Journal of Biomedical Informatics 58 (2015) 235–246

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Rough-set-based ADR signaling from spontaneous reporting data with missing values

Wen-Yang Lin [a,*], Lin Lan [a], Feng-Hsiung Huang [a,b], Min-Hsien Wang [a]

[a] Dept. of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung, Taiwan
[b] Dept. of Pharmacy, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan

ARTICLE INFO

Article history:
Received 30 April 2015
Revised 13 September 2015
Accepted 22 October 2015
Available online 29 October 2015

Keywords:
Adverse drug reaction
Missing data
Pharmacovigilance
Rough set theory
Spontaneous reporting data

ABSTRACT

Spontaneous reporting systems of adverse drug events have been widely established in many countries to collect as could as possible all adverse drug events to facilitate the detection of suspected ADR signals via some statistical or data mining methods. Unfortunately, due to privacy concern or other reasons, the reporters sometimes may omit consciously some attributes, causing many missing values existing in the reporting database. Most of research work on ADR detection or methods applied in practice simply adopted listwise deletion to eliminate all data with missing values. Very little work has noticed the possibility and examined the effect of including the missing data in the process of ADR detection.

This paper represents our endeavor towards the exploration of this question. We aim at inspecting the feasibility of applying rough set theory to the ADR detection problem. Based on the concept of utilizing characteristic set based approximation to measure the strength of ADR signals, we propose twelve different rough set based measuring methods and show only six of them are feasible for the purpose. Experimental results conducted on the FARES database show that our rough-set-based approach exhibits similar capability in timeline warning of suspicious ADR signals as traditional method with missing deletion, and sometimes can yield noteworthy measures earlier than the traditional method.

1. Introduction

Adverse Drug Reactions (ADRs) are uncomfortable or harmful reactions (side effects) in normal doses of drug usage. In other words, an ADR expresses the association between drugs and harmful side effects. Some serious ADRs may lead to death or life-threatening outcomes of patients. For example, in 1950 the new drug Thalidomide made in German caused more than 12,000 fetus limb deformities and more than 1300 people were suffering from polyneuritis for over 20 countries in Europe and Japan. Unfortunately, not all ADRs can be disclosed before the approval of drugs for marketing. Therefore, spontaneous reporting systems (SRSs) of adverse drug reactions have been widely established in many countries to collect as could as possible all adverse drug events to facilitate the detection of suspected ADR signals via some statistical or data mining methods.

Although different SRSs were running under different reporting regulations, most of them require, when the patient produces uncomfortable or harmful adverse reactions by normal drug of usage, the responsible hospitals, related pharmaceutical companies should, and/or the patient himself can report the events to SRSs. Unfortunately, the reporting data usually contain some missing values due to omitting or personal privacy concern. Data with missing values may affect results of analysis, which leads to the development of appropriate processing methods to increase the accuracy of signal detection.

Most of the reporting systems use listwise deletion [12] to process data with missing values; that is, simply deleting records with null values to maintain data completeness. The advantage of this simple method is easy to implement for data analysis, while it may affect the accuracy of the results, especially when the amount of data is relatively small. Indeed, small non-missing data is not uncommon for ADR reporting data. Firstly, records of rarely used or newly marketed drugs usually are in small amount. Secondly, the data size also decreases significantly when stratified signal detection is performed [13], e.g., considering a specific group of patients with dedicated age and/or sex. Although previous research work has shown that rough set theory can be used to handle data with missing values in the process of data analysis [11,17], e.g., data classification, there has been no work, to the best of our knowledge, conducted on applying rough set theory to the ADR

* Corresponding author.
E-mail addresses: wylin@nuk.edu.tw (W.-Y. Lin), m1015516@mail.nuk.edu.tw (L. Lan), ffhung@vghks.gov.tw (F.-H. Huang), m1025509@mail.nuk.edu.tw (M.-H. Wang).

http://dx.doi.org/10.1016/j.jbi.2015.10.013
1532-0464/© 2015 Elsevier Inc. All rights reserved.

detection problem. This motivates us to study if incorporating rough-set-based strategies to process the reporting data with missing values can be helpful for the detection of ADR signals.

In this paper, we present the concept of applying rough set theory to handling ADR detection from incomplete SRS dataset with missing data, and propose twelve different methods for measuring the strength of an ADR signal. We discuss the feasibility of the proposed twelve measuring methods and show only six of them are suitable for ADR signal measuring. We conducted preliminary experiments using the public FAERS datasets [9] to examine the effectiveness of rough-set-based ADR detection against traditional detection, from the viewpoint of timeline surveillance and warning of marketed drugs. The results show that most of the time the rough-set-based approach exhibits similar signaling capability to that of traditional approach. However, in some cases our approach, by providing an approximate range of signal strength, demonstrates better warning ability in timeline surveillance. This occurs especially when the amount of event cases with no missing value is relatively small, i.e., less than three, but the amount increases dramatically when missing values are included.

The remainders of this paper are organized as follows. In Section 2, we introduce background knowledge related to this work, including ADR detection and rough set theory. Section 3 presents our proposed rough set based method for measuring ADR signals from incomplete SRS data with missing values. In Section 4, we show and discuss the results of the experiments conducted over FAERS dataset. Finally, we describe conclusions and future work in Section 5.

2. Background and related work

2.1. ADR detection

Contemporary detection methods of ADR signals can be broadly divided into two categories [4]: frequentist methods and Bayesian methods.

Frequentist methods are widely used in most real ADR monitoring systems due to their simplicity to calculate and interpret. This category is mainly based on the statistical 2∗2 contingency table as shown in Table 1 to estimate the proportion of suspected ADRs in spontaneous reporting systems caused by the drug of interest vs. other drugs. If the ratio is higher than a threshold, then disproportionality occurs, which means the drug of interest is regarded to have a significant association with the suspected reaction. In the past decade, there have been various frequentist methods, each of which differs mainly on the metric for measuring the disproportionality. The most representative metrics are Proportional Reporting Ratio (PRR) [8] and Reporting Odds Ratio (ROR) [7]. Formulas of these two measures are defined as follows:

$$\mathrm{PRR} = \frac{a/(a+b)}{c/(c+d)}, \quad \mathrm{ROR} = \frac{a/c}{b/d}$$

Table 1
The 2 × 2 contingency table for ADR signal detection.

	Reaction of interest	Other reactions	Total
Drug of interest	a	b	$a+b$
Other drugs	c	d	$c+d$
Total	$a+c$	$b+d$	$N = a+b+c+d$

a: number of reports of the suspected drug lead to the suspected reaction.
b: number of reports of the suspected drug lead to all other reactions.
c: number of reports of all other drugs in the database lead to the suspected reactions.
d: number of reports of all other drugs lead to all other reactions.

Another category of more complex methods, Bayesian methods, were developed based on Bayesian statistics to estimate the (posterior) probability that the suspected adverse reaction occurs given the use of the suspected drug. Representatives of this category are Bayesian Confidence Propagation Neural Network (BCPNN) [2,3] and Multi-item Gamma Poisson Shrinker (MGPS) [1].

In the field of adverse drug reactions, most of detection methods can be used and every method has its own advantages and disadvantages. Therefore, one can select one or more suitable detection methods according to different analysis purposes.

2.2. Rough set theory

The rough set theory [16] is a useful tool for the analysis of imprecise, uncertainly or incomplete data. The theory is based on the concept of rough set, a formal approximation of a crisp set composed of objects represented by values of attributes. Classically, the set of objects concerned is represented as an information system or information table. In the following, we introduce the basic concepts of rough set theory and its extension to handle data with missing values.

(1) *Information system and decision table*: An information system is a pair $IS = \{U, A\}$, where U denotes a nonempty finite set of objects called the universe and A denotes a nonempty finite set of attributes. A decision table is a special form of information systems, in which the attribute set A is divided into a set of conditional attributes C and a decision attribute d, i.e. $A = C \cup \{d\}$. For example, in Table 2 there are three condition attributes $A = \{Height, Weight, Age\}$ and one decision attribute $d = \{Overweight\}$.

(2) *Indiscernibility relation*: Consider an information system $IS = \{U, A\}$. Let $B \subseteq A$ be a subset of attributes. The indiscernibility relation induced by B, denoted as I_B, is an equivalence relation defined as

$$(x, y) \in I_B \text{ if and only if for all } a \in B,\ \alpha(x, a) = \alpha(y, a),$$

where x and y are two cases in IS, and $\alpha(x, a)$ and $\alpha(y, a)$ denote the values of x and y, respectively, in attribute a. In other words, the indiscernibility relation induced by B defines a set of equivalence classes, within each of which the members have the same values in all attributes in B. For example, if B = {*Weight*, *Age*}, clearly $(2, 5) \in I_B$ since both cases have the same weight and age. And all cases in Table 2 will be divided into six equivalence classes, i.e., {1, 4}, {2, 5}, {3}, {6}, {7}, {8}.

