

A Proposal of Contraindication Database for Medicines

Ryo Okuya¹, Hirotsugu Ishida², Keita Nabeta², Masaomi Kimura¹,
Michiko Ohkura¹, and Fumito Tsuchiya³

¹ Shibaura Institute of Technology,

3-7-5 Toyosu, Koto Ward, Tokyo 135-8548 Japan

² Graduate School of Technology, Shibaura Institute of Technology,

3-7-5 Toyosu, Koto Ward, Tokyo 135-8548 Japan

³ International University of Health and Welfare,

2800-1 Kitakanemaru, Ohdawara City, Tochigi 324-8501 Japan

{107035, masaomi, ohkura}@shibaura-it.ac.jp,

{m110013, m709102}@shibaura-it.ac.jp, ftsuchiya@iuhw.ac.jp

Abstract. In recent years, despite various measures taken to reduce medical accidents as a result of confusions over drugs, cases of medical malpractice have occurred in Japan. As a countermeasure supported by a Health Labor Sciences Research Grant in 2009, drug information databases based on drug package inserts have been created for computer systems to prevent accidents caused by incorrect treatment of drug information [1]. However, the data in the databases remains problematic. In this study, we propose data item sets to be defined in drug information databases.

Keywords: Medical safety, Package inserts, Databases.

1 Introduction

Safe usage of medicines is one of the keys to preventing medicine-related accidents. One attempt to increase safety is a device to provide medicine-related data for doctors to verify whether the selection of medicines is appropriate for the purpose of treatment. In Japan, computerized prescriber order entry (CPOE) systems are widely used to prescribe medicines, and there have been many efforts to prevent the wrong input of medicine names: improved order in the list of medicines, highlighting frequently confused medicine names by adding certain symbol characters and so forth. In addition to such efforts, it is important to provide suitable medicine-related information for doctors in order to recognize the selection errors of medicines. With this in mind, we have conducted studies to create databases containing the information required for a CPOE prescription checking system.

In 2009, such databases for Japanese medicines were proposed with the support of a Health Labor Sciences Research Grant. The databases consist of tables such as the one storing contraindication data. The contraindication data is one of the key data described in package inserts to prevent the occurrence of fatal medical accidents. In spite of their effort to define table schema and their importance, the data therein are less structured and difficult to utilize in CPOE. This is because they inherit

characteristics of original package inserts, namely, the diversity of structures and expressions of descriptions. Original package insert data, though their contents are officially defined by the authorities, contain information expressed as sentences, tables, figures without strict constraints of expressions. Since the package inserts are an official source of medicine information, the contraindication part of the database is made based on them and are described in sentences or phrases. For such less constructed data, we need to employ a full-text search technique to find whether contraindication data include some keywords. Imagine the case that some medicine contraindicates use in ‘infants two months or younger’. If you confirm that the medicine has a contraindication to infants, you have to compare the string ‘infant’ with each word in the contraindication phrase. In order to search an exact match of keywords effectively, we have to turn the data into fully-structured ones based on code systems. The current version of the databases does not satisfy this requirement.

In this study, we decompose the phrases and the sentences in the contraindication part of the database into fundamental sets of keywords and group them to design code systems. For some tables, we decompose them by hand. Additional to these analyses, we also investigate the correspondence of terms defined in some other master data, e.g. ICD-10, the master data of diseases [2]. We separate the terms in the master data from the contraindication data and define the new attributes corresponding to the remaining parts. Based on the results of these analyses, we propose a database schema that is necessary for storing (structured) contraindication data.

2 Target Data

Our target databases are those developed with the support of a Health Labor Sciences Research Grant as outlined below.

- Contraindicated patient-type database
 - The database of contraindication descriptions related to patients such as elderly people, pregnant women, nursing women and babies (**Table 1**). The data are the summaries of precautionary statements and contraindication statements in package inserts that are related to such patients and are described in a few phrases. The databases are problematic, since they lump many kinds of contraindication information together. This occurs because various aspects of contraindication remain to be unarranged. For example, there exists the data ‘babies suffering from exomphalos’ and ‘babies two years or younger’. The former is the condition related to disease and the latter is related to age. Additional to this, plural conditions can co-exist in one record representing one medicine. In order to realize a search for an exact match, the data first need to be turned into normal form, namely, be broken into separate records for each condition. Several different spellings of terms are also problematic.
- Contraindicated disease database
 - The database of contraindication descriptions related to patients suffering from a certain disease (**Table 2**). The data are the summaries of contraindication statements in package inserts that focus on diseases except for their reasons and

are described in a few phrases. The problem in this database is that not only disease/symptom names but also other information, such as administration duration of other medicines (e.g. phrases like ‘during administration of catecholamine’) are contained in it, though it is intended to be a list of disease/symptom names. This occurs because the administration information is indirectly related to the disease information. As is in the contraindicated patient database, various kinds of contraindication information are also lumped together and several different spellings of terms co-exist.

