Baseline Regularization for Computational Drug Repositioning with Longitudinal Observational Data

Zhaobin Kuang[1], James Thomson[2], Michael Caldwell[3], Peggy Peissig[4], Ron Stewart[5], David Page[6]

University of Wisconsin-Madison[1,6], Morgridge Institute for Research[2,5], Marshfield Clinic[3,4]

zkuang@wisc.edu[1], JThomson@morgridge.org[2], caldwell.michael@marshfieldclinic.org[3], Peissig.Peggy@mcrf.mfldclin.edu[4], RStewart@morgridgeinstitute.org[5], page@biostat.wisc.edu[6]

Abstract

Computational Drug Repositioning (CDR) is the knowledge discovery process of finding new indications for existing drugs leveraging heterogeneous drug-related data. Longitudinal observational data such as Electronic Health Records (EHRs) have become an emerging data source for CDR. To address the high-dimensional, irregular, subject and time-heterogeneous nature of EHRs, we propose Baseline Regularization (BR) and a variant that extend the one-way fixed effect model, which is a standard approach to analyze small-scale longitudinal data. For evaluation, we use the proposed methods to search for drugs that can lower Fasting Blood Glucose (FBG) level in the Marshfield Clinic EHR. Experimental results suggest that the proposed methods are capable of rediscovering drugs that can lower FBG level as well as identifying some potential blood sugar lowering drugs in the literature.

1 Introduction

Computational Drug Repositioning (CDR) is the knowledge discovery process of finding new indications for existing drugs, leveraging heterogeneous drug-related data. It is a challenging and rapidly-growing application of artificial intelligence [Andronis *et al.*, 2011; Fakhraei *et al.*, 2013; Li and Lu, 2013]. Longitudinal observational data such as Electronic Health Records (EHRs) have become an emerging data source for CDR [Xu *et al.*, 2014]. In EHRs, detailed across-time clinical information on patients, such as drug prescriptions, conditions, lab test results, and demographics, are collected from a large and diverse population. This large-scale data source provides potentially valuable information to foster the study of correlations among various drugs, conditions, and lab results in diverse patient profiles, which is an important task for predictive analytics in the healthcare industry.

For the task of CDR, we build a predictive model that uses the drug prescription history of patients to predict their continuous numeric value of Fasting Blood Glucose (FBG) level. We examine the drugs (predictors) that have significant blood sugar lowering effects. If some of them are not known to lower blood sugar already, we can consider those drugs as potential candidates for repositioning to control blood sugar, with further inspection.

We observe that patients in the EHRs have extremely diverse FBG level profiles (e.g. some people tend to have higher FBG level than the others). Furthermore, different FBG level measurements taken far apart in time might have very different values. This is especially true if some persistent blood glucose altering events occur to a person, such as the diagnosis of diabetes.

Based on these observations, our Baseline Regularization (BR) model assumes that there is a patient-specific, time-varying, but unknown baseline FBG value for each patient over different time periods. The model further assumes that the *observed* FBG level of a particular patient taken at a particular time is influenced by the joint effects of the baseline value *and* the exposure statuses of various drugs for that patient at that time. We first build a sparse fixed-effect model to describe our assumptions, and then we impose regularization to the baseline model parameters, hence the name of our model, baseline regularization. For computational efficiency, we also propose an alternative formulation to the original baseline regularization model.

Our contributions are threefold:

- We introduce the baseline regularization model for the task of CDR, which generalizes the standard one-way fixed effect model [Frees, 2004].
- We propose an alternative formulation to the original baseline regularization model, which is equivalent to an L_1 regularized linear model and hence can be solved efficiently.
- Using our methodology, we discover some drugs with literature support indicating their potential glucose-lowering effects that are worthy of further investigation.

2 Background

2.1 Electronic Health Records (EHRs)

Figure 1 visualizes a simple EHR with two patients stored in a relational database. Two tables are presented. Table 1a represents *drug era* records. Each row records the identifier of a patient with the name, the start date, and the end date of a drug prescribed to this patient. We consider that the patient is under exposure of the drug during the time span from the start date to the end date of the drug era record. Table 1b

PATIENT_ID	DRUG_NAME	START_DATE	END_DATE
1	HUMALOG	Jan-28-2005	Mar-23-2005
1	HUMALOG	Jun-17-2005	Jul-20-2005
2	INSULIN	Mar-07-1998	May-14-1998

(a) Drug Era Records

PATIENT_ID	DATE	VALUE
1	Jan-28-2005	130
2	Apr-13-1998	95
2	Aug-12-1998	140

(b) Fasting Blood Glucose Records

Figure 1: Electronic Health Records (EHRs)

represents FBG level records. Each row records the identifier of a patient, with the measured date and the numeric value of a FBG measurement. Within each patient, drug era records of the same drug do not overlap in time. Furthermore, only one FBG measurement can be taken at a particular date within each patient.

2.2 Notation

Let there be N patients and M drugs in the EHR. We use y_{ij} to denote the numeric value of the j^{th} FBG measurement taken from the i^{th} patient, where $i \in \{1, 2, \cdots, N\}$ and $j \in \{1, 2, \cdots, J_i\}$. We denote $n = \sum_{i=1}^{N} J_i$. Furthermore, we use $\boldsymbol{x}_{ij}$ to denote an $M \times 1$ binary vector, with $x_{ijk} = 1$ representing that the i^{th} patient is under the exposure of the k^{th} drug at the time when the j^{th} FBG measurement was taken, where $i \in \{1, 2, \cdots, N\}$, $j \in \{1, 2, \cdots, J_i\}$, and $k \in \{1, 2, \cdots, M\}$. Similarly, $x_{ijk} = 0$ indicates that the i^{th} patient is *not* exposed to the k^{th} drug at the time when the j^{th} FBG measurement was taken.

3 The Baseline Regularization Model

3.1 The Fixed-Effect Model

We are interested in using $\boldsymbol{x}_{ij}$'s to predict y_{ij}'s. For this purpose, we consider the following one-way fixed-effect model:

$$y_{ij} \mid \boldsymbol{x}_{ij} = \alpha_i + \boldsymbol{\beta}^\top \boldsymbol{x}_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \overset{iid}{\sim} N\left(0, \sigma^2\right), \quad (1)$$

where α_i is a patient-specific, time-invariant, nonrandom parameter representing the constant baseline FBG level of the i^{th} patient, and $\boldsymbol{\beta}$ is an $M \times 1$ vector with β_k, $k \in \{1, 2, \cdots, M\}$, representing the effect of the k^{th} drug on the j^{th} measurement of the i^{th} patient if the patient is exposed to that drug at the time when the measurement was taken. ϵ_{ij}'s represent the independent and identically Gaussian distributed noises with zero mean and fixed but unknown variance σ^2. Based on the aforementioned definitions, intuitively, if $\boldsymbol{x}_{ij} = \boldsymbol{0}$, then

$$\mathbb{E}\left[y_{ij} \mid \boldsymbol{x}_{ij} = \boldsymbol{0}\right] = \alpha_i,$$

indicating that the baseline parameter α_i is the value of the expected FBG level of the i^{th} patient if the patient is exposed to no drugs.