(3) *Lower and upper approximations*: Let X represent a subset of elements of the universe U. The lower approximation indicates the set of elements certainly belonging to the set X, while the upper approximation indicates the set of elements possibly belonging to the set X. Given an information system $IS = (U, A)$ and $P \subseteq A$, the lower approximation of X induced by P in IS, denoted as $\underline{P}X$, and the upper approximation of X induced by P in IS, denoted as $\overline{P}X$, are defined as follows:

Table 2
An example of decision table.

Case	Height	Weight	Age	Overweight
1	170	75	18	Yes
2	165	50	30	Yes
3	165	60	18	No
4	145	75	18	No
5	145	50	30	No
6	170	45	45	Yes
7	145	50	45	No
8	170	45	30	Yes

$$\underline{P}X = \{e \in U | [e]_P \subseteq X\}, \quad \overline{P}X = \{e \in U | [e]_P \cap X \neq \emptyset\}$$

where $[e]_P$ denotes the equivalence class of e induced by attribute set P. For example, consider Table 2. Let $X = \{1, 2, 6, 8\}$ and $P = \{Weight, Age\}$. Below are the equivalence classes of each case.

$$[1]_P = \{1,4\}, \quad [2]_P = \{2,5\}, \quad [3]_P = \{3\}, \quad [4]_P = \{1,4\},$$
$$[5]_P = \{2,5\}, \quad [6]_P = \{6\}, \quad [7]_P = \{7\}, \quad [8]_P = \{8\}$$

And the lower and upper approximations of X induced by P are:

$$\underline{P}X = \{6,8\}, \quad \overline{P}X = \{1,2,4,5,6,8\}$$

(4) *Accuracy of approximations*: The accuracy of an approximation of X induced by P, denoted as $\sigma_P(X)$, is calculated as dividing the cardinality of the lower approximation by the cardinality of the upper approximation, i.e.,

$$\sigma_P(X) = \frac{|\underline{P}X|}{|\overline{P}X|}. \tag{1}$$

If $\sigma_P(X) = 1$, the lower and upper approximations are identical, and we say subset X is definable in U in terms of attribute set P. In other words, X can be regarded as not "imprecise" in terms of P, and so there is no room of vagueness needed to be captured by applying rough set theory. If $\sigma_P(X) < 1$, subset X can be defined by its lower and upper approximation and is roughly definable in U in terms of P.

2.3. Rough set strategies to data with missing data

In real world applications, a data collection usually contains missing values, making the data incomplete for analysis. Classically, the data is usually presented in the form of a decision table, where missing values can be interpreted from two aspects: lost and do not care. A "lost" missing value, denoted as "?", indicates that the value is important but is erased, and a "don't care" missing value, denoted as "$*$", indicates that the value is not important or redundant (see Table 3).

Various researchers have extended rough set theory for dealing with data with missing values [10,14,18]. We only present the concepts that are useful in our research, including the characteristic relation, characteristic set, and the refined lower and upper approximations.

The conventional rough set theory is under the assumption that information systems are complete, i.e., without missing data, and relies on the indiscernibility relation to derive the concept of lower and upper approximations. However, the indiscernibility relation is not applicable to data with missing values. Different extensions of the indiscernibility relation have been proposed, including the tolerance relation [14], similarity relation [18], and characteristic relation [10], which are described in what follows.

The tolerance relation was proposed by Kryszkiewicz to process data with "don't care" missing values, the similarity relation was proposed by Stefanowski and Tsoukias to process data with "lost" missing values, while the characteristic relation, proposed by Grzymala-Busse, considers both "lost" and "don't care" missing values. Since the characteristic relation is a general form of the tolerance and similarity relations, in this paper we adopt this term (denoted as R), and use subscripts T and S to denote the tolerance (R_T) and similarity versions (R_S), respectively.

Table 3
An example of an incomplete decision table containing "lost" (?) or "don't care" ($*$) missing values.

Case	Height	Weight	Gender	Overweight
1	170	50	Male	Yes
2	165	?/$*$	Female	No
3	170	80	?/$*$	No
4	165	50	Female	No
5	?/$*$	?/$*$	Male	Yes

Definition 1. Let $P \subseteq A$ be a subset of attributes. The similarity characteristic relation, denoted by $R_S(P)$, is defined as:

$(x,y) \in R_S(P)$ if and only if $\alpha(x,a) = \alpha(y,a)$ for all $a \in P$ such that $\alpha(x,a) \neq ?$.

And the similarity characteristic set is defined as $K_S(P,x) = \{y | (x,y) \in R_S(P)\}$, where x and y are two cases in the decision table, and $\alpha(x,a)$ denotes the value of x in attribute a.

Definition 2. Let $P \subseteq A$ be a subset of attributes. The tolerance characteristic relation, denoted by $R_T(P)$, is defined as:

$(x,y) \in R_T(P)$ if and only if $\alpha(x,a) = \alpha(y,a)$ or $\alpha(x,a) = {}^*$ or $\alpha(y,a) = {}^*$ for all $a \in P$.

And the tolerance characteristic set is $K_T(P,x) = \{y | (x,y) \in R_T(P)\}$.

Example 1. Consider Table 3. Let $P = \{Height, Weight, Gender\}$. Then the similarity and tolerance characteristic sets of all cases induced by P are:

$$\begin{array}{ll} K_S(P,1) = \{1\} & K_T(P,1) = \{1,5\} \\ K_S(P,2) = \{2,4\} & K_T(P,2) = \{2,4\} \\ K_S(P,3) = \{3\} & K_T(P,3) = \{3,5\} \\ K_S(P,4) = \{4\} & K_T(P,4) = \{2,4\} \\ K_S(P,5) = \{1,5\} & K_T(P,5) = \{1,3,5\} \end{array}$$

Based on the concept of characteristic relation and characteristic set, Grzymala-Busse [10] proposed three different extensions of the lower and upper approximations for processing data with missing values: *singleton*, *subset*, and *concept* approximations.

The first extension is called singleton approximation, which considers all cases in U and is similar to the original definitions of lower and upper approximations. Hereafter, for identification purpose we add subscripts g (singleton), s (subset), and c (concept) into the approximation, and add superscript K (stand for characteristic relation) to distinguish it from the conventional approximation derived by indiscernibility relation.

Definition 3. The singleton lower approximation of X induced by P, denoted by $\underline{P}_g^K X$, is the set of all cases whose characteristic set is contained in X, i.e.,

$$\underline{P}_g^K X = \{x \in U | K(P,x) \subseteq X\}$$

The singleton upper approximation of X in P, denoted by $\overline{P}_g^K X$, is the set of cases whose characteristic set having an non-empty intersection with X, i.e.,

$$\overline{P}_g^K X = \{x \in U | K(P,x) \cap X \neq \emptyset\}$$

Note that the characteristic set $K(P,x)$ presented in the above definition can be any types of characteristic sets. The second extension, called subset approximation, uses the union of characteristic sets to define approximation.

Definition 4. The subset lower approximation of X induced by P, $\underline{P}_s^K X$, is the union of characteristic sets that are contained in X, i.e.,

$$\underline{P}_s^K X = \cup\{K(P,x) | x \in U, K(P,x) \subseteq X\}$$

The subset upper approximation of X induced by $P, \overline{P}_s^K X$, is the union of characteristic sets which have an nonempty intersection with X, i.e.,

$$\overline{P}_s^K X = \cup\{K(P,x) | x \in U, K(P,x) \cap X \neq \emptyset\}$$

The third definition called concept approximation is more stringent than the subset version in that it only considers those cases in X.

Definition 5. The concept lower and upper approximations of X induced by P are defined as follows:

$$\underline{P}_c^K X = \cup\{K(P,x) | x \in X, K(P,x) \subseteq X\}$$

$$\overline{P}_c^K X = \cup\{K(P,x) | x \in X, K(P,x) \cap X \neq \emptyset\}$$

Example 2. Let X be the set of cases with *Overweight* = "Yes" in Table 3, i.e., X = {1, 5} and P = {*Height*, *Weight*, *Gender*}. The corresponding singleton, subset, and concept approximations of X are:

$$\underline{P}_g^K\{1,5\} = \{1\} \quad \underline{P}_s^K\{1,5\} = \{1,5\} \quad \underline{P}_c^K\{1,5\} = \{1\}$$

$$\overline{P}_g^K\{1,5\} = \{1,3,5\} \quad \overline{P}_s^K\{1,5\} = \{1,3,5\} \quad \overline{P}_c^K\{1,5\} = \{1,3,5\}$$

Note that for complete decision tables, all of the three approximations, singleton, subset, and concept, are amalgamated into the same definition. However, it is not true for incomplete decision tables.