Table 1. Sample data in the contraindicated patient-type database

Product names	Baby-related information	Nursing-woman-related information
Mercurochrome liquide	Babies suffering from exomphalos	
Selapina granule	Babies 2 years or younger	Nursing women (long-term continuous application)
Diapp suppository 4	Low-birth-weight baby, newborns	

Table 2. Examples of data in the contraindicated disease database

Product names	Contraindicated disease names
Prorner tablet 40	bleed
Junsi Ephedrine hydrochloride	during administration of catecholamine
Dilute iodine tincture. OY	iodine hypersensitivity
Calvital	hypersensitivity to iodine

3 Methods

3.1 Analysis of Contraindicated Patient-Type Database

First, we manually split the data stored in the database into phrases whose meaning can be interpreted. Then, we classified the phrase into a set of words representing each meaning. Phrases are split manually because words obtained by application of techniques such as morphological analysis often cannot be interpreted as medical information. **Fig. 1** shows an example of how to split the data. We classified the words with confirming the package inserts published by the Pharmaceuticals and Medical Devices Agency [3]. Since there are different expressions in the database, though they have the same meaning, we collected such expressions to create a dictionary and to standardize them.

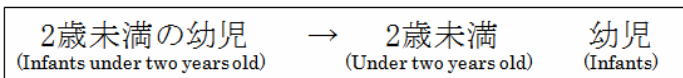


Fig. 1. An example of the data split

3.2 Analysis of the Contraindicated Disease Database

Since some data stored in this database is not a standard disease, we investigated whether the data match a disease name in standard disease master data published by The Medical Information System Development Center. We classified each of the data into four types of matching. The first type is data exactly matching one of the disease names in the master data, and is called *exactly matched*. The second type is data matching a combination of two or more words that are disease names or their modifiers in the master data, called *semi-exactly matched*. **Table 3** shows samples of the data in this type. The third type is data only whose substring matches a disease name or its modifier in the master data, called *partially matched*. The fourth type is data that are not applicable to these types, called *unmatched*.

We analyzed data in each type as follows.

We confirmed whether the exactly matched data is actually a disease name.

We investigated what the additional information to the semi-exactly matched or partially matched data is other than disease name.

Since the *unmatched* data do not match any disease names in the master, we investigated what information the data contain.

Table 3. Samples of the *semi-exactly matched* data

A phrase in data	Substrings matching the master data
急性冠動脈疾患 (Acute coronary artery disease)	急性(Acute)
	冠動脈疾患(Coronary artery disease)
重症糖尿病 (Severe diabetes)	重症(Severe)
	糖尿病(Diabetes)
特発性血管腫瘍 (Idiopathic vascular tumor)	特発性(Idiopathic)
	血管(Vascular)
	腫瘍(Tumor)

4 Results

4.1 Results of the Analysis of Contraindication Patient-Type Database

By classification of the phrases in the data in this database, we found that they contain not only information about medicine administration and the condition of patients but also the words/phrases representing patient categories (**Table 4**). We also found that the information about medicine administration can be categorized as either prohibited method, duration, body parts, dosage, purpose or frequency, and that the information about medicine administration can be categorized as either prohibited method, duration, the body parts, dosage, purpose or frequency. Moreover, the information was also categorized as either reason or condition of prohibited administration. The information about patient conditions was classified into prohibited age, duration, state, and disease

Table 4. Sample of patient categories

Patient category	Words/phrases	
高齢者(Geriatric)	高齢者(Geriatric)	老齡(Old age)
授乳婦(Nursing woman)	授乳婦(Nursing woman)	授乳中(Nursing)
乳幼小児 (Babies and children)	小児(Child)	幼児(Infant)
	新生児(Neonate)	患児(Diagnosed infant)
妊婦(Pregnant women)	妊婦(Pregnant women)	産婦(Parturient women)
	妊娠中 (During pregnancy)	妊娠又は妊娠している可能性 (Pregnant or possibly pregnant)