Fitting the fixed-effect model in (1) is equivalent to solving the following least square problem:

$$\arg\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \frac{1}{2} \left\| \boldsymbol{y} - \begin{bmatrix} \boldsymbol{Z} & \boldsymbol{X} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} \right\|_2^2, \quad (2)$$

where

$$\boldsymbol{y} = \begin{bmatrix} y_{11} & \cdots & y_{1J_1} & \cdots & y_{N1} & \cdots & y_{1J_N} \end{bmatrix}^\top,$$
$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_{11} & \cdots & \boldsymbol{x}_{1J_1} & \cdots & \boldsymbol{x}_{N1} & \cdots & \boldsymbol{x}_{NJ_N} \end{bmatrix}^\top,$$
$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_N \end{bmatrix}^\top, \ \boldsymbol{Z} = \text{diag}\left(\boldsymbol{1}_1, \cdots, \boldsymbol{1}_N\right),$$

where $\boldsymbol{Z}$ is a block diagonal matrix with $\boldsymbol{1}_i$ being an all-one $J_i \times 1$ vector.

3.2 Time-Varying Baseline

The introduction of the patient-specific baseline parameter $\boldsymbol{\alpha}$ in (2) explains the heterogeneous nature of the FBG levels measured from different patients. However, since $\boldsymbol{\alpha}$ is time-invariant, the model in (2) essentially assumes that the baseline FBG level of each patient does not change over time. This is a restrictive assumption for EHR data in that FBG records of a person can be collected over decades, and the baseline FBG level will usually change over such a long period of time [O'Sullivan, 1974; Ko *et al.*, 2006]. This is especially true if some persistent glucose altering events occur to a patient, such as the diagnosis of diabetes. To incorporate a time-varying baseline, we extend the fixed-effect model in (1) as follows:

$$y_{ij} \mid \boldsymbol{x}_{ij} = \alpha_i + t_{ij} + \boldsymbol{\beta}^\top \boldsymbol{x}_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \overset{iid}{\sim} N\left(0, \sigma^2\right), \ (3)$$

where t_{ij} can be considered as the *deviation* of the FBG level baseline from α_i at the time when the j^{th} measurement was taken from the i^{th} patient. Fitting model (3) is equivalent to solving the following least square problem:

$$\begin{aligned} &\arg\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{t}} \mathcal{L}\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{t}\right) \\ =&\arg\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{t}} \frac{1}{2} \left\| \boldsymbol{y} - \begin{bmatrix} \boldsymbol{Z} & \boldsymbol{X} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \\ \boldsymbol{t} \end{bmatrix} \right\|_2^2, \end{aligned} \quad (4)$$

where

$$\boldsymbol{t} = \begin{bmatrix} t_{11} & \cdots & t_{1J_1} & \cdots & t_{N1} & \cdots & t_{NJ_N} \end{bmatrix}^\top,$$

and $\boldsymbol{I}$ is an $n \times n$ identity matrix.

3.3 Parsimonious Representation

The parameter of interest in our task is $\boldsymbol{\beta}$. In contrast, we refer to $\boldsymbol{\alpha}$ and $\boldsymbol{t}$ as *nuisance parameters*. In this section, we consider deriving a parsimonious model from (4) that is $\boldsymbol{\alpha}$-free. Minimizing $\mathcal{L}\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{t}\right)$ with respect to $\boldsymbol{\alpha}$ is equivalent to:

$$\frac{\partial \mathcal{L}\left(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{t}\right)}{\partial \boldsymbol{\alpha}} = \boldsymbol{0} \Rightarrow \boldsymbol{\alpha} = \bar{\boldsymbol{y}} - \bar{\boldsymbol{X}}\boldsymbol{\beta} - \boldsymbol{Z}^\dagger \boldsymbol{t}, \quad (5)$$

where $\boldsymbol{Z}^\dagger = \left(\boldsymbol{Z}^\top \boldsymbol{Z}\right)^{-1} \boldsymbol{Z}^\top$, $\bar{\boldsymbol{y}}$ is an $N \times 1$ vector with $\bar{y}_i = \frac{1}{J_i} \sum_{j=1}^{J_i} y_{ij}$, and $\bar{\boldsymbol{X}}$ is an $N \times M$ matrix with the i^{th} row being $\bar{\boldsymbol{X}}_{i\cdot} = \frac{1}{J_i} \sum_{j=1}^{J_i} \boldsymbol{x}_{ij}^\top$. Substituting (5) into (4) yields,

$$\arg\min_{\boldsymbol{\beta},\boldsymbol{t}} \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{Z}\bar{\boldsymbol{y}} - \begin{bmatrix} \boldsymbol{X} - \boldsymbol{Z}\bar{\boldsymbol{X}} & \boldsymbol{I} - \boldsymbol{Z}\boldsymbol{Z}^\dagger \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{t} \end{bmatrix} \right\|_2^2, \quad (6)$$

which is free of $\boldsymbol{\alpha}$.

3.4 Sparsity and Baseline Regularization

The sample size in (6) is n; however, the number of parameters in the model is $M + n$. Overparameterization in (6) motivates us to impose regularization onto the parameters $\boldsymbol{\beta}$ and $\boldsymbol{t}$ so as to control their degree of freedom. A type of regularization that can be imposed on $\boldsymbol{\beta}$ is the lasso [Tibshirani, 1996] penalty that encourages sparsity, which yields,

$$\begin{aligned}\arg\min_{\boldsymbol{\beta},\boldsymbol{t}} \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{Z}\bar{\boldsymbol{y}} - \begin{bmatrix} \boldsymbol{X} - \boldsymbol{Z}\bar{\boldsymbol{X}} & \boldsymbol{I} - \boldsymbol{Z}\boldsymbol{Z}^{\dagger} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{t} \end{bmatrix} \right\|_2^2 \\ + \lambda_1 \left\| \boldsymbol{\beta} \right\|_1 ,\end{aligned} \tag{7}$$

where $\lambda_1 > 0$. The incorporation of the lasso penalty in (7) can potentially set most components of $\boldsymbol{\beta}$ to zero, which means the model makes the assumption that only a small subset of drugs in the EHRs will alter the FBG level.