3. Rough set based ADR detection

3.1. Problem description

As mentioned in Section 1, the SRS data may contain some missing values due to omitting or personal privacy problem. To facilitate the discussion, the reporting data is presented as an information system $IS = (U, A)$ containing missing values which can be either one of two categories: lost (?) or don't care ($*$). Our purpose is to examine the feasibility of rough set theory to the ADR detection, focusing on whether the inclusion of missing data through rough set based approximation can be helpful for the predicting capability of generated signals. Therefore, the problem can be described as given a SRS dataset that contains missing values and is represented in the form of a data table, we like to compute the strength (using PRR or ROR measure) of any given suspected ADR rule of the following form:

$$\text{Predc}, drug \rightarrow symptom \tag{2}$$

where Predc denotes extra conditions associated with the signal, e.g., *Sex* = "female", *Age* = ">18", and we will examine if the strength of this rule is over a specified threshold to becoming a noteworthy ADR signal.

Table 4
Description of the attributes selected from the FAERS dataset.

File name	Selected attribute name	Containing null values	Null probability (07Q2)
DEMO	ISR	×	0
	EVENT_DT	√	31.3
	AGE	√	38.8
	GNDR_COD	√	5.8
DRUG	DRUGNAME	×	0
REAC	PT	×	0

DEMO: to record personal information for each patient.
DRUG: to record the medicines taken by each patient.
REAC: to record the observed adverse reactions for each report.

In this study, the SRS data was obtained from the FDA Adverse Event Reporting System (FAERS) database [9]. The FAERS database is composed of seven data files, including DEMO, DRUG, REAC, OUTC, RPSR, THER, and INDI. We selected three data files that are essential for ADR signal detection, i.e., DEMO, DRUG, and REAC. From the DEMO data file we chosen four attributes about personal information of patients, including ISR (primary report id), EVENT_DT, AGE, and GNDR_COD. These attributes except ISR may contain null values. From the DRUG and REAC files we chosen the DRUGNAME and PT attributes, which do not contain null values. Details of the chosen attributes are presented in Table 4.

3.2. Rough set based measuring

Since all contemporary measures relies on the contingency 2×2 table, our basic idea is applying rough set theory to the calculation of the contingency 2×2 table. Consider the rule in (2) and the following corresponding contingency table.

Predc	*symptom*	other *symptoms*
drug	*a*	*b*
other *drugs*	*c*	*d*

If the information system is complete, then each of the cell values, a, b, c, d, on the contingency table are deterministic. Unfortunately, as we have shown previously, the attributes involved in the Predicate may contain missing values, causing the cell values imprecise. We thus adopt the concept of lower and upper approximations to obtain an approximate range for each cell values and in accordance compute the strength of the corresponding rule.

For simplicity, let X_a, X_b, X_c, and X_d denote the sets of cases satisfying the corresponding cell conditions in the contingency table. Clearly, for complete data we have $a = |X_a|$, $b = |X_b|$, $c = |X_c|$, and $d = |X_d|$. But for incomplete data we need to compute the lower and upper approximations for X_a, X_b, X_c, and X_d. Let P denote the set of attributes for the approximation computation. Each cell value can be denoted by a range, i.e.,

$$a^* : [\underline{a}, \bar{a}], b^* : [\underline{b}, \bar{b}], c^* : [\underline{c}, \bar{c}], d^* : [\underline{d}, \bar{d}] \tag{3}$$

Accordingly, we have

$$\begin{aligned} \underline{a} = |\underline{P}^K X_a|, \underline{b} = |\underline{P}^K X_b|, \underline{c} = |\underline{P}^K X_c|, \underline{d} = |\underline{P}^K X_d| \\ \bar{a} = |\overline{P}^K X_a|, \bar{b} = |\overline{P}^K X_b|, \bar{c} = |\overline{P}^K X_c|, \bar{d} = |\overline{P}^K X_d| \end{aligned} \tag{4}$$

Then the strength (range value) of the rule can be computed by performing a simple range calculation according to the formula of PRR and ROR. The resulting formulas are as follows:

$$\frac{\underline{a}(\underline{c} + \underline{d})}{\bar{c}(\bar{a} + \bar{b})} \leqslant \text{PRR} \leqslant \frac{\bar{a}(\bar{c} + \bar{d})}{\underline{c}(\underline{a} + \underline{b})}, \quad \frac{\underline{a} \times \underline{d}}{\bar{c} \times \bar{b}} \leqslant \text{ROR} \leqslant \frac{\bar{a} \times \bar{d}}{\underline{c} \times \underline{b}} \tag{5}$$

We consider two different options for defining the set P: global covering and local covering. The global covering specifies all attributes in the data to P, i.e., $P = A$. The local covering instead only considers the set of attributes forming the rule of concern (and so the contingency table). For convenience, we denote this attribute set as B, for $B \subseteq A$.

Since there are two different interpretations of missing values, i.e., lost or don't care, and three different versions of approximations, i.e., singleton, subset, and concept approximations, in total, we obtain twelve different ways for computing the cell values defined in (3) and (4), as shown in Fig. 1. Fig. 1 also depicts the research framework adopted in this study, inspecting the

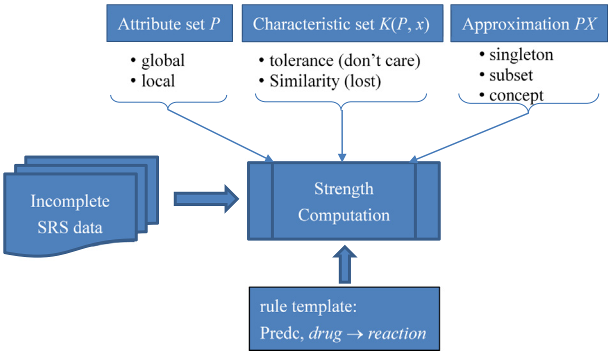


Fig. 1. The research framework for incomplete ADR signal detection.

feasibility for applying rough set theory to the ADR signals detection from an incomplete SRS dataset containing missing values. We assume that the template of the rule to be discovered is given, either by the user or generated by a pre-procedure of candidate rule generation. In the remainder of this section, we will examine the feasibility of the twelve ways (versions) for computing the cell values.

3.3. Feasibility analysis

We analyze the feasibility of the twelve different methods by examining whether each one of them can yield reasonable approximations for data with missing values. To facilitate the discussion, we first introduce the concept of satisfiable approximation and indistinguishable approximation.

Consider a rule of the form defined in (2) and the corresponding contingency table. Let C be the attribute set for defining the extra conditions, i.e., Predc, for forming the contingency table. For example, if Predc = {*Sex* = "female", *Age* = ">18"}, then C = {*Sex*, *Age*}.

Definition 6. Let y be any case in U. We say y satisfies the Predc condition if for each attribute t in C, $\alpha(y,t) = \alpha(\text{Predc}, t)$ or $\alpha(y,t) = ?$ or $\alpha(y,t) = *$, where $\alpha(\text{Predc}, t)$ denotes the condition value of attribute t in Predc.

Definition 7. An approximation of the contingency set X (X can be X_a, X_b, X_c, or X_d) defined on an attribute set P is a C-satisfiable approximation if all members in either the lower approximation $\underline{P}X$ or upper approximation $\overline{P}X$ satisfy the Predc condition specified by C.

Example 3. Consider the data with lost missing values in Table 3. We would like to compute the strength of the following rule:

$$Gender = \text{g1}, Drug = \text{d2} \rightarrow PT = \text{s1}$$

The corresponding contingency sets are X_a= {4}, $X_b = \emptyset$, X_c= {3, 7, 8}, and $X_d = \emptyset$, and C = {*Gender*}. Now assume the subset approximation with similarity characteristic set and global covering is applied. Then, we obtain the following characteristic sets of all cases in Table 5.

$$\begin{array}{ll} K_S(P,1) = \{1,3,6,7\} & K_S(P,5) = \{2,4,5\} \\ K_S(P,2) = \{2\} & K_S(P,6) = \{6\} \\ K_S(P,3) = \{3\} & K_S(P,7) = \{3,7\} \\ K_S(P,4) = \{4\} & K_S(P,8) = \{8\} \end{array}$$

Below are the lower and upper approximations of X_a, X_b, X_c, and X_d.

$$\underline{P}_s^K X_a = \{4\}, \quad \underline{P}_s^K X_b = \emptyset, \quad \underline{P}_s^K X_c = \{3,7,8\}, \quad \underline{P}_s^K X_d = \emptyset$$
$$\overline{P}_s^K X_a = \{4\}, \quad \overline{P}_s^K X_b = \emptyset, \quad \overline{P}_s^K X_c = \{1,3,6,7,8\}, \quad \overline{P}_s^K X_d = \emptyset$$

Table 5
An incompletely data table with lost missing values.

