Data Mining of Pharmaceutical Interactions

T. Okada¹, T. Ishizaki¹, E. Ichiishi²

¹Department of Informatics, Kwansei Gakuin University, Sanda, Japan
²New Industry Creation Hatchery Center, Tohoku University, Sendai, Japan

Abstract - Pharmaceutical interactions between a pair of specific drugs are surveyed using clinical records. Faint signals are detected in four pairs of drugs. Among the four effector drugs, two proton pump inhibitors showed broad spectrum interactions with other drugs. The key point for the success of this analysis has been in the evaluation of generic effects to each drug by a physician. The criterion of \( BSS \) (between-groups sum of squares) has shown its usefulness in the analysis of contingency tables.

Keywords – Drug, interaction, data mining, \( BSS \).

I. INTRODUCTION

As the electronic medical chart is spreading, there appears a flood of data in the clinical world. The technology of data mining is expected to play an important role to discover useful medical knowledge in a variety of aspects, such as diagnosis, medical treatments, gene therapies and tailor-made pharmaceuticals.

The objective of this paper is to find pharmaceutical interactions from medical records. Serious results caused by drugs are reported to governmental offices, and are reflected to the content of drug package inserts. However, physicians often hear patients' complaints on the effects of prescribed drugs. Their causes may come from the inherited genetic factor, living habits or psychological reasons. But, pharmaceutical interactions are believed to be one of the major reasons reducing the therapeutic effect. However, the combinatorial explosion in the number of drug pairs prohibits the analysis by physicians. Nonetheless, the detection of pharmaceutical interactions is an important subject to improve the quality of therapeutics.

Our research team consists of a physician (E.I.) and data miners, and we planned to detect faint signals of pharmaceutical interactions from a relatively small amount of clinical data. Patients visiting a hospital for the old are the source of data, as they usually have multiple symptoms and take many kinds of drugs simultaneously. In order to catch a faint signal, the effect of drugs were carefully coded and utilized as the guide to detect interactions. We succeeded to detect four signals from pairs of specific drugs. Further analysis led us to find the broad spectrum interactions caused by two proton pump inhibitors.

The next section of this paper introduces the contents of the source data collected, its preprocessing scheme as well as the selection of affected drugs. The method of analysis is explained and the resulting interactions are discussed in Sect. III. The last section summarizes the results and the task for the future.

II. DATA AND SELECTION OF AFFECTED DRUGS

A. Source Data and Preprocessing

An author (E.I.) collected medical records of patients who visited a hospital located in Kyoto prefecture, Japan, at the end of 2002. As the hospital is for the old, almost all patients suffer from multiple lifestyle-related diseases, and they need to take many sorts of drugs.

The source data consists of the following six tables.

1) Sample table: This is the master table which contains 125 patient records. Each record consists of age, sex, living habits and comments by the physician. The habits contain the amount and the term of drinking and smoking as well as the foods that a patient likes and dislikes. Also contained is a blood sample ID, which is to be used after the genetic test of samples. In the preprocessing step, ages are categorized to every 10 years, and the amount of drinking and smoking are categorized to 4 levels.

2) Blood test table: It contains the results of the blood test. 24 biochemical items as well as the blood pressure data are included. All data items are converted to 3 levels (normal, high, low) using the standard values for Japanese people.

3) Sick history table: This table contains disease information for patients. Each record has a patient ID and a disease ID as well as the starting date of the symptom. One patient has about 5 diseases on average, and the maximum number reaches 10.

4) Drug history table: It shows the drugs taken by patients. Each record has a patient ID and a disease ID as well as the starting date of the symptom. One patient has about 5 diseases on average, and the maximum number reaches 10.

5) Blood test table: It contains the results of the blood test. 24 biochemical items as well as the blood pressure data are included. All data items are converted to 3 levels (normal, high, low) using the standard values for Japanese people.

6) Sick history table: This table contains disease information for patients. Each record has a patient ID and a disease ID as well as the starting date of the symptom. One patient has about 5 diseases on average, and the maximum number reaches 10.