We now consider the regularization of the baseline parameter $\boldsymbol{t}$. For this purpose, we first consider an adjacent pair of FBG measurements from the same patient i, i.e. y_{ij} and $y_{i(j+1)}$, where $i \in \{1, 2, \cdots, N\}$, and $j \in \{1, 2, \cdots, J_i - 1\}$. We predefine a time threshold δ; furthermore, we denote the time when y_{ij} was taken as τ_{ij} and we define $\tau_{i(j+1)}$ accordingly. We consider only adjacent pairs of FBG measurements from the same patient that satisfies

$$\tau_{i(j+1)} - \tau_{ij} \leq \delta. \tag{8}$$

In words, the constraint in (8) means that we only focus on adjacent pairs of FBG measurements from the same patient that are measured close enough in time. For the pair of y_{ij} and $y_{i(j+1)}$, their corresponding baseline FBG levels are $\alpha_i + t_{ij}$ and $\alpha_i + t_{i(j+1)}$ respectively. A reasonable assumption is, if a pair of adjacent FBG measurements from the same patient are close enough in time, their FBG baseline levels should not be very different from each other; that is to say, we expect the quantity

$$|(\alpha_i + t_{i(j+1)}) - (\alpha_i + t_{ij})| = |t_{i(j+1)} - t_{ij}|$$

to be small. This assumption motivates us to incorporate a fused lasso penalty [Tibshirani and Taylor, 2011] to (7) in order to regulate the baseline parameter $\boldsymbol{t}$, yielding:

$$\begin{aligned}\arg\min_{\boldsymbol{\beta},\boldsymbol{t}} \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{Z}\bar{\boldsymbol{y}} - \begin{bmatrix} \boldsymbol{X} - \boldsymbol{Z}\bar{\boldsymbol{X}} & \boldsymbol{I} - \boldsymbol{Z}\boldsymbol{Z}^{\dagger} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{t} \end{bmatrix} \right\|_2^2 \\ + \lambda_1 \left\| \boldsymbol{\beta} \right\|_1 + \lambda_2 \left\| \boldsymbol{D}_{\delta} \boldsymbol{t} \right\|_1 .\end{aligned} \tag{9}$$

In (9), $\lambda_2 > 0$ and $\boldsymbol{D}_{\delta}$ is a matrix that depends on δ. It only contains 0, and ± 1 entries. When multiplied with $\boldsymbol{t}$, the role of each row of $\boldsymbol{D}_{\delta}$ is to subtract the earlier FBG baseline parameter from the later baseline parameter of an adjacent pair of records from the same patient that were collected within a time span of δ. We denote the number of adjacent pairs that satisfy the aforementioned criteria as s, then $\boldsymbol{D}_{\delta}$ is an $s \times n$ matrix. The fused lasso penalty in (9) penalizes adjacent baseline parameters that are very different from each other, and can potentially drive most differences to zero. Therefore the model helps to keep the value of many adjacent components in $\boldsymbol{t}$ the same or very closed to each other, which reduces the degree of freedom of $\boldsymbol{t}$ significantly. With regularization for both $\boldsymbol{\beta}$ and $\boldsymbol{t}$, the overparameterization problem in the time-varying baseline model is relieved substantially.

We define $\boldsymbol{\psi} = \left(\boldsymbol{I} - \boldsymbol{Z}\boldsymbol{Z}^{\dagger}\right)\boldsymbol{t}$ and hence:

$$\boldsymbol{\psi} = \begin{bmatrix} \psi_{11} & \cdots & \psi_{1J_1} & \cdots & \psi_{N1} & \cdots & \psi_{NJ_N} \end{bmatrix}^{\top},$$

where

$$\psi_{ij} = t_{ij} - \bar{t}_i, \quad \bar{t}_i = \frac{1}{J_i} \sum_{j'=1}^{J_i} t_{ij'} .$$

As illustrated in Section 3.2, t_{ij}'s can be considered as deviation of baseline FBG level from α_i. Therefore, it is reasonable to assume that the mean deviation, $\bar{t}_i$, is close to zero. Furthermore, as J_i's increase, $\forall i$, the entries in $\boldsymbol{Z}\boldsymbol{Z}^{\dagger}$ are closer and closer to zero too. Based on these observations, we approximate $\boldsymbol{I} - \boldsymbol{Z}\boldsymbol{Z}^{\dagger}$ in (9) with $\boldsymbol{I}$, and define the *baseline regularization* model as:

$$\begin{aligned}\arg\min_{\boldsymbol{\beta},\boldsymbol{t}} \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{Z}\bar{\boldsymbol{y}} - \begin{bmatrix} \boldsymbol{X} - \boldsymbol{Z}\bar{\boldsymbol{X}} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{t} \end{bmatrix} \right\|_2^2 \\ + \lambda_1 \left\| \boldsymbol{\beta} \right\|_1 + \lambda_2 \left\| \boldsymbol{D}_{\delta} \boldsymbol{t} \right\|_1 .\end{aligned} \tag{10}$$

The approximation made in (10) offers substantial computational advantages in solving the problem, which we will further illustrate in Section 4.

4 Optimization Procedure for BR

The baseline regularization model in (10) can be rewritten as a generalized lasso [Tibshirani and Taylor, 2011] problem:

$$\begin{aligned}\arg\min_{\boldsymbol{\beta},\boldsymbol{t}} \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{Z}\bar{\boldsymbol{y}} - \begin{bmatrix} \boldsymbol{X} - \boldsymbol{Z}\bar{\boldsymbol{X}} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{t} \end{bmatrix} \right\|_2^2 \\ + \lambda_1 \left\| \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{\lambda_2}{\lambda_1} \boldsymbol{D}_{\delta} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{t} \end{bmatrix} \right\|_1 .\end{aligned} \tag{11}$$

However, due to the large volume of data in EHRs, solving (11) directly using a generalized lasso solver might be slow or even infeasible. The difficulty of directly solving (11) motivates us to consider the following optimization strategy.