ISR	Age	Gender	Drug	PT
1	?	?	d1	s1
2	a2	?	d2, d3	s1, s2
3	a1	g1	d1	s1
4	a1	g1	d2, d3	s1, s2
5	?	?	d2, d3	s1, s2
6	?	g2	d1	s1
7	?	g1	d1	s1
8	a1	g1	d3	s1, s2

Note that case 6 in the upper approximation of X_c contradicts condition *Gender* = g1. Therefore, the subset approximation with similarity characteristic set and global covering is not C-satisfiable.

Definition 8. An approximation of the contingency set X (X can be X_a, X_b, X_c, or X_d) defined on an attribute set P is *indistinguishable* if the lower approximation of the contingency set X is always equal to the corresponding upper approximation, i.e., $\underline{P}^K X = \overline{P}^K X$.

Example 4. Consider Table 5 and the rule in Example 3 again. Assume that the concept approximation with similarity characteristic set and global covering is applied. Below are the lower and upper approximations of X_a, X_b, X_c, and X_d.

$$\underline{P}_c^K X_a = \{4\}, \quad \underline{P}_c^K X_b = \emptyset, \quad \underline{P}_c^K X_c = \{3,7,8\}, \quad \underline{P}_c^K X_d = \emptyset$$
$$\overline{P}_c^K X_a = \{4\}, \quad \overline{P}_c^K X_b = \emptyset, \quad \overline{P}_c^K X_c = \{3,7,8\}, \quad \overline{P}_c^K X_d = \emptyset$$

Since the lower and upper approximations are the same, this approximation is indistinguishable.

In what follows we present the important properties required to determine the feasibility of the twelve measurements. All detailed proofs are presented in Appendix to keep the content more concise and readable.

Lemma 1. *The subset approximation defined by tolerance characteristic set K_T for contingency sets X_a, X_b, X_c, and X_d, is not C-satisfiable with respect to P, for $P \supseteq B$.*

Lemma 2. *The subset approximation defined by similarity characteristic set K_S for contingency sets X_a, X_b, X_c, and X_d, is not C-satisfiable with respect to P, for $P \supseteq B$.*

Lemma 3. *The concept approximation defined by similarity characteristic set K_S is indistinguishable for contingency sets X_a, X_b, X_c, and X_d, with respect to P, for $P \supseteq B$, i.e., $\underline{P}_c^{K_S} X = \overline{P}_c^{K_S} X$, for X being X_a, X_b, X_c, or X_d.*

Lemma 4. *The concept approximation defined by tolerance characteristic set K_T is not indistinguishable for contingency sets X_a, X_b, X_c, and X_d, with respect to P, for $P \supseteq B$, i.e., $\underline{P}_c^{K_T} X \neq \overline{P}_c^{K_T} X$, for X being X_a, X_b, X_c, or X_d.*

Finally, we show that the singleton approximation defined either by similarity or tolerance characteristic set is C-satisfiable and not indistinguishable.

Lemma 5. *The singleton approximation defined by similarity characteristic set K_S for contingency sets X_a, X_b, X_c, and X_d, is C-satisfiable with respect to P, for $P \supseteq B$.*

Lemma 6. *The singleton approximation defined by tolerance characteristic set K_T is not indistinguishable for contingency sets X_a, X_b, X_c, and X_d, with respect to P, for $P \supseteq B$, i.e., $\underline{P}_g^{K_T} X \neq \overline{P}_g^{K_T} X$, for X being X_a, X_b, X_c, or X_d.*

Table 6
Summarization of the feasible and infeasible approximation methods.

	Lost		Don't care	
	Global	Local	Global	Local
Singleton	√	√	√	√
Subset	✘	✘	✘	✘
Concept	x	x	√	√

√: feasible approximation.
✘: infeasible approximation, due to unsatisfiable property.
x: infeasible approximation, due to indistinguishable property.

In summary, the twelve approximation methods can be divided into two categories, the feasible methods and the infeasible methods, as shown in Table 6. For convenience, we denote the six feasible methods in terms of characteristic sets (similarity or tolerance), attribute covering (global or local), and approximation definition (singleton, subset, or concept) as follows:

Method 1 M(s, g, g): Similarity set, global covering, singleton approximation.
Method 2 M(s, l, g): Similarity set, local covering, singleton approximation.
Method 3 M(t, g, g): Tolerance set, global covering, singleton approximation.
Method 4 M(t, l, g): Tolerance set, local covering, singleton approximation.
Method 5 M(t, g, c): Tolerance set, global covering, concept approximation.
Method 6 M(t, l, c): Tolerance set, local covering, concept approximation.

3.4. Comparative analysis

We further conducted a comparative analysis of the derived six rough-set-based measurements in terms of the accuracy derived from (1). To this purpose, we first define the *Accuracy* $(\cdot)$ of a rough set based measuring method as the ratio of lower and upper signal values. That is,

$$\frac{\mathrm{PRR}_l}{\mathrm{PRR}_u}=\frac{\underline{a}(\underline{c}+\underline{d})\underline{c}(\underline{a}+\underline{b})}{\bar{a}(\bar{c}+\bar{d})\bar{c}(\bar{a}+\bar{b})},\quad \frac{\mathrm{ROR}_l}{\mathrm{ROR}_u}=\frac{\underline{a}\times\underline{d}\times\underline{c}\times\underline{b}}{\bar{a}\times\bar{d}\times\bar{c}\times\bar{b}} \tag{6}$$

We showed that there exists a proper order of the six rough-set-based methods in terms of the measurement accuracy, summarized in Theorem 7, where M1–M6 refer to the six methods. Details of the proof are described in Appendix.

Table 7
The resulting CS_list, corresponding to tolerance characteristic set, in terms of attribute set {Height, Gender} for the example in Table 3.

Value Pair	ID list
165, Male	5
165, Female	2, 4
165, *	2, 4
170, Male	1, 3, 5
170, Female	3
170, *	1, 3, 5
*, Male	1, 3, 5
*, Female	2, 3, 4
*, *	1, 2, 3, 4, 5

Theorem 7. *Accuracy(M1) ⩾ Accuracy(M2) ⩾ Accuracy(M5) ⩾ Accuracy(M3) ⩾ Accuracy(M4) = Accuracy(M6).*

3.5. The detection method

Given a SRS dataset with missing values, we assume that the rule representing the ADR signal to be discovered is provided by the user. Our algorithm, as shown in Fig. 2, computes the strength of the rule according to the following parameters, attribute covering (global or local), characteristic set (tolerance or similarity), approximation (singleton, subset, or concept), and the signal measure (PRR or ROR).

The most expensive step of our algorithm is Step 1. Computing the characteristic set of a given case requires pairwise case comparison throughout the whole SRS table. For a SRS table consisting of n cases, this procedure consumes $O(n^2)$ case comparisons, and so we developed a more efficient method. The basic idea is as follows.

We introduced a structure called *CS_list* (Characteristic Set list) to store the case*_IDs* that belong to the same characteristic set, i.e., indistinguishable in terms of the attributes of concern. For example, consider Table 3. The resulting *CS_list*, corresponding to tolerance characteristic set, in terms of attribute set {*Height*, *Gender*} is depicted in Table 7, where column Value Pair denotes the possible combinations in terms of {*Height*, *Gender*} and ID List the set of cases belong to the characteristic set of this value pair. Once the *CS_list* has been pre-computed, the characteristic set of a given case can be generated in a minute. As such, we performed a preprocessing to generate six different *CS_lists* corresponding to the six rough-set-based measurements. Note this process is performed only once and can be computed offline. With the *CS_list* available, we can achieve a near on-line signal measuring for each rule of interest, as will be demonstrated in the experiments.

Input:

- *STab*: the SRS data table;
- *RTemp*: the rule template;
- *ACtype*: the type of attribute coverings;
- *CStype*: the type of characteristic sets;
- *APtype*: the type of approximations;
- *MStype*: the type of measures.

Output: The rule and the strength.

Steps:

1. Compute the characteristic sets of all records in *STab* according to *ACtype* and *CStype*;
2. Generate the four contingency sets, X_a, X_b, X_c, and X_d, according to the rule template *RTemp*;
3. Generate the lower and upper approximations of X_a, X_b, X_c, and X_d, according to the selected approximation *APtype*;
4. Compute the rule strength using the approximate contingency a^*, b^*, c^*, and d^*, according to the measure *MStype*;
5. Return the rule with the computed strength;

Fig. 2. Algorithmic framework of the proposed ADR detection method.