Table 5. Samples of the information about medicine administration

Categories	Phrases	
方法(Method)	密封法 (Occlusive dressing technique)	点滴静注 (Intravenous infusion)
期間(Duration)	長期 (A long term)	長期間 (A long term)
部位(Body parts)	指(Finger)	子宮腔内 (The insides of vagina of womb)
投与量(Dosage)	大量 (Large dose)	過量投与 (Overdose)
目的(Purpose)	広範なテスト(Extensive test)	子宮卵管撮影 (Hysterosalpingography)
頻度(Frequency)	頻回 (Frequent)	
理由(Reason)	ヒスタミン遊離 (Histamine release)	生殖発生毒性試験 (Reproduction test)
条件(Condition)	リバビリンとの併用 (Ribavirin combination)	

Table 6. Samples of the conditions of a patient

Categories	Phrases	
年齢(Age)	生後3カ月未満 (Under three months of age)	2歳以下 (Two years or younger)
期間(Duration)	3カ月以内 (Within three months)	12週未満 (Less than twelve weeks)
	初期 (Early phase)	末期 (Last phase)
状態(State)	胎位異常 (Fetal malpresentation)	継続が危険 (Continuation is dangerous)
病名(Disease name)	腸アトニー (Intestinal atony)	高度黄疸 (Severe jaundice)
	高ビリルビン血症 (Hyperbilirubinemia)	肺動脈閉鎖 (Pulmonary atresia)

Table 7. A List of sets of synonyms

Combination of the similar phrases	
乳幼児(Infants)	乳・幼児(Infants)
妊婦又は妊娠している可能性 (Pregnant or possibly pregnant)	妊娠又は妊娠している可能性 (Pregnant or possibly pregnant)
妊婦及び妊娠している可能性 (Pregnant and possibly pregnant)	妊娠又は妊娠の疑われる (Pregnant or doubt of pregnant)

name. **Table 5** and **Table 6** show samples of the information about medicine administration and the condition of the patient and their categories. **Table 7** shows a list of synonym sets. The phrases in the same cell are synonyms that have similar meanings such as ‘妊娠’(pregnant) and ‘妊婦’(pregnant woman), ‘可能性’(potential) and ‘疑われる’(in doubt). Such combinations must be unified in a word.

We propose data items that should be included in a table schema based on the classification results and combination of similar phrases. The data items are as follows:

- A patient category
- An administration method
- An administration period
- A target body part
- Dosage
- Frequency of administration
- The purpose of administration
- The reason for contraindication
- Patient age
- Patient state
- Duration of patient state
- A disease name

Additional to the definition of these data items, we created a master table for each of them, each of whose data has ID that is identical to their meanings but whose branch number indicates difference of expression (**Table 8**). Based on these data items and codes, we created a database whose samples are shown in **Table 9**. In order to avoid sparse assignment of data in each record, the column named *Contraindication condition codes* stores the data in items other than patient categories.

In order to evaluate the data items newly proposed in this study, we randomly selected 1,000 out of 7,152 records in the contraindicated patient database and confirmed that the table can store the information included in them without omission.

We also found that our schema do not admit a query with multiple conditions to match, since it assumes each piece of contraindication information is dealt with independently. From the viewpoint of safety, this might be sufficient to cover necessary contraindication information, since a database user should decide whether he/she adopts its restrictive conditions.

Table 8. Sample codes related to contraindicated patient types

Codes	Contraindicated patient types
A001-1	高齢者 (geriatric)
A002-1	老齡 (old age)
B001-1	授乳婦 (nursing women)
B001-2	授乳中 (nursing)
B001-3	授乳期 (lactation)
C001-1	小児 (a child)
C002-1	新生児 (an infant)
C003-1	低出生体重児 (low-birth-baby)
C003-2	未熟児 (immature baby)

Table 9. Sample data in the resultant database

YJ code	HOT9 code	Patient cat.	Contraindication condition codes
1119401A1036	100316101	C001-1	P1005-1 (2 years or younger)
1119401A1036	100316101	C004-1	P1005-1
7219413A1023	111867401	A001-1	null

4.2 Results of Analysis of the Contraindicated Disease Database

Since all of the exactly matched data were disease names, we can deal with them as *contraindicated disease names*.