Notice that $\boldsymbol{\beta}$ and $\boldsymbol{t}$ are *separable* in the lasso and fused lasso penalty in (10), i.e. the penalty part of the BR model in (10) is a summation of two functions that involves only $\boldsymbol{\beta}$ and $\boldsymbol{t}$ respectively in each function. We therefore can perform blockwise minimization over $\boldsymbol{\beta}$ and $\boldsymbol{t}$ alternatively [Tseng, 2001] in order to solve (10). Specifically, denote $\boldsymbol{\beta}^{(k)}$ as the estimation of $\boldsymbol{\beta}$ generated after the k^{th} iteration, at iteration $k+1$, we consider $\boldsymbol{\beta}^{(k)}$ as a constant and optimize with respect to $\boldsymbol{t}$ for $\boldsymbol{t}^{(k+1)}$ by solving the $\boldsymbol{t}$-step subproblem:

The $\boldsymbol{t}$-Step:

$$\begin{aligned}\boldsymbol{\xi}^{(k)} &= \boldsymbol{y} - \boldsymbol{Z}\bar{\boldsymbol{y}} - \left(\boldsymbol{X} - \boldsymbol{Z}\bar{\boldsymbol{X}}\right)\boldsymbol{\beta}^{(k)}, \\ \boldsymbol{t}^{(k+1)} &= \arg\min_{\boldsymbol{t}} \frac{1}{2} \left\| \boldsymbol{\xi}^{(k)} - \boldsymbol{t} \right\|_2^2 + \lambda_2 \left\| \boldsymbol{D}_{\delta} \boldsymbol{t} \right\|_1 .\end{aligned} \tag{12}$$

The problem in (12) is a fused lasso problem with identity design matrix, which can be solved efficiently by the `genlasso` [Arnold *et al.*, 2014] package in `R`. Notice that the reason why we have an identity design matrix in (12) is due to the approximation made in (10). Had the approximation not been made, we need to solve a fused lasso problem with a general design matrix, which is much less efficient than solving the problem in (12).

After computing $\boldsymbol{t}^{(k+1)}$ from (12), we fix $\boldsymbol{t}^{(k+1)}$ and optimize with respect to $\boldsymbol{\beta}$ for $\boldsymbol{\beta}^{(t+1)}$ by solving the $\boldsymbol{\beta}$-step subproblem:

The $\boldsymbol{\beta}$-Step:

$$\begin{aligned} \boldsymbol{\phi}^{(k+1)} &= \boldsymbol{y} - \boldsymbol{Z}\bar{\boldsymbol{y}} - \boldsymbol{t}^{(k+1)}, \quad \boldsymbol{\Delta} = \left(\boldsymbol{X} - \boldsymbol{Z}\bar{\boldsymbol{X}}\right), \\ \boldsymbol{\beta}^{(k+1)} &= \arg\min_{\boldsymbol{\beta}} \frac{1}{2} \left\| \boldsymbol{\phi}^{(k+1)} - \boldsymbol{\Delta}\boldsymbol{\beta} \right\|_2^2 + \lambda_1 \left\|\boldsymbol{\beta}\right\|_1 . \end{aligned} \tag{13}$$

The $\boldsymbol{\beta}$-step is an L_1 regularized linear model, which can be solved efficiently by the `glmnet` [Friedman *et al.*, 2010] package in `R`. The blockwise minimization algorithm can therefore iteratively alternate between the $\boldsymbol{t}$-step and the $\boldsymbol{\beta}$-step to solve for the BR model defined in (10). The algorithm is summarized in Algorithm 1.

Algorithm 1 Blockwise Minimization

```
Require: $\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{D}_\delta, \lambda_1 > 0, \lambda_2 > 0$
Ensure: solution $\boldsymbol{\beta}, \boldsymbol{t}$
Initialize $\boldsymbol{t}^{(0)} = \boldsymbol{0}$, solve
    $\boldsymbol{\beta}^{(0)} = \arg\min_{\boldsymbol{\beta}} \frac{1}{2} \left\|\boldsymbol{y} - \boldsymbol{Z}\bar{\boldsymbol{y}} - \left(\boldsymbol{X} - \boldsymbol{Z}\bar{\boldsymbol{X}}\right)\boldsymbol{\beta}\right\|_2^2 + \lambda_1 \left\|\boldsymbol{\beta}\right\|_1 .$
for $k \leftarrow 0, 1, 2, \cdots$ do
    Solve the $\boldsymbol{t}$-Step in (12) for $\boldsymbol{t}^{(k+1)}$
    Solve the $\boldsymbol{\beta}$-Step in (13) for $\boldsymbol{\beta}^{(k+1)}$
    if stop criterion satisfied then
        return $\boldsymbol{\beta}^{(k+1)}, \boldsymbol{t}^{(k+1)}$
    end if
end for
```

5 An Alternative Formulation of BR

Let δ be given, we can define $\boldsymbol{D}_\delta$ accordingly. Consider the following modification of the unregulated problem in (4):

$$\arg\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{t}} \frac{1}{2} \left\| \boldsymbol{D}_\delta \boldsymbol{y} - \boldsymbol{D}_\delta \left[\boldsymbol{Z} \quad \boldsymbol{X} \quad \boldsymbol{I}\right] \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \\ \boldsymbol{t} \end{bmatrix} \right\|_2^2 , \tag{14}$$

which multiples $\boldsymbol{D}_\delta$ with the residue in (4) and considers the square of norm of the transformed residue. In (14), $\boldsymbol{D}_\delta \boldsymbol{Z}\boldsymbol{\alpha} = \boldsymbol{0}$, and hence the problem in (14) is $\boldsymbol{\alpha}$-free. Equipped the problem in (14) with the lasso and the fused lasso penalty, and let $\boldsymbol{\gamma} = \boldsymbol{D}_\delta \boldsymbol{t}$, we consider an optimization problem with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$:

$$\begin{aligned} \arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} \quad & \frac{1}{2} \left\|\boldsymbol{D}_\delta \boldsymbol{y} - \boldsymbol{D}_\delta \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\gamma}\right\|_2^2 + \lambda_1 \left\|\boldsymbol{\beta}\right\|_1 + \lambda_2 \left\|\boldsymbol{\gamma}\right\|_1 , \\ \text{s.t.} \quad & \boldsymbol{\gamma} = \boldsymbol{D}_\delta \boldsymbol{t}, \quad \boldsymbol{t} \in \mathbb{R}^n. \end{aligned}$$