Table 8
Selected drugs marketed in US and associated ADRs.

Rule no.	Drug name	Adverse reaction	Group (age or gender)	Marked year	Withdrawn or warning year
ADRs of withdrawn drugs					
R1-1	AVANDIA	Myocardial infarction	18∼	1999	2010
R1-2		Death			
R1-3		Cerebrovascular accident			
R2	TYSABRI	Progressive multifocal leukoencephalopathy	18∼	2004	2005
R3	ZELNORM	Cerebrovascular accident	Female	2002	2007
ADRs of non-withdrawn drugs					
R4	WARFARIN	Myocardial infarction	60∼	1940	2014
R5	REVATIO	Death	∼18	2008	2014

Table 9
The accuracy of each rough set based method (For X_a).

Rule	Method					
	M1	M2	M3	M4	M5	M6
R1-1	0.926	0.894	0.854	0.818	0.855	0.818
R1-2	0.909	0.881	0.851	0.822	0.852	0.822
R1-3	0.959	0.947	0.903	0.885	0.903	0.885
R2	0.945	0.940	0.900	0.885	0.900	0.885
R3	1.000	0.994	0.996	0.994	0.996	0.994
R4	1.000	1.000	1.000	1.000	1.000	1.000
R5	0.922	0.918	0.881	0.877	0.881	0.877
Average	0.952	0.939	0.912	0.897	0.912	0.897

4. Experiments

We conducted a series of experiments to inspect the effectiveness of our methods. All of the available collections of the FAERS dataset, from 2004Q1 to 2013Q3, were used. Each quarterly collection contains around 60,000 to 230,000 reports. All experiments were performed on a PC with CPU i7-2600, 3 GB RAM, and 120 GB SSD.

We aim to compare the predicting capability of our rough-set-based methods with deletion method, the most common technique for handling missing data, on timeline warning of serious ADR signals. Two variants of deletion method were considered: listwise deletion and pairwise deletion. The listwise deletion method eliminates any records containing missing value in at least one attribute, while pairwise deletion withholds the records with missing values not occurring on the attributes of concern.

Two groups of drugs were used in these experiments, including three withdrawn drugs [5,9], AVANDIA, TYSABRI, and ZELNORM, and two non-withdrawn drugs but labeled in the FDA warning list (MedWatch) [15], WARFARIN and REVATIO. Other criteria for choosing these drugs are: (1) There are enough cases associated with these drugs reported in the FAERS dataset (yearly number of reports > 3); (2) These drugs yield known ADRs associated with specific populations. Table 8 lists detailed information of these drugs and the associated ADRs. For convenience, each ADR is denoted by a rule. The AGE attribute was discretized into three levels, "<18", "18–60", and ">60", in accordance with these rules.

All signals were measured by two commonly used criteria, PRR and ROR, though we only show the results measured by PRR since similar phenomena were observed for ROR. The threshold for an ADR rule being significant followed the widely adopted setting, PRR $\geqslant 2$ and $a \geqslant 3$ [6], where a denotes the number of reports satisfy the rule.

4.1. Accuracy comparison

We first compared the accuracy of the six rough-set-based methods. For this purpose, we computed for each method the average accuracy of each rule's strength over all quarters, and then obtained the final average over all rules. Since the accuracies of all contingency sets and rule signals exhibit similar phenomenon, we only show the results for contingency set X_a (see Table 9). Obviously, the results are consistent with the analysis presented in Section 3.4; Method 1 outperforms all the others, while Methods 4 and 6 exhibit the worst performance.

4.2. Prediction comparison

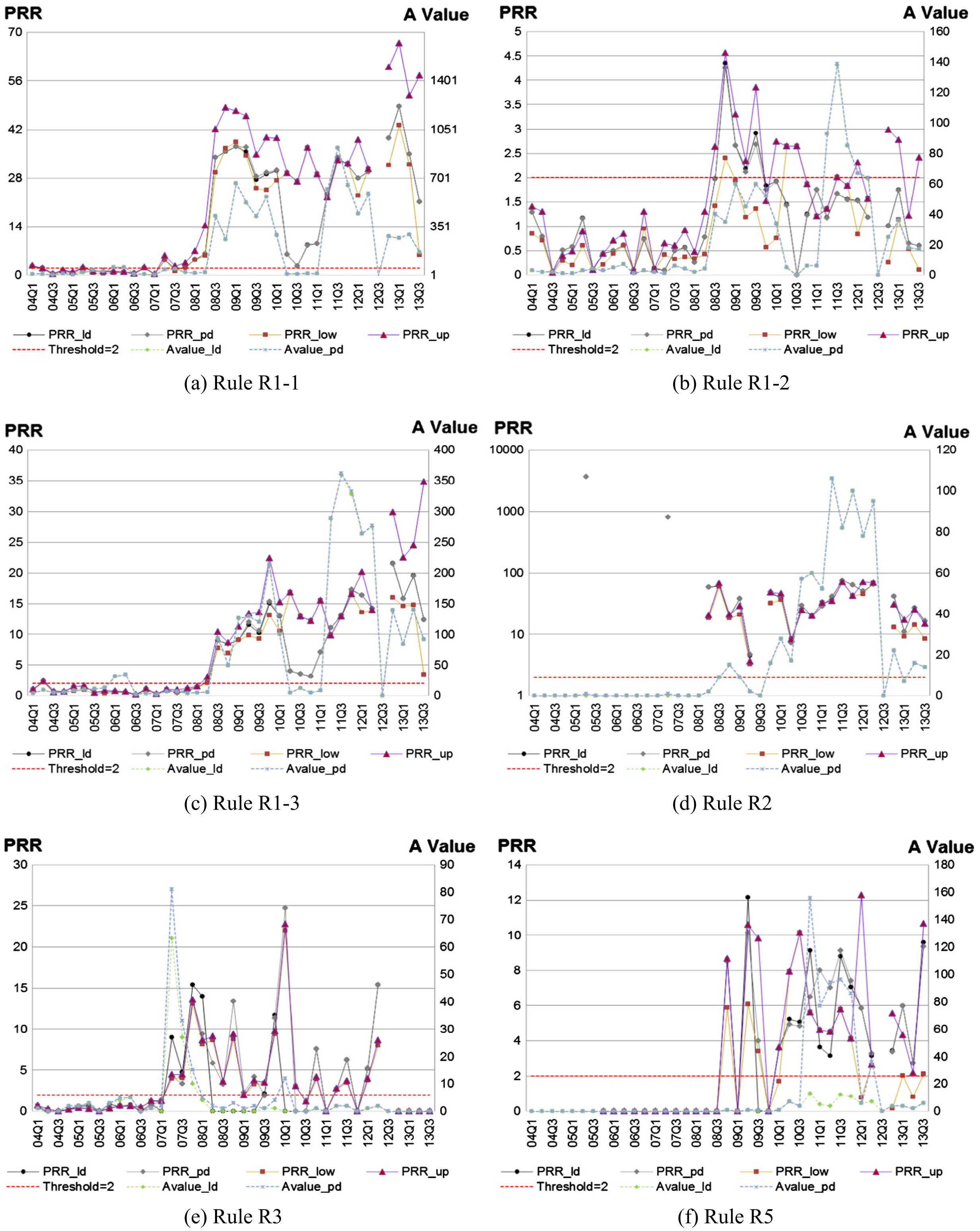
Since Method 1 exhibits the best accuracy, we then compared Method 1 with listwise deletion and pairwise deletion. The results are displayed in Fig. 3, where PRR_ld, PRR_pd, PRR_low, and PRR_up denote the PRRs generated by listwise deletion, pairwise deletion, lower and upper approximation by our method, respectively. Note rule R4 is omitted because all methods for this rule failed to generate significant strength (with PRR $\geqslant 2$). The contingency a values are shown for convenience to inspect the condition $a \geqslant 3$.

As the results demonstrate, most of the time our method exhibit similar capability of timeline warning as that of deletion method, both predicting ADR signals earlier than the time FDA issued warning or withdrawal announcement. However, in some cases our method, by providing an approximate range of signal strength, can predict the signal earlier than both the listwise and pairwise deletions. For example, for R1-1 our method generates stable strengths higher than threshold starting from 2007Q2 while the listwise and pairwise methods do so from 2007Q4, and for R1-2 the result is 2008Q3 (our method) vs. 2008Q4 (listwise and pairwise deletions).

