5) via $\|\mathbf{v}^{(u)}\|_2$, where the d^{th} component of $\mathbf{v}^{(u)}$ is

$$v_d^{(u)} \triangleq \begin{cases} s_d^{(u)} - 1, & \beta_d^{(u)} > 0 \\ s_d^{(u)} + 1, & \beta_d^{(u)} < 0 \\ \max\left\{0, |s_d^u| - 1\right\}, & \beta_d^{(u)} = 0 \end{cases},$$

we know about how far away the current solution is to optimality. Such a measurement can be used as a stopping criterion. In our experiment, we set $\|\mathbf{v}^{(u)}\|_2 \leq 0.01$ as our stopping criterion.

6 Conclusion

We have presented an algorithm to predict the effects of drugs on numeric physical measurements in the EHR such as fasting blood glucose. Drugs with a strong effect to decrease the measurement are potential repurposing targets. Our method inherits from self-controlled case series [13] the ability to take into account inter-patient variation. By addition of a time-varying baseline it can also address intra-patient variation over time. And by use of dyadic influence functions it can avoid the need to decide drug eras and can model different effect times for different drugs.

Acknowledgement

The authors would like to gratefully acknowledge the NIH BD2K Initiative grant U54 AI117924, the NIGMS grant 2RO1 GM097618, NIH CTSA at UW-Madison 1UL1TR002373, NSF grant CCF-1418976, and ARO grant W911NF-17-1-0357.

References

[1] M. ABE, S. NAKAMURA, T. HIGA, J. OKUBO, and M. KAKINOHANA. Frequent hypoglycemia after prescription of pregabalin in a patient with painful diabetic neuropathy. *Journal of Japan Society of Pain Clinicians*, advpub, 2015. doi: 10.11321/jjspc.14-0035.

[2] T. Arnold, V. Sadhanala, and R. J. Tibshirani. glmgen: Fast generalized lasso solver, 2014.

[3] Y. Bao, Z. Kuang, P. Peissig, D. Page, and R. Willett. Hawkes process modeling of adverse drug reactions with longitudinal observational data. In *Machine Learning for Healthcare Conference*, pages 177–190, 2017.

[4] DiabetesInControl. Drugs that can affect blood glucose levels. http://www.diabetesincontrol.com/wp-content/uploads/2010/07/www.diabetesincontrol.com_images_tools_druglistaffectingbloodglucose.pdf, 2015. (Visited on 03/12/2018).

[5] FDA. Cipro medication guide. https://www.fda.gov/downloads/Drugs/DrugSafety/UCM088572.pdf, . (Visited on 03/12/2018).

[6] FDA. Lipitor (atorvastatin calcium) tablets. https://www.accessdata.fda.gov/drugsatfda_docs/label/2009/020702s057lbl.pdf, . (Visited on 03/12/2018).

[7] FDA. Zestril (lisinopril) label. https://www.accessdata.fda.gov/drugsatfda_docs/label/2009/019777s054lbl.pdf, . (Visited on 03/12/2018).

[8] J. Friedman, T. Hastie, and R. Tibshirani. glmnet: Lasso and elastic-net regularized generalized linear models. *R package version*, 1(4), 2009.

[9] J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

[10] E. Girardin and D. Raccah. Interaction between converting enzyme inhibitors and hypoglycemic sulfonamides or insulin. *Presse medicale (Paris, France: 1983)*, 27(37):1914–1923, 1998.

[11] N. A. Johnson. A dynamic programming algorithm for the fused lasso and l 0-segmentation. *Journal of Computational and Graphical Statistics*, 22(2):246–260, 2013.

[12] Z. Kuang, J. Thomson, M. Caldwell, P. Peissig, R. Stewart, and D. Page. Baseline regularization for computational drug repositioning with longitudinal observational data. In *IJCAI: proceedings of the conference*, volume 2016, page 2521. NIH Public Access, 2016.

[13] Z. Kuang, J. Thomson, M. Caldwell, P. Peissig, R. Stewart, and D. Page. Computational drug repositioning using continuous self-controlled case series. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 491–500. ACM, 2016.

[14] Z. Kuang, P. Peissig, V. Santos Costa, R. Maclin, and D. Page. Pharmacovigilance via baseline regularization with large-scale longitudinal observational data. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1537–1546. ACM, 2017.

[15] P. M. Nadkarni. Drug safety surveillance using de-identified emr and claims data: issues and challenges. *Journal of the American Medical Informatics Association: JAMIA*, 17(6):671, 2010.

[16] P. Neerati and J. Gade. Influence of atorvastatin on the pharmacokinetics and pharmacodynamics of glyburide in normal and diabetic rats. *European Journal of Pharmaceutical Sciences*, 42(3):285–289, 2011.

[17] G. Page, D. Laight, and M. Cummings. Thiamine deficiency in diabetes mellitus and the impact of thiamine replacement on glucose metabolism and vascular disease. *International journal of clinical practice*, 65(6):684–690, 2011.

[18] R. R. Poudel and N. K. Kafle. Verapamil in diabetes. *Indian journal of endocrinology and metabolism*, 21(5):788, 2017.

[19] P. Raman. Hypoglycemia induced by pregabalin. *Journal of The Association of Physicians of India*, 64, 2016.

[20] P. Ryan. Establishing a drug era persistence window for active surveillance. foundation for the national institutes of health, 2010, 2015.

[21] S. E. Simpson, D. Madigan, I. Zorych, M. J. Schuemie, P. B. Ryan, and M. A. Suchard. Multiple self-controlled case series for large-scale longitudinal observational databases. *Biometrics*, 69(4):893–902, 2013.

[22] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[23] R. Tibshirani, J. Bien, J. Friedman, T. Hastie, N. Simon, J. Taylor, and R. J. Tibshirani. Strong rules for discarding predictors in lasso-type problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(2):245–266, 2012.

[24] R. J. Tibshirani and J. Taylor. The solution path of the generalized lasso. *The Annals of Statistics*, pages 1335–1371, 2011.

[25] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of optimization theory and applications*, 109(3):475–494, 2001.

[26] J. W. Tukey. *Exploratory data analysis*, volume 2. Reading, Mass., 1977.

[27] K. vinh quoc Luong and L. T. H. Nguyen. The impact of thiamine treatment in the diabetes mellitus. *Journal of clinical medicine research*, 4(3):153, 2012.

[28] J. Wang, W. Fan, and J. Ye. Fused lasso screening rules via the monotonicity of subdifferentials. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1806–1820, 2015.