Note that we can verify that $\boldsymbol{D}_\delta$ is a row full rank matrix. Therefore, $\forall \boldsymbol{\gamma} \in \mathbb{R}^s$, $\exists \boldsymbol{t} \in \mathbb{R}^n$, s.t. $\boldsymbol{\gamma} = \boldsymbol{D}_\delta \boldsymbol{t}$. This property indicates that the equality constraint can always be satisfied and hence can be dropped, resulting in an **A**lternative formulation of the BR model (ABR):

$$\arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} \frac{1}{2} \left\|\boldsymbol{D}_\delta \boldsymbol{y} - \boldsymbol{D}_\delta \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\gamma}\right\|_2^2 + \lambda_1 \left\|\boldsymbol{\beta}\right\|_1 + \lambda_2 \left\|\boldsymbol{\gamma}\right\|_1 , \tag{15}$$

which is also an L_1 regularized linear model that can be solved efficiently. The intuition behind this modification is to take the difference between a pair of adjacent records that are close enough with each other in time from the same patient in order to eliminate $\boldsymbol{\alpha}$ for a parsimonious representation. However, in practice, the model in (15) generates different yet sensible estimation of $\boldsymbol{\beta}$ than that generated by (10), which will be further illustrated in Section 6.

6 Experiments

We conduct experiments to empirically evaluate the proposed methods in two aspects:

- **Application**: we would like to demonstrate that our algorithms are not only able to rediscover drugs with *known* FBG-lowering indications (Section 6.2) but also able to identify *potential* drugs that might be repositioned to control blood sugar (Section 6.3).
- **Methodology**: we will show that the incorporation and regularization of time-varying baseline parameters in both BR and ABR models help to improve predictive performance compared with models without such parameters.

The time-invariant counterpart of BR can be derived by setting $\boldsymbol{t} = \boldsymbol{0}$ in (10), degenerating to an L_1 regularized linear model named Continuous Self-Controlled Case Series (CSCCS) in [Kuang *et al.*, 2016]:

$$\arg\min_{\boldsymbol{\beta}} \frac{1}{2} \left\|\boldsymbol{y} - \boldsymbol{Z}\bar{\boldsymbol{y}} - \left(\boldsymbol{X} - \boldsymbol{Z}\bar{\boldsymbol{X}}\right)\boldsymbol{\beta}\right\|_2^2 + \lambda_3 \left\|\boldsymbol{\beta}\right\|_1 . \tag{16}$$

One can also set $\boldsymbol{\gamma} = \boldsymbol{0}$ in (15), resulting in yet another L_1 regularized linear model named CSCCS for Adjacent records (CSCCSA) [Kuang *et al.*, 2016]:

$$\arg\min_{\boldsymbol{\beta}} \frac{1}{2} \left\|\boldsymbol{D}_\delta \boldsymbol{y} - \boldsymbol{D}_\delta \boldsymbol{X}\boldsymbol{\beta}\right\|_2^2 + \lambda_4 \left\|\boldsymbol{\beta}\right\|_1 . \tag{17}$$

Notice that the model in (17) is still time-varying, because the least square part of the objective function depends on δ. It essentially assumes that two adjacent records from the same patient within δ amount of time share the same FBG baseline and the difference between the two measurements is only dependent on their difference of drug exposure statuses. In our experiments, we set $\delta = 4$ years. We present our experimental results in details in subsequent sections.

6.1 Dataset

EHRs from Marshfield Clinic are used in our experiments. 64515 patients are admitted in the cohort with 219306 FBG measurements in total. 2980 drugs are considered in the experments.

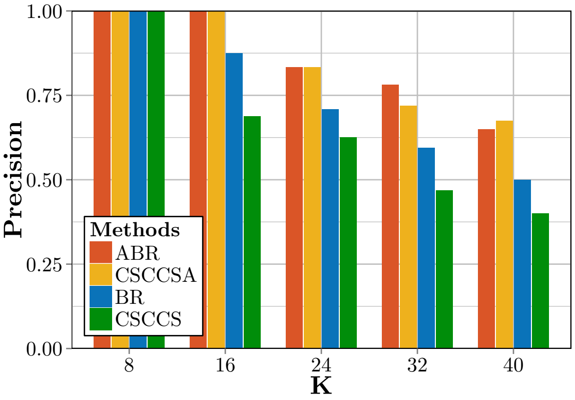

Figure 2: Precision at K among the top-forty drugs generated by the four models

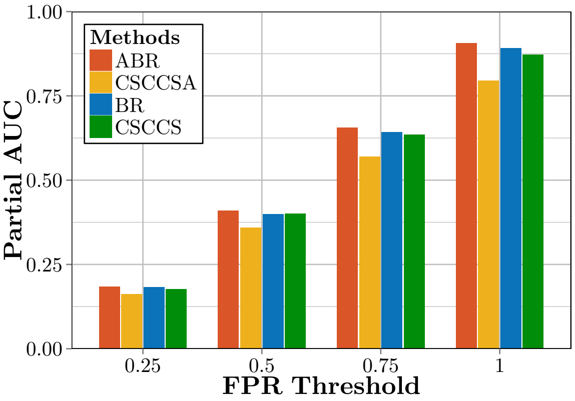

Figure 3: Partial AUCs on the top-forty drugs generated by the four models.

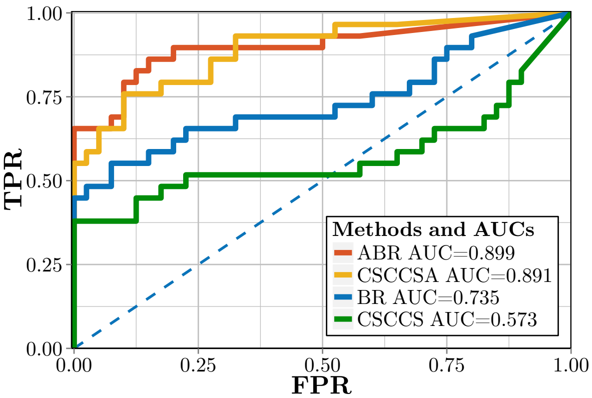

Figure 4: ROC curves on the union list of drugs generated by combining the top-forty drugs from each model

6.2 Rediscovering FBG-Lowering Drugs

Model Selection

Table 1: λ_1, λ_2 Used in (10)

INDX	λ_1	λ_2	# DRUGS
1	0.003923	291.141	205
2	0.003257	290.046	252
3	0.003923	228.185	197
4	0.003257	227.232	245
5	0.003923	177.043	196
6	0.003257	176.502	238
7	0.003923	161.502	194
8	0.003257	161.087	231

Table 2: Summary of Models Used in the Experiments

Name	Definition	Time-Varying
BR	(10)	Yes
ABR	(15)	Yes
CSCCS	(16)	No
CSCCSA	(17)	Yes

Table 1 summarizes the configurations of λ_1 and λ_2 of the BR model defined in (10) and their corresponding number of drugs selected in our experiments. The pairs of λ_1 and λ_2 are chosen in a way to ensure that approximately two hundred drugs or more are selected. We use Bayesian Information Criterion (BIC) [Zou *et al.*, 2007; Tibshirani and Taylor, 2011] to select the best model from the eight candidates and use the best model as a representative of BR. In this case, model 7 is selected and shaded in gray in Table 1. For this model, 194 drugs are selected. We choose λ_3 in (16) adaptively so that the number of drugs selected by (16) is also equal to 194. We use this model as the representative of CSCCS.