We also observe that listwise deletion and pairwise deletion yields nearly the same signal strength over all cases, though pairwise sometimes generate higher strength of rule due to larger a values it maintains, making pairwise a little better than listwise in signal prediction. For example, see the results for rule R3 during 07Q3 to 10Q1, and 10Q4 to 12Q2 for R5. This enhancement on the other hand would bias the signal of concern. For example, for rule R2 we observe two extraordinary large values yielded by pairwise deletion at 05Q2 and 07Q2, for which a report containing missing value that is dismissed by listwise deletion and our method is taken into account.

Another important phenomenon is solely using quarterly generated patterns for monitoring ADR signals is not reliable. For example, the noteworthy signal for R1-2 generated by our method during 08Q3 to 10Q3 is faded out and appear again during 12Q4 to 13Q1. This is mainly contributed by the well-known underreporting problem [19], causing large variances among the number of cases reported over different quarters. For this reason, cumulative ADR measurements usually are adopted as auxiliaries.

Fig. 4 shows the cumulative counterpart of Fig. 3. In general, the results coincide with those exhibited in Fig. 3. The curves generated by the deletion method, either listwise or pairwise, are situated between those by our method, which also reinforces the superiority of our method in earlier detection of high-profile



(a) Rule R1-1

(b) Rule R1-2

(c) Rule R1-3

(d) Rule R2

(e) Rule R3

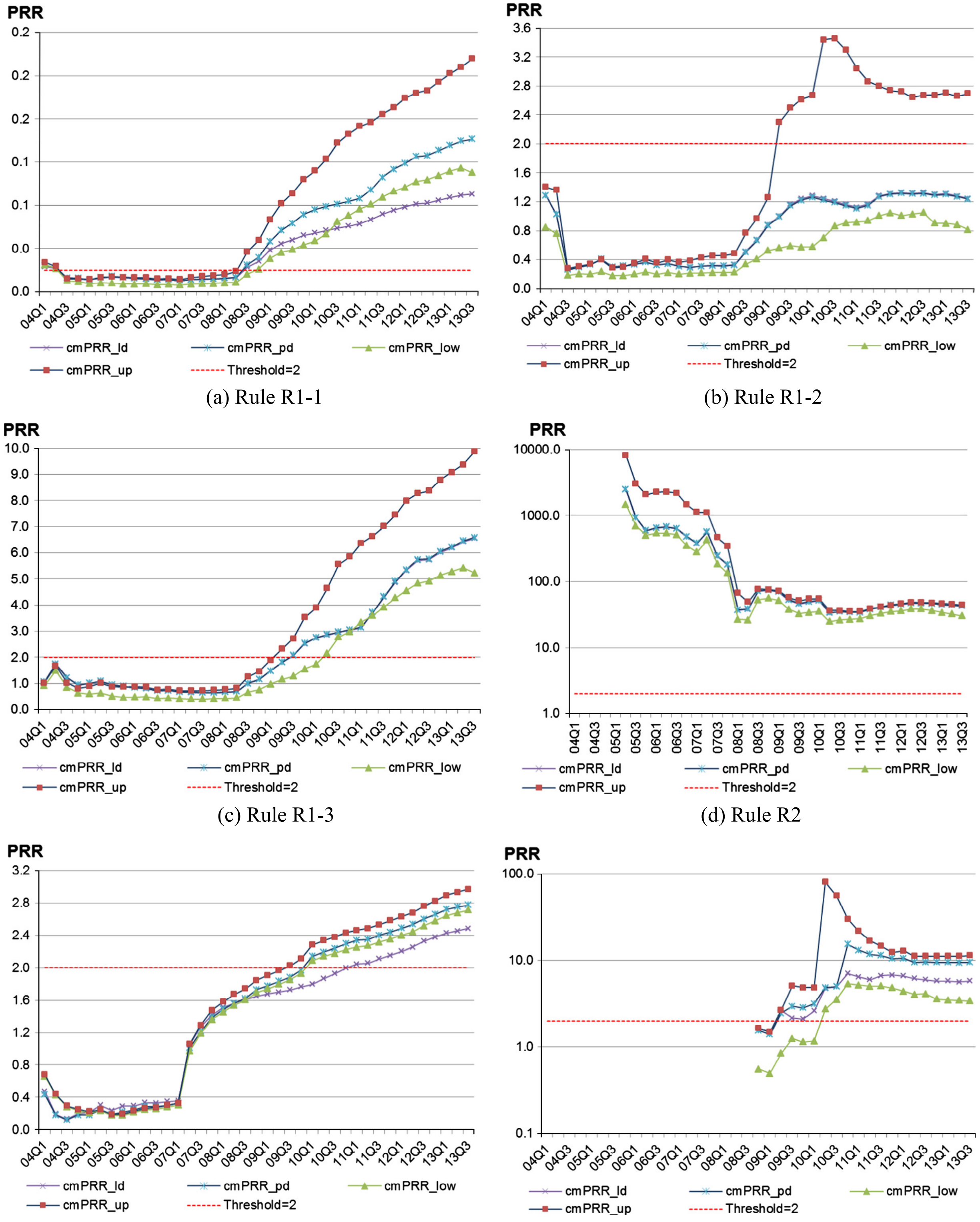
(f) Rule R5

Fig. 3. Comparison of our method and traditional method on quarterly generated signal strengths.

signals. A noteworthy observation occurs to Fig. 4(b), where both deletion methods fail to identify Rule 1–2 as significant, while our method exhibits the potential for this ADR signal.

4.3. Performance comparison

Finally, we compared the execution times of our method with listwise and pairwise deletions. Two different implementations of our method were considered, including the naïve approach and *CS_list* pre-computed approach. Since the results are similar for all ADR rules measured by the six methods, we only show the results of running rule R1-1 with method M1 over 2004Q2 and 2012Q1, where 2004Q2 represents the smallest dataset containing around 60,000 records, while 2012Q1 is the largest dataset containing around 230,000 records. The preprocessing times for generating *CS_list* from 2004Q2 and 2012Q1 requires two and eight



(a) Rule R1-1

(b) Rule R1-2

(c) Rule R1-3

(d) Rule R2

(e) Rule R3

(f) Rule R5

Fig. 4. Comparison of our method and traditional method on cumulative quarterly signal strengths.

minutes, respectively. From the results shown in Table 10, our *CS_list* version, even including the *CS_list* generation time, significantly outperforms naïve implementation, achieving 150x speedup. Both implementations of our method are relatively expensive compared with listwise and pairwise deletions. However, the computation overhead caused by using our method,

Table 10
Execution time compariosn for rule R1-1 with M1.

Dataset	Listwise (s)	Pairwise (s)	RS method	
			Naïve	*CS_list* (s)
2004Q2	0.4	0.5	40 min	19
2012Q1	1.7	1.8	20 h	104

especially the *CS_list* version, shall be acceptable compared with quarterly duration for data collection and signal reporting.

5. Conclusions

Although it is well known that the SRS dataset contains lots of missing data, most of published research work on this topic adopted listwise deletion to eliminate data with missing values. No work has noticed the possibility and examined the effect of including the missing data in the process of ADR detection. In this paper, we have inspected the feasibility of applying rough set theory to the ADR detection problem. Specifically, we have proposed twelve different rough-set-based measuring methods and showed that, in terms of two novel concepts, satisfiable and indistinguishable properties, only six of them are feasible for the purpose. We also have conducted a comparative analysis of these six methods in terms of measurement accuracy, showing Method 1 the most effective.

We have compared our method with traditional frequentist methods with listwise deletion or pairwise deletion in timeline warning of noteworthy ADR signals. Experimental results show that most of the time our method exhibits similar capability of timeline warning as that of traditional method but in some cases it can yield noteworthy measures earlier. From the preliminary results, we conclude that rough-set-based ADR signal measuring method that takes missing data into account is feasible and may be regarded as an auxiliary for the traditional measuring method.

In the future, we will conduct more comprehensive experiments on other drugs and improve the efficiency of our algorithm. Besides, our method only applicable to ADR rules with extra condition, that is, there is at least one incomplete attribute other than Drug and PT involved in the rule condition. We will also devise other rough-set-based approaches to eliminate this limitation.

Another restriction of our rough-set-based approach goes into continuous attributes. The rough set theory is based on the concept of indiscernibility relation, from which lower and upper approximations are derived. Although theoretically this relation applies both for attributes with discrete and continuous domains, in practice it is only valuable for discrete attributes, because in a decision table all objects may have been discernible with respect to continuous attributes. This is why discretization is needed for continuous attributes before employs rough set theory. Since the characteristic relation for dealing with missing values is derived from the indiscernibility relation, discretization for continuous attributes is also necessary. It will be a challenging issue for how to deal with continuous attributes without discretization.

