

# Arresting Treatment Patterns for Individual Patients in Clinical Big Data: An Exploratory Procedure

Mizuki Morita<sup>1</sup>, Masanori Shiro<sup>1</sup>, Shotaro Akaho<sup>1</sup>, Toshihiro Kamishima<sup>1</sup>, Hideki Asoh<sup>1</sup>  
Eiji Aramaki<sup>2</sup>, Koiti Hasida<sup>3</sup>, Takahide Kohro<sup>4,5</sup>

<sup>1</sup>National Institute of Advanced Industrial Science and Technology (AIST), Japan

<sup>2</sup>Kyoto University, Japan

<sup>3</sup>The University of Tokyo, Japan

<sup>4</sup>Jichi Medical University, Japan

<sup>5</sup>The University of Tokyo Hospital, Japan  
morita.mizuki@aist.go.jp

## Abstract

Data mining of clinical data that are stored continually in the course of daily medical practice will contribute to the advancement of healthcare. However, real-world clinical data are characteristically noisy, sparse, irregular, and biased, which makes it difficult to perform data mining. This study assesses an exploratory approach to ascertain how physicians tackle the worsening of a patient's condition using clinical data from a hospital. It yielded reasonable results.

## Introduction

A large-scale randomized controlled trial, which provides clinical evidence, is the current gold standard in evidence-based medicine. Although an overall trend can be ascertained from such trials, it is sometimes not trivial to decide how to apply the results to individual patients in daily medical practice. For example, evidence of treatment for patients in a severe condition generally can not be obtained from clinical trials because such patients are excluded from clinical trials by ethical and technical constraints, yet treatment tactics for such patients are exactly those which physicians in clinics are eager to know.

Clues for such treatment might be found in clinical records stored continuously from daily medical practice. Although clinical record data have shortcomings when applied to research (e.g., they are noisy, sparse, irregular, and biased), growth in data volume, which is associated with the prevalence of electronic health records, can usher in an era of Clinical Big Data, with new approaches to process data and to create evidence.

This study examined heart failure outpatient data. Treatment for patients with both heart and kidney failure is difficult, but treatment guidelines for such patients do not

exist. Therefore, physicians must plan treatments using a trial-and-error process (Shlipak and Massie 2004). A list showing how other physicians acted in similar cases to date would facilitate the selection of a treatment policy. Here, using stored clinical data, we propose an exploratory procedure for the efficient extraction of a clinical treatment of interest that was used by other physicians in the past.

## Materials and Methods

Clinical data used in this study were obtained, after institutional review board approval, from a medical database of heart failure outpatients at The University of Tokyo Hospital. Records of 302 patients with both heart and kidney failure were extracted from the database.

The change in laboratory test values for patients was used as an indicator of the change in patient status. The change in drug prescription patterns of physicians was used for the physician's response. The results of two laboratory tests were used to assess patient status: brain natriuretic peptide (BNP) for heart failure, and creatinine (CRE) for kidney failure. Extracted patients were those patients for whom both BNP and CRE had surpassed thresholds (300.0 for BNP, 1.5 for CRE) at least once. Prescription records of 52 drugs (chemical compounds) commonly used at this hospital in treatment for heart failure were applied to our analyses. These drugs fall into three major groups: antihypertensives, cardiotonics, and diuretics.

From time series outpatient clinical data, every adjacent time point was extracted. Their mutual differences in laboratory test values and prescription patterns were taken, which provides two variable spaces: change in laboratory test values and change in drug prescription patterns. Changes in laboratory test values were discretized into 16

Table 1: Selected changes in drug prescription when heart and/or kidney condition worsens from severe

(a) Prior <i>heart</i> status was <i>severe</i> and change was <i>worsen</i>			
Prior <i>kidney</i> status: <i>not severe</i>		Prior <i>kidney</i> status: <i>severe</i>	
Change: <i>not worsen</i>	Change: <i>worsen</i>	Change: <i>not worsen</i>	Change: <i>worsen</i>
decrease azosemide <sup>D</sup>		decrease ubidecarenone <sup>C</sup>	start triamterene <sup>D</sup>
start triamterene <sup>D</sup>		increase pimobendan <sup>C</sup>	end trichlormethiazide <sup>D</sup>
start lisinopril <sup>A</sup>		increase furosemide <sup>D</sup>	
end pimobendan <sup>C</sup>		start trichlormethiazide <sup>D</sup>	
start trichlormethiazide <sup>D</sup>		start spironolactone <sup>D</sup>	
increase furosemide <sup>D</sup>			
start docarpamine <sup>C</sup>			
(b) Prior <i>kidney</i> status was <i>severe</i> and change was <i>worsen</i>			
Prior <i>heart</i> status: <i>not severe</i>		Prior <i>heart</i> status: <i>severe</i>	
Change: <i>not worsen</i>	Change: <i>worsen</i>	Change: <i>not worsen</i>	Change: <i>worsen</i>
end digitoxin <sup>C</sup>		decrease furosemide <sup>D</sup>	start triamterene <sup>D</sup>
increase telmisartan <sup>A</sup>			end trichlormethiazide <sup>D</sup>

<sup>A</sup> = antihypertensives, <sup>C</sup> = cardiotonics, <sup>D</sup> = diuretics. Conditions on the right edge in both (a) and (b) are identical.

classes: (lab test: BNP | CRE)  $\times$  (prior status: *severe* | *not severe*)  $\times$  (change in status: *worsen* | *not worsen*). The thresholds in discretizing laboratory test values for the prior status were 300.0 for BNP and 1.5 for CRE. The thresholds in discretizing change in laboratory test values were the third quartile of slope values. The number of changes in drug prescribing patterns was 208: 52 drugs  $\times$  (change: start | end | increase | decrease).

The purpose of analysis here is to extract how physicians respond to the patient's worsening heart and/or kidney status. To put it more precisely, we sought to ascertain how physicians change a drug prescription when a patient *worsens* from *severe* status. To address this issue, investigation of the association between two variable spaces should be performed to ascertain the notable change in prescription to be explored. For this study, we used least absolute shrinkage and selection operation (LASSO) linear regression (Tibshirani 1996) for selecting variables (change in prescription), using change in laboratory test values as the response variable and change in drug prescription patterns as the explanatory variable.

To explore the validity of selected change in prescription, we assessed the original time series clinical data and discussed their interpretation with a physician specializing in cardiovascular internal medicine.

## Results and Discussion

Of all 208 changes in drug prescriptions, 90 were selected by LASSO as associated variables with change in patient status. We further analyzed 17 prescription changes associated with the patient status change of interest: *worsen* from *severe* status (Table 1).

For cases in which '*worsen*' from '*not severe*', where '*worsen*' from '*severe*' was found for another organ, no drug prescription change was selected (second column in Table 1). For these cases, it was assumed that physicians maintained their prescription with close monitoring. However, when the change is '*not worsen*' for one organ, several drug prescription changes were selected (first and third columns in Table 1).

Analyzing each prescription change produced clinically interesting cases that compelled physicians go into deep thought to choose a treatment policy. It also brought noisy results that were selected by LASSO, but which were not clinically interesting.

This study used only two adjacent data points in time series data to detect changes. Despite its simplicity, clinically interesting cases were detected. However, judging whether a selected case was interesting or merely noise required confirmation of the original time series data, suggesting that expansion to processing the history of laboratory data can improve this approach.

## Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 23240043, 25120011.

## References

- Shlipak, M. G., and Massie, B. M. 2004. The Clinical Challenge of Cardiorenal Syndrome. *Circulation*, 110:1514-1517.
- Tibshirani, R. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B* 58:267-288.