For the ABR model, we first define $r = \frac{\lambda_2}{\lambda_1}$, where λ_1 and λ_2 are defined in (15). We let r range from $1/16$ to 1 with an incremental step length of $\frac{1}{400}$, resulting in 376 different r's. We choose λ_1 (and hence λ_2 because $\lambda_2 = r\lambda_1$) adaptively so as to select two hundred drugs approximately for each r. Now that we have 376 pairs of distinct $\boldsymbol{\beta}$'s and $\boldsymbol{\gamma}$'s, one pair for each r, we use BIC again to select the best model among the 376 candidates. The minimum BIC is achieved when $r = 0.075$, and 201 drugs are selected in this case. We use this model as a representative of ABR. To compare ABR with its degenerated counterpart in (17), we tune λ_4 in (17) such that 201 drugs are also chosen by (17). We use this model as a representative of (17). After model selection, we have four representatives, one for each model. Each representative returns a list of approximately two hundred drugs. The four models used in our experiments are summarized in Table 2.

Top-Forty Drugs

We sort the drugs returned by each model in ascending order according to their coefficients that correspond to components in $\boldsymbol{\beta}$. We then manually label the top-forty drugs of each list, considering drugs with *known* FBG-lowering indications as positives while others as negatives. The labeling process is somewhat analogous to web search, where the search algorithm should return a short list of potential results for further human determination. In this scenario, the results at the top K positions might matter the most. We hence report the precision-at-K metric for each of the four models in Figure 2.

Table 3: Top-twenty potential drugs generated by BR

INDX	CODE	NAME	SCORE
1	2919	DIVALPROEX SODIUM	-1.818
2	8025	PROZAC	-1.073
3	7768	PREMARIN	-1.023
4	4616	HYDROXYCHLOROQUINE SULFATE	-0.883
5	5204	LANCETS	-0.873
6	1216	BLOOD SUGAR DIAGNOSTIC	-0.816
7	5822	METHOTREXATE SODIUM	-0.708
8	7963	PROPOXYPHENE NAP/ACETAMINOPHEN	-0.684
9	10392	ZOLOFT	-0.654
10	5636	MAVIK	-0.631
11	3471	ESTROGEN CON/M-PROGEST ACET	-0.583
12	5319	LEXAPRO	-0.570
13	467	AMITRIPTYLINE	-0.565
14	7831	PRILOSEC	-0.545
15	2367	COUMADIN	-0.474
16	2426	CYANOCOBALAMIN (VITAMIN B-12)	-0.446
17	3686	FERROUS SULFATE	-0.389
18	3646	FENOFIBRATE MICRONIZED	-0.381
19	6610	NORGESTIMATE-ETHINYL ESTRADIOL	-0.356
20	1455	CALCIUM CARBONATE/VITAMIN D3	-0.355

Table 4: Top-twenty potential drugs generated by ABR

INDX	CODE	NAME	SCORE
1	5204	LANCETS	-1.610
2	2919	DIVALPROEX SODIUM	-1.490
3	1216	BLOOD SUGAR DIAGNOSTIC	-1.344
4	5319	LEXAPRO	-1.232
5	5822	METHOTREXATE SODIUM	-1.086
6	2426	CYANOCOBALAMIN (VITAMIN B-12)	-0.715
7	8025	PROZAC	-0.604
8	3686	FERROUS SULFATE	-0.589
9	10392	ZOLOFT	-0.529
10	7768	PREMARIN	-0.526
11	7963	PROPOXYPHENE NAP/ACETAMINOPHEN	-0.502
12	9034	SULFAMETHOXAZOLE/TRIMETHOPRIM	-0.500
13	1455	CALCIUM CARBONATE/VITAMIN D3	-0.433
14	1200	BLOOD-GLUCOSE METER	-0.372
15	10215	WELLBUTRIN SR	-0.334
16	7831	PRILOSEC	-0.327
17	8073	PSYLLIUM SEED (WITH SUGAR)	-0.253
18	9531	TRAMADOL HCL	-0.214
19	7496	PLAVIX	-0.212
20	494	AMOXICILLIN/POTASSIUM CLAV	-0.209

In Figure 2, all four models return a reasonable amount of true positives at the top-forty positions, indicating their capabilities of identifying FBG-lowering drugs. We also notice that introducing and regulating time-varying baseline parameters substantially improve the precision of BR at all K thresholds, compared with CSCCS; while the integration of such parameters does not change the performance of ABR much in terms of precision, compared with CSCCSA. However, in Figure 3, we notice that ABR has a substantially higher Area Under Curve (AUC) than CSCCSA, dominating CSCCSA at every FPR threshold. This phenomenon suggests that making more intrinsic time-varying baseline assumptions in ABR might be potentially beneficial to classification performance.

Union List

We now consider generating a union list of drugs by combining all the top-forty drugs from the four models. The Receiver Operating Characteristics (ROC) curves of all models on the union list are reported in Figure 4. For each model, if a drug from the union list is not selected by the model, we consider it as an example that will be identified as positive at the every end, which results in the straight line parts of the curves in the liberal region. In Figure 4, BR dominates CSCCS at every FPR threshold, demonstrating the potential benefits of incorporating time-varying baseline information. Although the AUC of ABR is very close to that of CSCCSA, we notice that ABR has a higher partial AUC than CSCCSA in the conservative region, from which the operating points of most medical classifiers are usually selected.

6.3 Identifying Potential Drugs

We now turn to discuss identifying drugs that have potential glucose lowering effects. From experiments in Section 6.2, we notice the substantial improvement of rediscovery performance related to the introduction of the time-varying parameters. We therefore will focus on the results generated by the two models with such parameters, BR and ABR. For each of the two models, we follow the model selection procedure described in Section 6.2. However, we only select approximately one hundred drugs for each model and evaluate the potential of the top-twenty drugs in each drug list. We exclude all the records from patients that have any one of the FBG-lowering drugs discovered in Section 6.2, and run our algorithms only on the subset of data after exclusion. Table 3 and Table 4 summarize the findings from BR and ABR respectively, with gray rows representing drugs discovered by both methods. We evaluate the potential of each drug based on literature review in the subsequent sections.