Conflict of interest

None declared.

Acknowledgment

This work was supported by the National Science Council of Taiwan under Grant No. NSC101-2221-E-390-027.

Appendix A

In this appendix, we present the detailed proof of all theorems. First, we prove Lemmas 1 to 6.

Lemma 1. *The subset approximation defined by tolerance characteristic set K_T for contingency sets X_a, X_b, X_c, and X_d, is not C-satisfiable with respect to P, for $P \supseteq B$.*

Proof. We only consider the case of X_a and $P = B$. It is easy to apply similar strategies to prove other cases. To prove $\underline{P}_s^{K_T}X_a$ is not C-satisfiable, we will show that indeed, the upper approximation $\overline{P}_s^{K_T}X_a$ is not C-satisfiable. □

Lemma 2. *The subset approximation defined by similarity characteristic set K_S for contingency sets X_a, X_b, X_c, and X_d, is not C-satisfiable with respect to P, for $P \supseteq B$.*

Proof. The proof is similar to that in Lemma 1. □

Lemma 3. *The concept approximation defined by similarity characteristic set K_S is indistinguishable for contingency sets X_a, X_b, X_c, and X_d, with respect to P, for $P \supseteq B$, i.e., $\underline{P}_c^{K_S}X = \overline{P}_c^{K_S}X$, for X being X_a, X_b, X_c, or X_d.*

Proof. Again, we only consider the case of X_a and $P = B$. Recall the following definitions for $\underline{P}_c^{K_S}X_a$ and $\overline{P}_c^{K_S}X_a$.

$$\underline{P}_c^{K_S}X_a = \cup\{K_S(P,x)|x \in X_a, K_S(P,x) \subseteq X_a\}$$

$$\overline{P}_c^{K_S}X_a = \cup\{K_S(P,x)|x \in X_a, K_S(P,x) \cap X_a \neq \emptyset\}$$

According to the definition of $K_S(P,x)$, if a case $y \in K_S(P,x)$, then $\alpha(x,t) = \alpha(y,t)$ for any attribute $t \in P$ and $\alpha(x,t) \neq ?$. Since $x \in X_a$, it follows that all attribute values of x in B are not lost, i.e., for all $t \in B$, $\alpha(x,t) \neq ?$, and so are y. This means if $y \in K_S(P,x)$ then $y \in X_a$ as well. In other words, $K_S(P,x) \subseteq X_a$ and we have

$$\underline{P}_c^{K_S}X_a = \overline{P}_c^{K_S}X_a = X_a,$$

which proves the lemma. □

It is interesting to note that in the proof of Lemma 3, if the concept approximation is defined by the tolerance characteristic set, then a member $y \in K_T(P,x)$ may not belong to X_a. This is because y may contain some don't care attributes in B, which hinders it from a member of X_a. This leads to the proof of Lemma 4.

Lemma 4. *The singleton approximation defined by similarity characteristic set K_S for contingency sets X_a, X_b, X_c, and X_d, is C-satisfiable with respect to P, for $P \supseteq B$.*

Proof. Once again, we only consider the case of X_a and $P = B$. The other cases can be proved by similar strategies. According to the definitions, we have

$$\underline{P}_g^{K_S}X_a = \{x \in U|K_S(P,x) \subseteq X_a\}$$

$$\overline{P}_g^{K_S}X_a = \{x \in U|K_S(P,x) \cap X_a \neq \emptyset\}$$

Assume that $\underline{P}_g^{K_S}X_a$ is not C-satisfiable, which implies there exists at least one case $x \in \underline{P}_g^{K_S}X_a$ and $\alpha(x,t) \neq ?$ such that $\alpha(x,t) \neq \alpha(\text{Predc},t)$ for some attribute $t \in C$. According to the definition of $K_S(P,x)$, every member y in $K_S(P,x)$ should have the same value on attribute t and so $\alpha(y,t) \neq \alpha(\text{Predc},t)$, which contradicts the fact that $y \in X_a$. Similarly, if $\overline{P}_g^{K_S}X_a$ is not C-satisfiable, we will conclude $K_S(P,x) \cap X_a = \emptyset$, also a contradiction. □

Lemma 5. *The singleton approximation defined by tolerance characteristic set K_T is not indistinguishable for contingency sets X_a, X_b, X_c, and X_d, with respect to P, for $P \supseteq B$, i.e., $\underline{P}_g^{K_T}X \neq \overline{P}_g^{K_T}X$, for X being X_a, X_b, X_c, or X_d.*

Proof. Recall the definition of $K_T(P,x)$. A case y in $K_T(P,x)$ may have a null value on some attribute $t, t \in B$, while $\alpha(x,t) \neq *$. Clearly, x does not belong to $\underline{P}_g^{K_S}X_a$ because y is not in X_a, which invalidates condition $K_T(P,x) \subseteq X_a$. However, this does not hinder x from being a member of $\overline{P}_g^{K_S}X_a$. The lemma follows. □

The next part of this appendix details the proof of Theorem 7. To this purpose, we first show the subsumptive properties of the two types of characteristic sets. Let A and B denote the set of attributes defined in Section 3.2, $B \subseteq A$, and K denote the similarity (K_S) or tolerance characteristic set (K_T).

Lemma A-1. *$K(A,x) \subseteq K(B,x)$.*

Proof. Consider any case $y \in K(A,x)$. Clearly $y \in K(B,x)$ as well according to the definition of characteristic set. On the other hand, consider $y \in K(B,x)$. If $\alpha(y,t) \neq \alpha(x,t)$ and $\alpha(x,t) \neq ?$ or $*$ for some attribute $t \in A - B$, then $y \notin K(A,x)$. The lemma follows. □

Let X be X_a, X_b, X_c, or X_c. We further analyze the subsumptive relations between different types of rough-set-based approximations. For simplicity, we only show the relations facilitating comparative analysis of our proposed six measurements.

Lemma A-2.

(a) $\underline{A}_g^{K}X \supseteq \underline{B}_g^{K}X$.

(b) $\overline{A}_g^{K}X \subseteq \overline{B}_g^{K}X$.

Proof. Consider $x \in \underline{A}_g^{K}X$. According to the definition, we have $K(A,x) \subseteq X$. Since $K(A,x) \subseteq K(B,x), K(B,x) \subseteq X$ may not be held, which implies x may not be in $\underline{B}_g^{K}X$. On the other hand, consider $x \in \underline{B}_g^{K}X$. By definition and $K(A,x) \subseteq K(B,x)$, we have $K(A,x) \subseteq X$, leading to $x \in \underline{A}_g^{K}X$. Thus, $\underline{A}_g^{K}X \supseteq \underline{B}_g^{K}X$. By similar approach, we can show $\overline{A}_g^{K}X \subseteq \overline{B}_g^{K}X$. □

Then, according to Lemmas A-2, it is easy to obtain the following result.

Corollary A-3. *Accuracy(M1) $\geqslant$ Accuracy(M2), Accuracy(M3) $\geqslant$ Accuracy(M4).*

Next, we compare method M3 with M5, and M4 with M6.

Lemma A-4.

(a) $\underline{B}_g^{K_T}X = \underline{B}_c^{K_T}X$.

(b) $\overline{B}_g^{K_T}X = \overline{B}_c^{K_T}X$.

Proof.

(a) Consider any case $x \in \underline{B}_g^{K_T}X$. According to the definition, we have $K_T(B,x) \subseteq X$ and so $x \in X$. To show $x \in \underline{B}_c^{K_T}X$, we have to show that there exist some $y, x \in K_T(B,y)$ such that $y \in X$ and $K_T(B,y) \subseteq X$. This is trivial since x itself satisfies the condition. Hence, $x \in \underline{B}_c^{K_T}X$. Next, we consider any case $x \in \underline{B}_c^{K_T}X$. According to definition, x must belong to some characteristic set, say $K_T(B,y)$, for $y \in X$ and $K_T(B,y) \subseteq X$. Clearly $x \in X$. Then for all $t \in B$, $\alpha(x,t) = \alpha(y,t)$, leading to $K_T(B,x) = K_T(B,y) \subseteq X$. That it, $x \in \underline{B}_g^{K_T}X$, which completes the proof for $\underline{B}_g^{K_T}X = \underline{B}_c^{K_T}X$.