4) Drug history table: It shows the drugs taken by patients. Each record has a patient ID, a commercial name of a drug, its therapeutic effects as well as the amount and the starting date of its dose. One patient has 5-6 drugs on average, while the maximum goes to 18. The subjective and objective effects of a drug were graded according to ten ranks by the patient and by a medical check, and they are summarized to give a generic effect, T or F, by a single physician. Since individual evaluation for each drug is hard,
the effects of drugs for a single patient often take the same value. However, different ranks and generic effects are given to drugs, when it is possible to evaluate them separately. This generic effect has been used as the guide to the efficacy of a drug. The resulting generic effects contained 120 F values among 703 doses.

5) **Drug table**: This table denotes a generic name of a drug for each commercial drug. A list of metabolic enzyme names is also included. There appear 143 commercial drugs and 132 generic names in the table. Currently, no grouping of drugs is done based on therapeutic indications, and the generic names are used in the analysis.

6) **Disease table**: A disease ID and its name constitute a record of this table. There appear 173 kinds of diseases. Currently, no grouping of diseases is employed.

### III. DETECTION OF PHARMACEUTICAL INTERACTIONS

#### A. Detection of Interactions from Specific Drugs

In the research of pharmaceutical interactions, we supposed that the generic effect of Drug A changes by the coadministration of Drug B. We call Drug A an affected drug (AD), while Drug B is named an effector drug (ED). The following 2x2 contingency table was created to check the strength of the interaction.

<table>
<thead>
<tr>
<th>AD</th>
<th>Nicardipine Hydrochloride</th>
<th>BSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED</td>
<td>take</td>
<td>T</td>
</tr>
<tr>
<td>Manidipine Hydrochloride</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>19</td>
</tr>
<tr>
<td>Sum</td>
<td>20</td>
<td>4</td>
</tr>
</tbody>
</table>

\[ \chi^2: 11.8, \ p: .0006 \]

This sample table shows an AD-ED pair with the strongest interaction. When a patient is coadministrated by manidipine hydrochloride, the percentage of (generic effect = T) for (AD = nicardipine hydrochloride) is 25%, which is much smaller compared with the observation of 95% in cases of no coadministration.

There appear 3 measures to show the strength of the interaction in Table II. The first two are well-known chi-square statistic and its \( p \) value shown below the table. The last is BSS values for each group and their sum shown at the right of the table. They are calculated using the following formulae [1].

\[
BSS_g = \frac{n_g}{2} \sum_a \left( p_g(a) - p_0(a) \right)^2 \tag{1}
\]

\[
BSS = \sum_g BSS_g \tag{2}
\]

where \( n_g \) denotes the number of samples in group \( g \), and \( p_g(a) \) and \( p_0(a) \) show the probability of taking the value \( a \) in group \( g \) and before grouping, respectively.

BSS value shows the between-groups sum of squares for categorical data. In this case, total sum of squares from the generic effects of AD for 24 doses is computed to be 3.33, in which 1.63 (1.36 + 0.27) is explained by the between-groups sum of squares when the samples are divided into two groups by the presence/absence of the ED. The rest of sum of squares remains as within-group sum of squares of the two groups. BSS value for (ED take = Y) accounts for more than 80% of the sum of BSS, and we can judge that the distribution of generic effects moves away more distinctly in this group.

Our survey of EDs is limited to drugs used for more than 5 patients. They are shown in Table III.
Examination of contingency tables detected a few pairs of interacting drugs. Table IV shows seven AD-ED pairs with the descending order of BSS values for (ED\_take = Y) group.

Pretty large BSS values are observed in top four pairs, and the contingency tables for these pairs are shown in Table V. Here, four tables at the top show those constructed from AD-ED pairs in Table IV, where BSS values are denoted below each table. The pharmaceutical interactions might happen in these pairs even if the roles of ADs and EDs are interchanged. Therefore, four tables interchanging AD and ED are also shown at the bottom.