Potential Decrease

In both Table 3 and Table 4, for propoxyphene nap and acetaminophen, a case has been reported relating to propoxyphene-induced hypoglycemia [Shah *et al.*, 2006]. Zoloft, which is a type of antidepressant, is related to increase of insulin level [Kesim *et al.*, 2011]. Lack of vitamin B12 is related to hyperglycemia in a rat model reported by [Chow and Stone, 1957], and diabetic patients with metformin prescriptions might be vitamin B12 deficient [Ting *et al.*, 2006].

In Table 3, Hydroxychloroquine sulfate is known to cause severe hypoglycemia [SanofiAventisCanadaInc., 2015]. Fenofibrate micronized might also be helpful to lower FBG level, based on the findings in [Damci *et al.*, 2003]. In Table 4, Sulfamethoxazole and trimethoprim, or Bactrim, is known to induce hypoglycemia [NIH, 2016]. Wellbutrin SR is another antidepressant that might be potentially beneficial to control blood sugar level [Lustman *et al.*, 2007]. Psyllium is also reported to improve glycemic control in a study with type 2 diabetes male patients [Anderson *et al.*, 1999]. A recent study suggests that tramadol HCl might increase the risk of hospitalization due to hypoglycemia [Fournier *et al.*, 2015]. Finally, Plavix is reported to cause hypoglycemia due to interactions with Prandin [GovernmentOfCanada, 2015].

Mixed Evidence

In both Table 3 and Table 4, divalproex sodium might increase blood sugar level, according to [DiabetesInControl, 2015]. Prozac could both increase or decrease blood glucose level [DiabetesInControl, 2015]. A hyperglycemia case has also been reported due to Lexapro [Zuccoli *et al.*, 2013]. However, Lexapro is a type of antidepressants called Selective Serotonin Reuptake Inhibitor (SSRI), which is linked to hypoglycemia in certain situations [Zammit, 2012]. The effects of calcium carbonate and vitamin D3 on blood glucose level have been widely studied, but mixed conclusions are reported [De Boer *et al.*, 2008; Mitri *et al.*, 2011;

Mitri and Pittas, 2014].

6.4 BR versus ABR

In Section 6.3, from Table 3 and Table 4, we notice that BR and ABR provide somewhat different potential drug lists. For both lists, we can find literature support to substantiate the blood sugar lowering potentials of many drugs. This experiment indicates that both methods might be viable options to aid the knowledge discovery process of drug repositioning. On the other hand, in Section 6.2, although experiments suggest that regulating time-varying baseline help to improve the rediscovery performance, ABR outperforms BR (and the other methods) in general in the rediscovery task. This observation points to several directions for future work that may extend beyond the CDR application.

First, since the submission of this paper, we have discovered an unpublished manuscript [Hess *et al.*, 2013] that describes a modeling framework similar to BR that is applied to econometric analysis. The empirical advantage we observe for ABR over BR in the present paper for one application of longitudinal data analysis, CDR, suggests investigating whether ABR enjoys a similar advantage for some econometric forecasting applications.

Second, CSCCS and CSCCSA are extensions–to predict continuous variables such as FBG, rather than binary variables–of self-controlled case series, arguably the current leading approach to adverse drug event discovery [Simpson *et al.*, 2013] from clinical data. Hence our results suggest that ABR might have advantages for ADE discovery as well. Testing this conjecture is a direction for further research.

More generally, an important direction for further work is the application of ABR to areas involving longitudinal data, especially when time intervals may be irregularly sampled as in the EHR data we use for CDR. Another direction for future research is to further validate the predictions of ABR for the CDR application.

7 Conclusion

We have introduced the baseline regularization model and its variant for the task of computational drug repositioning. Our proposed methods take into account the high dimensionality, irregularity, subject-heterogeneity and time-heterogeneity of longitudinal observational data and generalize the standard fixed effect model. Experimental results suggest that our methods are not only able to rediscover drugs with confirmed indications, but also able to identify drugs that might be potentially helpful to control FBG level. We therefore believe that the proposed methods can potentially aid the knowledge discovery process of drug repositioning.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive comments and advices. The authors would also like to gratefully acknowledge the NIH BD2K Initiative grant U54 AI117924, and the NIGMS grant 2RO1 GM097618.

References

[Anderson *et al.*, 1999] James W Anderson, Lisa D Allgood, Jan Turner, Peter R Oeltgen, and Bruce P Daggy. Effects of psyllium on glucose and serum lipid responses in men with type 2 diabetes and hypercholesterolemia. *The American Journal of Clinical Nutrition*, 70(4):466–473, 1999.

[Andronis *et al.*, 2011] Christos Andronis, Anuj Sharma, Vassilis Virvilis, Spyros Deftereos, and Aris Persidis. Literature mining, ontologies and information visualization for drug repurposing. *Briefings in Bioinformatics*, 12(4):357–368, 2011.

[Arnold *et al.*, 2014] Taylor B Arnold, Ryan J Tibshirani, Maintainer Taylor Arnold, and TRUE ByteCompile. Package genlasso. *Statistics*, 39(3):1335–1371, 2014.

[Chow and Stone, 1957] Bacon F Chow and Howard H Stone. The relationship of vitamin b12 to carbohydrate metabolism and diabetes mellitus. *The American Journal of Clinical Nutrition*, 5(4):431–439, 1957.

[Damci *et al.*, 2003] Taner Damci, Serkan Tatliagac, Zeynep Osar, and Hasan Ilkova. Fenofibrate treatment is associated with better glycemic control and lower serum leptin and insulin levels in type 2 diabetic patients with hypertriglyceridemia. *European Journal of Internal Medicine*, 14(6):357–360, 2003.

[De Boer *et al.*, 2008] Ian H De Boer, Lesley F Tinker, Stephanie Connelly, J David Curb, Barbara V Howard, Bryan Kestenbaum, Joseph C Larson, JoAnn E Manson, Karen L Margolis, David S Siscovick, et al. Calcium plus vitamin d supplementation and the risk of incident diabetes in the women's health initiative. *Diabetes Care*, 31(4):701–707, 2008.

[DiabetesInControl, 2015] 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 09/28/2015).