(b) Now, consider $x \in \overline{B}_g^{K_T}X$. According to definition, we know $K_T(B,x) \cap X \neq \varnothing$. To show that $x \in \overline{B}_c^{K_T}X$ we have to show that there exists a case $y, x \in K_T(B,y)$, such that $y \in X$ and $K_T(B,y) \cap X \neq \varnothing$. Let $y \in K_T(B,x) \cap X$. It is easy to show $x \in K_T(B,y)$. That is, y is just the case we need. Hence, $x \in \overline{B}_c^{K_T}X$. On the other hand, let $x \in \overline{B}_c^{K_T}X$. Then, there must exist some $y \in X$ and $K_T(B,y) \cap X \neq \varnothing$, such that $x \in K_T(B,y)$. It is easy to show $y \in K_T(B,x)$. Since $y \in X$, we obtain $K_T(B,x) \cap X \neq \varnothing$, leading to $x \in \overline{B}_c^{K_T}X$. This proves $\overline{B}_g^{K_T}X = \overline{B}_c^{K_T}X$. □

Corollary A-5. *Accuracy(M4) = Accuracy(M6).*

Lemma A-6.

(a) $\underline{A}_g^{K_T}X \subseteq \underline{A}_c^{K_T}X$.

(b) $\overline{A}_g^{K_T}X = \overline{A}_c^{K_T}X$.

Proof. The proof is similar to that for Lemmas A-4. The only difference is that if $x \in \underline{A}_c^{K_T}X$, then x may not be in $\underline{A}_g^{K_T}X$. This is because the fact that $x \in X$ does not guarantee $K_T(A,x) \subseteq X$ even we know $x \in K_T(A,y)$, for $y \in X$ and $K_T(A,y) \subseteq X$. □

Corollary A-7. *Accuracy(M3) $\leqslant$ Accuracy(M5).*

Lemma A-8.

(a) $\underline{B}_g^{K_S}X \supseteq \underline{A}_c^{K_T}X$.

(b) $\overline{B}_g^{K_S}X = \overline{A}_c^{K_T}X$.

Proof.

(a) Consider $x \in \underline{B}_g^{K_S}X$. By definition, we know $K_S(B,x) \subseteq X$ and so $x \in X$. To show that $x \in \underline{A}_c^{K_T}X$, we have to show that there exists a case $y, x \in K_T(A,y)$, such that $y \in X$ and $K_T(A,y) \subseteq X$. First, suppose ydoes exist, i.e., $x \in K_T(A,y)$, which implies that $\forall t \in B$, $\alpha(x,t) = \alpha(y,t) \neq *$ or $\alpha(y,t) = *$. In the latter case, clearly $y \notin X$, while the first case does not guarantee $K_T(A,y) \subseteq X$; for example, consider a case $w \in K_T(A,y)$ with $\alpha(w,t) = *$ for some $t \in B$. Therefore, we conclude that x may not belong to $\underline{A}_c^{K_T}X$. On the other hand, suppose $x \in \underline{A}_c^{K_T}X$, which implies $\exists y \in X$ and $K_T(A,y) \subseteq X$, such that $x \in K_T(A,y)$. Clearly, $x \in X$, implying $\alpha(x,t) \neq ?$ for all $t \in B$, and so $\forall w \in K_S(B,x)$, $\alpha(w,t) = \alpha(x,t)$ or all $t \in B$. Thus, $K_S(B,x) \subseteq X$, and so $x \in \underline{B}_g^{K_S}X$. It follows that $\underline{B}_g^{K_S}X \supseteq \underline{A}_c^{K_T}X$.

(b) Consider $x \in \overline{B}_g^{K_S}X$. By definition, we have $K_S(B,x) \cap X \neq \varnothing$. Let $y \in K_S(B,x) \cap X$. The fact that $y \in K_S(B,x)$ means $\forall t \in B$, $\alpha(x,t)$ = $\alpha(y,t)$ = ? or $\alpha(x,t)$ = ?, and so we have $x \in K_T(B,y)$ (Note that '?' is interpreted as '$*$' as we consider K_T). Since $y \in K_T(B,y), K_T(B,y) \cap X \neq \varnothing$. Hence, $x \in \overline{A}_c^{K_T}X$. Next, consider $x \in \overline{A}_c^{K_T}X$. By definition, there must exist some $y \in X$ and $K_T(A,y) \cap X \neq \varnothing$, such that $x \in K_T(A,y)$. Clearly, $x \in K_T(B,y)$ by Lemmas A-1. By similar statement, we know $y \in K_S(B,x)$ (In this case, '$*$' is interpreted as '?'). Since $y \in X$ and $y \in K_S(B,x)$, i.e., $K_S(B,x) \cap X \neq \varnothing$, we conclude $x \in \overline{B}_g^{K_S}X$. It follows that $\underline{B}_g^{K_S}X = \underline{A}_c^{K_T}X$. □

Corollary A-9. *Accuracy(M2) $\geqslant$ Accuracy(M5).*

Finally, combing the results in Corollaries A-3, A-5, A-7, and A-9 we obtain the result in Theorem 7.

References

[1] J.S. Almenoff, K.K. LaCroix, N.A. Yuen, D. Fram, W. DuMouchel, Comparative performance of two quantitative safety signalling methods: implications for use in a pharmacovigilance department, Drug Safety 29 (10) (2006) 875–887.

[2] A. Bate, Bayesian confidence propagation neural network, Drug Safety 30 (7) (2007) 623–625.

[3] A. Bate, M. Lindquist, I.R. Edwards, S. Olsson, R. Orre, A. Lansner, R.M. De Freitas, A Bayesian neural network method for adverse drug reaction signal generation, Eur. J. Clin. Pharmacol. 54 (4) (1998) 315–321.

[4] B.K. Chen, Y.T. Yang, Post-marketing surveillance of prescription drug safety: past, present, and future, J. Legal Med. 34 (2) (2013) 193–213.

[5] P.M. Coloma, G. Trifirò, V. Patadia, M. Sturkenboom, Postmarketing safety surveillance: where does signal detection using electronic healthcare records fit into the big picture?, Drug Safety 24 (6) (2013) 343–348

[6] G. Deshpande, V. Gogolak, S.W. Smith, Data mining in drug safety: review of published threshold criteria for defining signals of disproportionate reporting, Pharmac. Med. 24 (1) (2010) 37–43.

[7] A.C. Egberts, R.H. Meyboom, E.P. van Puijenbroek, Use of measures of disproportionality in pharmacovigilance: three Dutch examples, Drug Safety 25 (6) (2002) 453–458.

[8] S.J. Evans, P.C. Waller, S. Davis, Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports, Drug Safety 10 (6) (2001) 483–486.

[9] FDA Adverse Event Reporting System, <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm083765.htm>.

[10] J.W. Grzymala-Busse, Rough set strategies to data with missing attribute values, in: Proc. Workshop on Foundations and Novel Approaches in Data Mining, IEEE ICDM2003, vol. 9, 2006, pp. 197–212.

[11] A.E. Hassanien, A. Abraham, J.F. Peters, G. Schaefer, Rough sets in medical informatics applications, in: J. Mehnen, M. Koeppen, A. Saad, A. Tiwari (Eds.), Applications of Soft Computing, vol. 58, Springer-Verlag, Heidelberg, 2009, pp. 23–30.

[12] R. Harpaz, W. DuMouchel, N.H. Shah, D. Madigan, P. Ryan, C. Friedman, Novel data mining methodologies for adverse drug event discovery and analysis, Clin. Pharmacol. Therapeut. 91 (6) (2012) 1010–1021.

[13] J. Hopstadius, G.N. Norén, A. Bate, I.R. Edwards, Impact of stratification on adverse drug reaction surveillance, Drug Safety 31 (11) (2008) 1035–1048.

[14] M. Kryszkiewicz, Rough set approach to incomplete information systems, in: Proc. 2nd Annual Joint Conf. on Information Sciences, 1995, pp. 194–197.

[15] MedWatch: The FDA Safety Information and Adverse Event Reporting Program, <http://www.fda.gov/Safety/MedWatch/>.

[16] Z. Pawlak, Rough sets, Int. J. Comput. Inform. Sci. 11 (5) (1982) 341–356.

[17] S. Rissino, G.L. Torres, Rough set theory — fundamental concepts, principals, data extraction, and applications, in: J. Ponce, A. Karahoca (Eds.), Data Mining and Knowledge Discovery in Real Life Applications, I-Tech Education and Publishing, Vienna, 2009, pp. 36–58.

[18] J. Stefanowski, A. Tsoukias, On the extension of rough sets under incomplete information, in: Proc. 7th Int. Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing, 1999, pp. 73–81.

[19] P.G.M. van der Heijden, E.P. van Puijenbroek, S. van Buuren, J.W. van der Hofstede, On the assessment of adverse drug reactions from spontaneous reporting systems: the influence of under-reporting on odds ratios, Stat. Med. 21 (14) (2002) 2027–2044.