All statistics indicate that these four AD-ED pairs have pretty strong interactions, but no interactions are observed when the roles of ED and AD are interchanged. Since the number of interacting cases is very limited in all these pairs, we cannot draw a decisive conclusion, but we think that a
I. INTRODUCTION

Long-term monitoring is worth in these pharmaceutical interactions.

B. Analysis from a Viewpoint of Effector Drugs

Most pharmaceutical interactions are supposed to arise from the inhibition or the induction of cytochrome P450 (CYP450) enzymes in the process of hepatic metabolism. The preceding analysis has pointed out the possibility of these metabolic phenomena by the administration of 4 kinds of EDs. If these metabolic effects are going on, these EDs might affect the generic effects of many other drugs.

We surveyed the generic effects of all drugs, which are simultaneously administrated with these 4 EDs. The resulting contingency tables are shown in Table VI.

Here, we can observe that all statistic measures indicate the stronger effects by omeprazole and lansoprazole compared with those in Table V, while the interactions become weaker in manidipine hydrochloride and atorvastatin. Two drugs with strong generic effects are proton pump inhibitors, and they have similar chemical structures. In case that simultaneous administration of other drugs is necessary, lansoprazole can be said to have a relatively smaller generic effect, if one is forced to compare these two drugs.

IV. CONCLUDING REMARKS

We have succeeded to detect faint signals of pharmaceutical interactions from a small clinical dataset. They consist of four pairs of specific drugs as well as two broad spectrum effects by proton pump inhibitors.

The method of analysis is not so different from usual statistical ones, though we introduced BSS criterion that has been used in data mining [2, 3].

The principal reason of this success is the attribute of the generic effect in source data. Using this objective variable as the guide of survey, we could reach meaningful results. But, the evaluation task for the drug necessitates a lot of efforts by the physician. Further, we cannot get such data, if patients refrain to criticize the therapeutic effect of drugs prescribed by a physician. Therefore, extensive survey of this kind of clinical records is prohibitive, and we will need a new technique which enables the detection of generic effects from other source of data.

ACKNOWLEDGMENT

T.O. & T.I. thanks Dr. Masumi Yamakawa of Kwansei Gakuin University for his helpful suggestions about drug research.

REFERENCES


TABLE VI

<table>
<thead>
<tr>
<th>ED</th>
<th>Generic Effect</th>
<th>take</th>
<th>T</th>
<th>F</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omeprazole</td>
<td>Y</td>
<td>182</td>
<td>39</td>
<td>221</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>370</td>
<td>71</td>
<td>441</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>552</td>
<td>110</td>
<td>662</td>
<td></td>
</tr>
</tbody>
</table>

χ²: 0.25, p: 0.61,
BSS: 0.023+0.012

<table>
<thead>
<tr>
<th>ED</th>
<th>Generic Effect</th>
<th>take</th>
<th>T</th>
<th>F</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atorvastatin</td>
<td>Y</td>
<td>24</td>
<td>9</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>556</td>
<td>108</td>
<td>663</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>580</td>
<td>117</td>
<td>697</td>
<td></td>
</tr>
</tbody>
</table>

χ²: 2.73, p: 0.099,
BSS: 0.36+0.02

<table>
<thead>
<tr>
<th>ED</th>
<th>Generic Effect</th>
<th>take</th>
<th>T</th>
<th>F</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lansoprazole</td>
<td>Y</td>
<td>24</td>
<td>18</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>553</td>
<td>100</td>
<td>653</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>577</td>
<td>118</td>
<td>695</td>
<td></td>
</tr>
</tbody>
</table>

χ²: 21.2, p: < 0.0001,
BSS: 2.81+0.18

<table>
<thead>
<tr>
<th>ED</th>
<th>Generic Effect</th>
<th>take</th>
<th>T</th>
<th>F</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omeprazole</td>
<td>Y</td>
<td>31</td>
<td>29</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>545</td>
<td>88</td>
<td>633</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>576</td>
<td>117</td>
<td>693</td>
<td></td>
</tr>
</tbody>
</table>

χ²: 46.3, p: < 0.0001,
BSS: 5.83+0.56