[Fakhraei *et al.*, 2013] Shobeir Fakhraei, Louiqa Raschid, and Lise Getoor. Drug-target interaction prediction for drug repurposing with probabilistic similarity logic. In *Proceedings of the 12th International Workshop on Data Mining in Bioinformatics*, pages 10–17. ACM, 2013.

[Fournier *et al.*, 2015] Jean-Pascal P Fournier, Laurent Azoulay, Hui Yin, Jean-Louis L Montastruc, and Samy Suissa. Tramadol use and the risk of hospitalization for hypoglycemia in patients with noncancer pain. *JAMA Internal Medicine*, 175(2):186–193, 2015.

[Frees, 2004] Edward W Frees. Longitudinal and panel data: analysis and applications in the social sciences. 2004.

[Friedman *et al.*, 2010] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

[GovernmentOfCanada, 2015] GovernmentOfCanada. Gluconorm (repaglinide) - new contraindication for concomitant use with clopidogrel - recalls & alerts - healthy canadians website. http://healthycanadians.gc.ca/recall-alert-rappel-avis/hc-sc/2015/54454a-eng.php, 2015. (Visited on 01/29/2016).

[Hess *et al.*, 2013] Wolfgang Hess, Maria Persson, Stephanie Rubenbauer, and Jan Gertheiss. Using lasso-type penalties to model time-varying covariate effects in panel data regressions–a novel approach illustrated by the'death of distance'in international trade. 2013.

[Kesim *et al.*, 2011] Murat Kesim, Ahmet Tiryaki, Mine Kadioglu, Efnan Muci, Nuri Ihsan Kalyoncu, and Ersin Yaris. The effects of sertraline on blood lipids, glucose, insulin and hba1c levels: A prospective clinical trial on depressive patients. *Journal of Research in Medical Sciences: the Official Journal of Isfahan University of Medical Sciences*, 16(12):1525, 2011.

[Ko *et al.*, 2006] Gary TC Ko, Hendena PS Wai, and Joyce SF Tang. Effects of age on plasma glucose levels in non-diabetic hong kong chinese. *Croatian Medical Journal*, 47(5):709–713, 2006.

[Kuang *et al.*, 2016] Zhaobin Kuang, James Thomson, Michael Caldwell, Peggy Peissig, Ron Stewart, and David Page. Computational drug repositioning using continuous self-controlled case series. *arXiv preprint*, 2016.

[Li and Lu, 2013] Jiao Li and Zhiyong Lu. Pathway-based drug repositioning using causal inference. *BMC Bioinformatics*, 14(16):1, 2013.

[Lustman *et al.*, 2007] Patrick J Lustman, Monique M Williams, Gregory S Sayuk, Billy D Nix, and Ray E Clouse. Factors influencing glycemic control in type 2 diabetes during acute-and maintenance-phase treatment of major depressive disorder with bupropion. *Diabetes Care*, 30(3):459–466, 2007.

[Mitri and Pittas, 2014] Joanna Mitri and Anastassios G Pittas. Vitamin d and diabetes. *Endocrinology and Metabolism Clinics of North America*, 43(1):205–232, 2014.

[Mitri *et al.*, 2011] Joanna Mitri, Bess Dawson-Hughes, Frank B Hu, and Anastassios G Pittas. Effects of vitamin d and calcium supplementation on pancreatic β cell function, insulin sensitivity, and glycemia in adults at high risk of diabetes: the calcium and vitamin d for diabetes mellitus (caddm) randomized controlled trial. *The American Journal of Clinical Nutrition*, 94(2):486–494, 2011.

[NIH, 2016] NIH. Drug-induced hypoglycemia: Medlineplus medical encyclopedia. https://www.nlm.nih.gov/medlineplus/ency/article/000310.htm, 2016. (Visited on 01/29/2016).

[O'Sullivan, 1974] John B O'Sullivan. Age gradient in blood glucose levels: magnitude and clinical implications. *Diabetes*, 23(8):713–715, 1974.

[SanofiAventisCanadaInc., 2015] SanofiAventisCanadaInc. Product monograph plaquenil. http://products.sanofi.ca/en/plaquenil.pdf, 2015. (Visited on 01/29/2016).

[Shah *et al.*, 2006] Pankaj Shah, Jarek Aniszweski, and John F Service. Propoxyphene-induced hypoglycemia in renal failure. *Endocrine Practice*, 12(2):170–173, 2006.

[Simpson *et al.*, 2013] Shawn E Simpson, David Madigan, Ivan Zorych, Martijn J Schuemie, Patrick B Ryan, and Marc A Suchard. Multiple self-controlled case series for large-scale longitudinal observational databases. *Biometrics*, 69(4):893–902, 2013.

[Tibshirani and Taylor, 2011] Ryan J. Tibshirani and Jonathan Taylor. The solution path of the generalized lasso. *The Annals of Statistics*, 2011.

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

[Ting *et al.*, 2006] Rose Zhao-Wei Ting, Cheuk Chun Szeto, Michael Ho-Ming Chan, Kwok Kuen Ma, and Kai Ming Chow. Risk factors of vitamin b12 deficiency in patients receiving metformin. *Archives of Internal Medicine*, 166(18):1975–1979, 2006.

[Tseng, 2001] Paul Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 109(3):475–494, 2001.

[Xu *et al.*, 2014] Hua Xu, Melinda C Aldrich, Qingxia Chen, Hongfang Liu, Neeraja B Peterson, Qi Dai, Mia Levy, Anushi Shah, Xue Han, Xiaoyang Ruan, Min Jiang, Ying Li, Jamii St Julien, Jeremy Warner, Carol Friedman, Dan M Roden, and Joshua C Denny. Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. *Journal of the American Medical Informatics Association*, pages amiajnl–ami2014, 2014.

[Zammit, 2012] Paul Zammit. Ssri-induced hypoglycemia causing confusion in a nondiabetic octogenarian. *Annals of Long-Term Care: Clinical Care and Aging*, 20:28–30, 2012.

[Zou *et al.*, 2007] Hui Zou, Trevor Hastie, Robert Tibshirani, et al. On the degrees of freedom of the lasso. *The Annals of Statistics*, 35(5):2173–2192, 2007.

[Zuccoli *et al.*, 2013] ML Zuccoli, PC Brasesco, G Milano, A Martelli, S Leone, F Mattioli, and C Fucile. A case report on escitalopram-induced hyperglycaemia in a diabetic patient. *The International Journal of Psychiatry in Medicine*, 46(2):195–201, 2013.