Semi-exactly matched data were found to contain not only disease names but also their additional information such as the extent of the disease, frequency, body parts, and the conditions of a patient, as is shown in **Table 10** and **Table 11**.

Table 10. Samples of information that include the extent of disease

Phrases	Disease names	The extent of disease
軽症熱傷(mild burn)	熱傷(Burn)	軽症(Mild)
高度肝障害(Severe hepatopathy)	肝障害(Hepatopathy)	高度(Severe)

Table 11. Samples of information that include the conditions of a patient

Phrases	Disease names	The Conditions of a patient
妊娠ヘルペス (Gestational herpes)	ヘルペス(Herpes)	妊娠(Pregnant)

The partially-matched data and the unmatched data also contained additional disease information such as the extent of diseases, body parts, the conditions of a patient, reactive drugs, medical devices, treatments and related diseases. Some data did not contain disease but the condition of patients, reactive drugs and treatments. We should note their categories, and they must be dealt with as distinct items in the database. Their samples are shown from **Table 12** to **Table 14**.

Table 12. Samples of reactive medicines

Phrases	Timings	Drugs reactive with the target
経口避妊薬服用中 (During taking oral contraceptive pill)	服用中 (During taking)	経口避妊薬 (Oral contraceptive pill)
MAO阻害剤投与中 (During taking MAO inhibitor)	投与中 (During taking)	MAO阻害剤 (MAO inhibitor)

Table 13. Samples of treatments

Phrases	Treatments
血液透析中(During hemodialysis)	血液透析(Hemodialysis)
PUVA療法実施中(During PUVA therapy)	PUVA療法(PUVA therapy)

Table 14. Samples of the conditions of patients

Phrases	The condition of a patient
白血球数2,000 / mm ³ 以下 (The white blood cell count is 2,000 per cubic millimeter or less)	白血球数2,000 / mm ³ 以下 (The white blood cell count is 2,000 per cubic millimeter or less)

We propose the data items to describe contraindicated diseases as follows:

- Drug ID
 - YJ code
 - HOT9 code
 - A product name
- Contraindication information
 - A contraindicated disease name
 - A contraindicated patient state
 - A contraindicated treatment
 - A contraindicated drug reactive with the target drug
- Additional information
 - Details of a disease
 - The extent of a disease
 - The phase of a disease
 - Reactive drugs
 - A body part
 - A treatment
 - A medical device
 - A causative disease and/or symptom

- A causative action
- A disease not eligible
- An accompanying disease

5 Conclusion

In this study, in order to contribute to designing databases for a CPOE prescription checking system, we investigated the databases for Japanese medicines storing contraindication data. The target databases were the contraindicated patient database, the contraindicated disease database and the contraindicated drug combination database.

As for the contraindicated patient-type database, we confirmed that many kinds of contraindication information related to patients are lumped together in it and that several different spellings of terms exist therein. We also found that we can decompose and sort the contraindication information in each data into items such as medicine administration information, patient-related conditional information and patient categories. The medicine administration information is categorized into an administration method, the body part to be applied and quantity. The patient-related conditional information consists of age, disease names and so forth.

Although the contraindicated disease database is expected to be a list of contraindicated disease/symptom names of each medicine, it also includes medical treatment, patient status and mutual interaction between medicines. As for contraindicated disease/symptom names, we found that there is additional information such as the parts of the body, the extent and period of disease and so forth. We defined data items to store these and also created master data for each kind of information.

Since the data in these databases are described in natural language, there is variance in the expressions of terms. In order to unify the expressions, we identified terms that have the same meaning and defined a coding scheme to reflect the identification on.

In the future, we will extend our database schema to enable processing of more flexible queries by reflecting the correspondence relationships between pieces of contraindication information. We will also implement contraindication databases other than the contraindicated patient database.

References

1. Fumito, T., Masaomi, K.: Studies on the utilization of computerized drug package inserts. Summary Research Report 2009, Health Labor Sciences Research Grant (2009)
2. The Medical Information System Development Center, <http://www.medis.or.jp/>
3. Pharmaceuticals and Medical Device Agency, <http://www.info.pmda.go.jp/>